Hindawi
Mathematical Problems in Engineering
Volume 2018, Article ID 7952974, 14 pages
https://doi.org/10.1155/2018/7952974

Research Article

Key-Skeleton-Pattern Mining on 3D Skeletons Represented by Lie Group for Action Recognition

Guang Li, Kai Liu, Wenwen Ding, Fei Cheng, and Boyang Chen

School of Computer Science and Technology, Xidian University, Xi'an, China

Correspondence should be addressed to Kai Liu; kailiu@mail.xidian.edu.cn

Received 11 April 2018; Revised 28 October 2018; Accepted 7 November 2018; Published 5 December 2018

Academic Editor: Paolo Spagnolo

The human skeleton can be considered as a tree system of rigid bodies connected by bone joints. In recent researches, substantial progress has been made in both theories and experiments on skeleton-based action recognition. However, it is challenging to accurately represent the skeleton and precisely eliminate noisy skeletons from the action sequence. This paper proposes a novel skeletal representation, which is composed of two subfeatures to recognize human action: static features and dynamic features. First, to avoid scale variations from subject to subject, the orientations of the rigid bodies in a skeleton are employed to capture the scale-invariant spatial information of the skeleton. The static feature of the skeleton is defined as a combination of the orientations. Unlike previous orientation-based representations, the orientation of a rigid body in the skeleton is defined as the rotations between the rigid body and the coordinate axes in three-dimensional space. Each rotation is mapped to the special orthogonal group $SO(3)$. Next, the rigid-body motions between the skeleton and its previous skeletons are utilized to capture the temporal information of the skeleton. The dynamic feature of the skeleton is defined as a combination of the motions. Similarly, the motions are represented as points in the special Euclidean group $SE(3)$. Therefore, the proposed skeleton representation lies in the Lie group $(SE(3)\times\cdots\times SE(3), SO(3)\times\cdots\times SO(3))$, which is a manifold. Using the proposed representation, an action can be considered as a series of points in this Lie group. Then, to recognize human action more accurately, a new pattern-growth algorithm named MinP-PrefixSpan is proposed to mine the key-skeleton-patterns from the training dataset. Because the algorithm reduces the number of new patterns in each growth step, it is more efficient than the PrefixSpan algorithm. Finally, the key-skeleton-patterns are used to discover the most informative skeleton sequences of each action (skeleton sequence). Our approach achieves accuracies of 94.70%, 98.87%, and 95.01% on three action datasets, outperforming other relative action recognition approaches, including LieNet, Lie group, Grassmann manifold, and Graph-based model.

1. Introduction

Human action recognition is currently the most dynamic research topic in the field of computer vision, owing to its applications in intelligent surveillance, video games, robotics, and other fields. Several approaches have been proposed to recognize human action from RGB video sequences over the past few decades [1], but their performance is unsatisfactory because RGB data are very sensitive to factors such as perspective changes, occlusions, and background clutter. Although significant research results have been achieved, human action recognition remains a challenging problem.

Because the human skeleton can generally be regarded as an articulated system of rigid segments, which are connected by joints, human action can be viewed as a continuous evolution of the spatial configuration, which is constructed by these rigid segments [2]. Therefore, if human skeleton sequences can be accurately extracted from RGB videos, action recognition can be performed by classifying these sequences. However, it is very difficult to accurately extract a skeleton sequence from RGB videos [3]. With the advent of cost-effective RGB-D cameras, it has become easier to extract the three-dimensional (3D) human skeleton from depth maps. Although this improves the appearance and viewpoint variations to a certain extent [4–7], the following two challenges cause large intraclass variations and remain unresolved. First, different people can perform the same action in different ways. Second, the 3D human skeleton

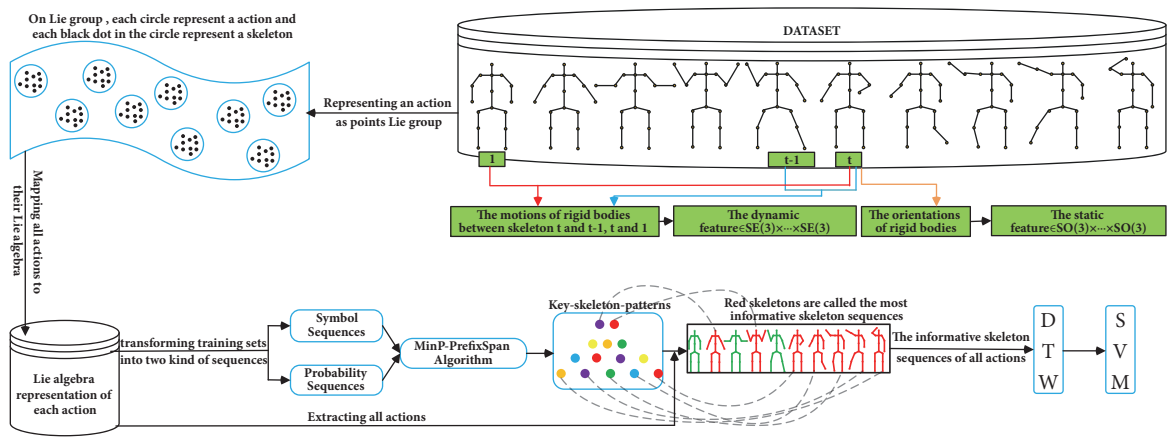

Figure 1: The general framework of the proposed method.

is sometimes imprecise because depth maps include noisy information. However, a psychological research found that humans can easily recognize an action from a pose sequence [8]. According to the work of [8], Yang et al. considered that actions can be classified by a single key pose [9]. This suggests that a set of key skeletons can be used to perform action classification rather than the entire skeleton sequence. Since the representation of the key poses is robust to outlier poses, this approach should improve the accuracy of action recognition as long as the key poses are accurate.

The general framework of the proposed approach is shown in Figure 1. By observing human action in daily life, the orientations and motions of rigid bodies can include a lot of useful information for action recognition. In this paper, a new skeletal representation, which is composed of the static feature and dynamic feature, is proposed for 3D skeleton-based action recognition. The static feature is used to represent the spatial information in a given skeleton t. To capture the scale-invariant spatial information, the orientations of the rigid bodies in the skeleton are employed to construct its static feature. In this work, the orientation of a rigid body in the skeleton is represented as six rotation matrices between the rigid body and the three coordinate axes in 3D space. The rotation matrices are mapped to the special orthogonal group $SO(3)$[10]. Next, the dynamic feature is employed to represent the temporal information of skeleton t. The dynamic feature is composed of the rigid-body motions between skeletons t and t-1 and those between skeletons t and 1 (the three skeletons belong to the same sequence or action). The motions are represented as points in the special Euclidean group $SE(3)$[11]. Hence, skeleton t is represented by a point in Lie Group $(SE(3) \times SE(3) \cdots \times SE(3), SO(3) \times SO(3) \cdots \times SO(3))$,where the operation $\times$ represents the direct product between groups in group theory. Using the proposed skeleton representation, a human action (skeleton sequence) can be represented as points in the Lie group. However, it is typically a very complicated task to classify human actions represented by a Lie group directly. Many standard classification approaches, such as the support vector machine (SVM)[12] approach, are not directly applicable to Lie groups. To overcome the classification difficulties, the actions (skeleton sequences) are mapped from the Lie Group to its Lie algebra $(\mathfrak{se}(3) \times \mathfrak{se}(3) \times \cdots \times \mathfrak{se}(3), \mathfrak{so}(3) \times \mathfrak{so}(3) \times \cdots \times \mathfrak{so}(3))$, which are the elements of the tangent space of the manifold at the identity element. The Lie algebra is a vector space, which makes action classification easier.

An action (skeleton sequence) usually includes many noisy skeletons, which can reduce the action recognition accuracy. In this study, the key-skeleton-patterns are used to eliminate noisy skeletons from an action, and the remaining skeletons in the action are called the most informative skeleton sequence. First, a pattern is defined as a short skeleton sequence, which is not necessarily adjacent in the original skeleton sequences. If the short skeleton sequence appears in many skeleton sequences of an action class, the pattern is called the key-skeleton-pattern in that class. Next, to mine the key-skeleton-patterns, k-means is used to learn the symbolic dictionary from all skeletons in the dataset. Each symbol in the dictionary represents a class of similar skeletons, which means that each skeleton is quantized (represented) by a symbol in the dictionary. Then, a skeleton sequence can be represented as a symbol sequence. In this paper, probability is used to measure the distance between a skeleton and its corresponding symbol in order to minimize the effect of quantization errors (e.g., two different skeletons are quantized by the same symbol). Hence, each skeleton is represented by a distance-based probability, and an action is represented as a probability sequence. Then, a new pattern-growth algorithm named MinP-PrefixSpan is proposed to mine the key-skeleton-patterns of an action class from the symbol sequences and the probability sequences that correspond to the action class. Compared with the PrefixSpan algorithm, our algorithm achieves higher efficiency by reducing the number of new skeleton patterns in each growth step. Finally, the key-skeleton-patterns are utilized to eliminate noisy skeletons from the action in order to capture the most informative skeleton sequence of the action. An SVM is employed to classify the most informative skeleton sequences.

The main contributions of this study are as follows. (1)To capture the scale-invariant skeletal information, the

orientations of rigid bodies in a skeleton are utilized to construct the static feature. Different from previous orientation-based approaches, in this study, a rigid-body orientation is represented as six rotation matrices, and each rotation matrix is represented as a point in $SO(3)$. (2) Traditional approaches based on Lie groups [5, 13, 14] only consider the spatial information of a skeleton but ignore the temporal information between different skeletons. Therefore, our approach employs the rigid-body motions between different skeletons to describe the temporal variation. Likewise, the motions can be represented as points on $SE(3)$. (3) Traditional approaches also ignore the influence of noisy skeletons in an action on the accuracy of the action recognition. In this study, based on the PrefixSpan algorithm [15] in data mining, a new pattern-growth algorithm is proposed to mine the key-skeleton-patterns of each action class, and the key-skeleton-patterns are used to eliminate noisy skeletons.

2. Related Work

A brief overview of the related work on human action recognition approaches based on skeletons is provided in this section, and various sequential pattern-mining algorithms are reviewed.

The existing skeleton-based action recognition approaches can be classified into three main categories. The first class of approach ignores the influence of noisy skeletons on the accuracy of action recognition. Slama et al. represented an action by an observability matrix, which was characterized by an element of a finite Grassmann manifold [16]. However, their method does not eliminate noisy skeletons from an action, and it is insufficient to estimate the approximation of an extended observability sequence with a finite Grassmann manifold. Ding et al. divided actions into subactions and used the profile hidden Markov model(HMM) to align them [13]. Although their approach accurately extracts the spatial features of an action, it does not solve the following two problems: eliminating noisy skeletons and reducing the time complexity of the profile HMMs. Liu et al. proposed a new spatiotemporal representation, called "Skepxels," to transform skeleton videos into images of flexible dimensions, and employed the resulting images to build a CNN-based framework for effective human action recognition[17]. Likewise, their approach does not eliminate noisy skeletons from an action. In this study, the key-skeleton-patterns of an action are utilized to eliminate noisy skeletons from the action as an approach to improve the accuracy of action recognition.

The second class of approach ignores scale variations from subject to subject, which means that the spatial feature of an action cannot be accurately represented. Chaudhry et al. hierarchically divided the human skeleton into smaller parts and employed certain bio-inspired shape features to represent each part [18]. The temporal evolutions of these bio-inspired features are modeled by linear dynamical systems (LDSs). Although their approach takes full advantage of the correlation between the skeletal parts, it ignores the feature of the rigid bodies in a skeleton and the scale variations between different subjects. Xia et al. proposed a view-invariant representation of the human skeleton using histograms of 3D joint locations [19]. The temporal evolutions of this skeletal representation are modeled by a discrete HMM. However, their approach not only ignores the relativity between the rigid bodes in a skeleton but also the normalization of the skeleton data. Li et al. represented an action by a special graph based on the top-K relative variance of joint relative distance (RVJRD) [20]. One potential limitation of this approach is that the graph-based model does not handle scale variations, which may cause incorrect spatial information to be selected by the top-K RVJRD. In contrast, our proposed approach uses the orientations of the rigid bodies in a skeleton to capture scale-invariant skeletal features.

The third class of approach ignores the temporal information of an action and treats the poses in the action independently. Evangelidis et al. used a local skeleton descriptor to encode the relative positions of joint quadruples [21]. The descriptor of an action was represented by a multilevel Fisher vector composed of the local skeleton descriptor in the action. However, the action descriptor not only ignores the temporal information between different skeletons but also has high time complexity. Huang et al. combined the Lie group structure with a deep network framework [22]. Their learning structure (LieNet) has a rotation mapping layer transforming the Lie group features into the traditional neural network model. One main limitation of this approach is that LieNet ignores the rich temporal information of human actions. Vemulapalli et al. described the relative geometry between the rigid-body parts using special Euclidean group $SE(3)$[5]. Therefore, the entire skeleton in an action can be represented as a point in $SE(3)$. An action is represented as a curve in the Lie group $(SE(3) \times SE(3) \cdots \times SE(3))$. Although their approach can accurately extract the spatial information of a skeleton, it ignores the temporal cues between the skeletons in an action and does not eliminate noisy skeletons from the action. Our proposed dynamic feature models the temporal structures of an action using the rigid-body motions between different skeletons in the action.

Sequential pattern mining aims to discover frequent subsequences as patterns in a sequence database. Traditional sequential pattern mining algorithms [23–26] are usually used to mine frequent sequential patterns from deterministic databases. However, those approaches cannot be indirectly applied to uncertain data (or probabilistic data). Unfortunately, the existing pattern mining algorithm on an uncertain dataset [27, 28] is not adopted to our probabilistic sequence model. Therefore, considering the amount of noise in our uncertain datasets (probabilistic datasets), a new pattern-growth algorithm is proposed to mine the key-skeleton-patterns from the datasets.

3. Proposed Framework

3.1. Fundamental Concepts. In this subsection, a brief overview of the special Euclidean group $SE(3)$ and the special orthogonal group $SO(3)$ is presented, which is necessary for further understanding of the Lie group. We refer the readers to [2, 10, 11] for a general introduction to Lie groups. Important notations are shown in Table 1.

Table 1: Notations involving Lie groups.

Notation	Description
$SO(3)$	Special orthogonal group
$\mathfrak{so}(3)$	Lie algebra of $SO(3)$
$SO(3)$	Special Euclidean group
$\mathfrak{se}(3)$	Lie algebra of $SE(3)$
$\times$	Direct product of Lie group

3.1.1. Special Orthogonal Group. The special orthogonal group $SO(3)$ is a Lie group, which can be represented by all 3×3 orthogonal matrices shown as follows:

$$SO(3) = \left\{A \in R^{3\times3}; A^T A = I_3, \det A = 1\right\} \tag{1}$$

where I_3 denotes the 3×3 identity matrix and A is a rotation matrix. In 3D space, a rotation A is an element of $SO(3)$ and can transform a vector $x = [x_1, x_2, x_3]^T$ to x' by

$$x' = Ax \tag{2}$$

Every group $SO(3)$ has an associated Lie algebra of $SO(3)$ that is the tangent space around the identity element I_3. The Lie algebra of $SO(3)$, denoted by $\mathfrak{so}(3)$, is a set of all real 33 skew-symmetric matrices as follows:

$$\mathfrak{so}(3) = \left\{M \in R^{3\times3}; M^t = -M\right\}. \tag{3}$$

Given an element

$$\Omega = \begin{bmatrix} 0 & -m_3 & m_2 \\ m_3 & 0 & -m_1 \\ -m_2 & m_1 & 0 \end{bmatrix} \in \mathfrak{so}(3) \tag{4}$$

its vector form $vec(\Omega)$ is

$$vec(\Omega) = [m_1, m_2, m_3]^T \tag{5}$$

The exponential map $exp_{SO(3)}$ from $\mathfrak{so}(3)$ to $SO(3)$ and the logarithm map $log_{SO(3)}$ from $SO(3)$ to $\mathfrak{so}(3)$, respectively, are

$$\begin{aligned} exp_{SO(3)}(\Omega) &= e^{\Omega}; \\ log_{SO(3)}(A) &= log(A). \end{aligned} \tag{6}$$

3.1.2. Special Euclidean Group. The special Euclidean group SE(3) is a Lie group, which is a set of 4 by 4 matrices

$$SE(3) = \left\{H \in R^{4\times4}, H = \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix}, A \in SO(3), b \in R^3\right\}. \tag{7}$$

The matrix representation also makes SE(3) action on points $c \in R^3$ by rotating and translating them:

$$c' = H \cdot c = \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix} \cdot c = \begin{bmatrix} Ac + b \\ 1 \end{bmatrix}. \tag{8}$$

Every Lie Group SE(3) can be associated with a Lie algebra $\mathfrak{se}(3)$, which is the tangent space of the Lie group SE(3) at the identity matrix I_4. Note that $\mathfrak{se}(3)$ is a 6D vector space that can be formed by 4 by 4 matrices

$$\mathfrak{se}(3) = \left\{\begin{bmatrix} M & v \\ 0 & 0 \end{bmatrix}, M \in R^{3\times3}, v \in R^3, M^T = M\right\}. \tag{9}$$

Given an element

$$\Omega = \begin{bmatrix} 0 & -m_3 & m_2 & v_1 \\ m_3 & 0 & -m_1 & v_2 \\ -m_2 & m_1 & 0 & v_3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \in \mathfrak{se}(3), \tag{10}$$

its vector form $vec(\Omega)$ is

$$vec(\Omega) = [m_1, m_2, m_3, v_1, v_2, v_3]^T. \tag{11}$$

The exponential map $exp_{SE(3)}$ from $\mathfrak{se}(3)$ to SE(3) and the logarithm map $log_{SE(3)}$ from SE(3) to $\mathfrak{se}(3)$, respectively, are

$$\begin{aligned} exp_{SE(3)}(\Omega) &= e^{\Omega}; \\ log_{SE(3)}(H) &= log(H). \end{aligned} \tag{12}$$

$\mathbf{SO(3)} \times \cdots \times \mathbf{SO(3)}$ and $\mathbf{SE(3)} \times \cdots \times \mathbf{SE(3)}$: the direct product $\times$ is used to combine multiple $SO(3)$, which form a new Lie group $K = SO(3) \times \cdots \times SO(3)$ with identity element $(I_3, \ldots, I_3)$ and its Lie algebra $k = \mathfrak{so}(3) \times \cdots \times \mathfrak{so}(3)$. The exponential of $(\Omega_1, \ldots, \Omega_N) \in k$ and the logarithm map of $(A_1, \ldots, A_N) \in K$, respectively, are given by

$$exp_K(\Omega_1, \ldots, \Omega_N) = \left(e^{\Omega_1}, \ldots, e^{\Omega_N}\right) \tag{13}$$

$$log_K(A_1, \ldots, A_N) = (log(A_1), \ldots, \log(A_N)). \tag{14}$$

The vector form of $log_K((A_1, \ldots, A_N))$ is

$$\begin{aligned} &vec\left(log_K(A_1, \ldots, A_N)\right) \\ &\quad = vec(log(A_1), \ldots, vec(log(A_N)). \end{aligned} \tag{15}$$

Similarly, a new Lie group $K = SE(3) \times \cdots \times SE(3)$ with identity element$(I_4, I_4, \ldots, I_4)$ and its Lie algebra $k = \mathfrak{se}(3) \times \mathfrak{se}(3) \times \cdots \times \mathfrak{se}(3)$ are formed by using the direct product $\times$. The exponential map of $(\Omega_1, \ldots, \Omega_N) \in k$ and the logarithm map of $(H_1, \ldots, H_N) \in K$, respectively, are given by

$$exp_K(\Omega_1, \ldots, \Omega_N) = \left(e^{\Omega_1}, \ldots, e^{\Omega_N}\right) \tag{16}$$

$$log_K(H_1, \ldots, H_N) = (log(H_1), \ldots, \log(H_N)). \tag{17}$$

The vector form of $log_K((H_1, \ldots, H_N))$ is

$$\begin{aligned} &vec\left(log_K(H_1, \ldots, H_N)\right) \\ &\quad = vec(log(H_1), \ldots, vec(log(H_N)). \end{aligned} \tag{18}$$

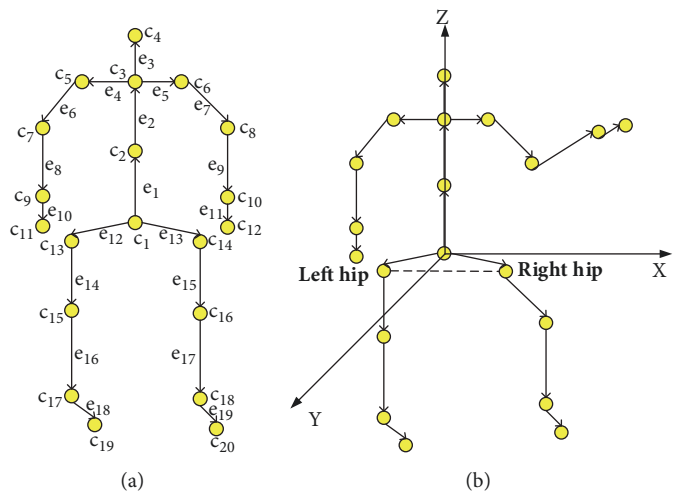

Figure 2: (a) Human skeleton with 19 rigid body parts and 20 bone joints. (b) Placing the hip center to the origin and aligning the x-axis with the ground plane projection of the vector from the left hip to the right hip.

3.1.3. Explanation of Fundamental Concepts. According to the concepts described in Section 3.2.1, the orientation of a rigid body is represented by six rotation matrices. Mathematically, a rotation matrix is a point in $SO(3)$; therefore, the orientation of a rigid body in skeleton t can be represented as six points in $SO(3)$. Then the static feature, composed of the orientations of the rigid bodies in the skeleton, is represented as a point in the Lie group $(SO(3) \times SO(3) \cdots \times SO(3))$, as shown in Figure 1.

The motions of a rigid body is generally regarded as its rotations and translations in 3D space. Mathematically, the rotations and translations of a rigid body are defined as $SE(3)$; therefore, a rigid-body motion between skeletons t and $t-1$ (or skeletons t and 1) can be represented as a point in $SE(3)$. Then, the dynamic feature, which is composed of the rigid-body motions between skeletons t and $t-1$ and those between skeletons t and 1, is represented as a point in the Lie group $(SE(3) \times SE(3) \cdots \times SE(3))$, as shown in Figure 1. A skeletal representation, composed of the static feature and dynamic feature, can be represent as a point in the Lie group $(SE(3) \times SE(3) \cdots \times SE(3), SO(3) \times SO(3)$.

The wavy surface represents a Lie group $(SE(3) \times SE(3) \cdots \times SE(3), SO(3) \times SO(3) \cdots \times SO(3))$ in Figure 1. A whole circle in the wavy surface represents an action (skeleton sequence). Each black dot in the circle represents a skeleton. Then, an action can be represented as points in the Lie group (the points are included in the same circle). To overcome the classification difficulties, an action (or a whole circle) is mapped from the Lie group to its Lie algebra $(\mathfrak{se}(3) \times \mathfrak{se}(3) \times \cdots \times \mathfrak{se}(3), \mathfrak{so}(3) \times \mathfrak{so}(3) \times \cdots \times \mathfrak{so}(3))$, as shown in Figure 1. In fact, the Lie algebra is a vector space.

3.2. Extraction of Skeleton Features. In this subsection, the static and dynamic features of a skeleton are represented as a point in the Lie group. Let $\mathbf{Z} = (\mathbf{C}, \mathbf{E})$ be a skeleton. The set of bone joints is denoted by $\mathbf{C} = \{c_1, c_2, \ldots, c_N\}$, and the set of rigid-body parts is denoted by $\mathbf{E} = \{e_1, e_2, \ldots, e_M\}$, where $c, e \in R^3$. Figure 2(a) shows an example of the human skeleton with 19 rigid-body parts and 20 bone joints.

3.2.1. Static Feature of Skeleton. By observing human action, the orientations of rigid bodies in a skeleton (pose) can include a lot of valuable information for action recognition. To describe the orientation of a given rigid body e_m, the global coordinate system is translated to the local coordinate system. $A_{m,x}$, $A_{m,y}$, and $A_{m,z}$ represent the three rotations that transform the rigid body e_m to the three coordinate axes, as shown in Figure 3(c). Their rotation relationship is shown as follows:

$$\begin{aligned} r_x &= A_{m,x} e_m, \\ r_y &= A_{m,y} e_m, \\ r_z &= A_{m,z} e_m, \end{aligned} \tag{19}$$

where $A_{m,x}$, $A_{m,y}$, and $A_{m,z} \in SO(3)$. Similarly, $A_{x,m}$, $A_{y,m}$, and $A_{z,m}$ represent the three rotations that transform r_x, r_y, and r_z to the rigid body e_m, respectively, as shown in Figure 3(d). Their rotation relationship is shown as follows:

$$\begin{aligned} e_m &= A_{x,m} r_x, \\ e_m &= A_{y,m} r_y, \\ e_m &= A_{z,m} r_z, \end{aligned} \tag{20}$$

where $A_{x,m}$, $A_{y,m}$, and $A_{z,m} \in SO(3)$. The six rotations can be used to describe the orientation of the rigid bodies.

Given a skeleton t, $o_m(t) = \{A_{m,x}(t), A_{x,m}(t), A_{m,y}(t), A_{y,m}(t), A_{m,z}(t), A_{z,m}(t)\}$ is used to represent the orientation of $e_m(t)$ in the skeleton. In this work, the skeletal **static feature** is defined as a set of the orientations of the rigid bodies in the skeleton as follows:

$$\begin{aligned} f_o(t) &= \{o_1(t), o_2(t), \ldots, o_M(t)\} \\ &\in SO(3) \times SO(3) \times \cdots \times SO(3), \end{aligned} \tag{21}$$

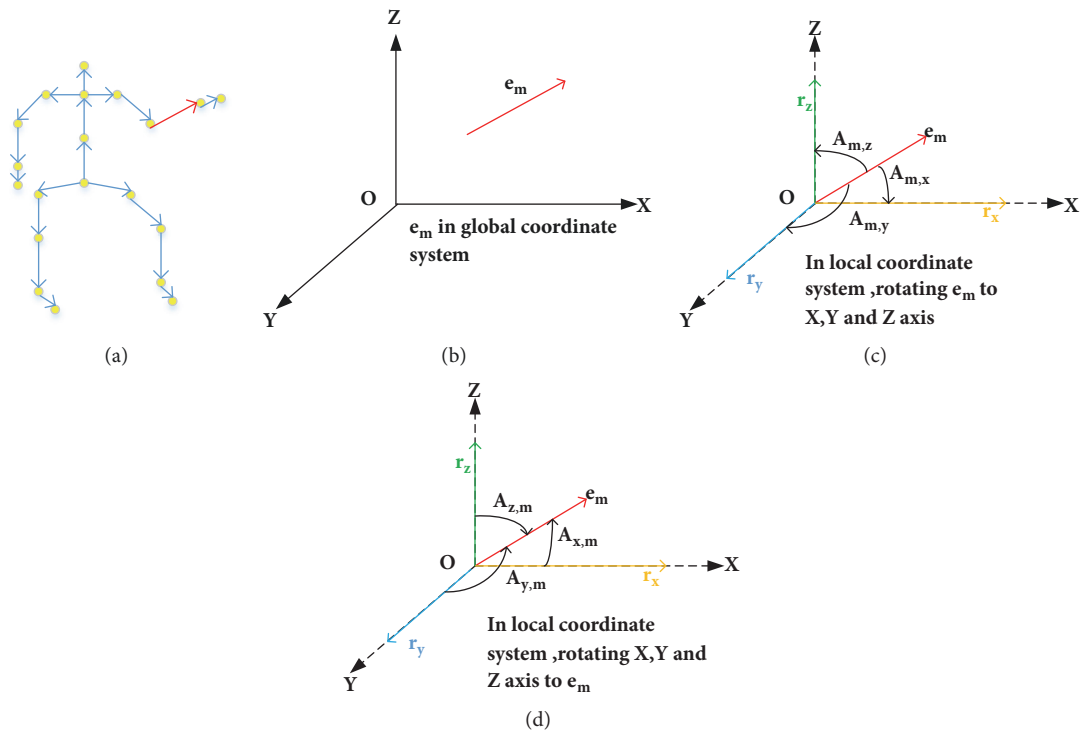

Figure 3: Representation of the orientation of rigid body e_m.

where M is the total number of rigid bodies in the human skeleton.

3.2.2. Dynamic Feature of Skeleton. Rigid-body motion is generally regarded as rotations and translations in 3D space. Mathematically, the rotations and translations of a rigid body can be denoted by $SE(3)$. In this study, $SE(3)$ is employed to describe rigid-body motions between different skeletons. Let $e_m(i) \in R^3$ be rigid body e_m in skeleton i and $e_m(j) \in R^3$ be rigid body e_m in skeleton $j(i \neq j)$.

Given a point $k(i) \in e_m(i)$ and $k(j) \in e_m(j)$ corresponding to $k(i)$, we have

$$\begin{bmatrix} k(j) \\ 1 \end{bmatrix} = H_m^{(i,j)} \begin{bmatrix} k(i) \\ 1 \end{bmatrix} = \begin{bmatrix} A_m^{i,j} & b_m^{i,j} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k(i) \\ 1 \end{bmatrix}, \quad (22)$$

where $H_m^{(i,j)} \in SE(3)$ and $A_m^{(i,j)}$ and $b_m^{(i,j)}$ are the rotation and translation, which can transform $e_m(i)$ to the position and orientation of $e_m(j)$, respectively, as shown in Figure 4(b).

Similarly, given a point $k(j) \in e_m(j)$ and $k(i) \in e_m(i)$ corresponding to $k(j)$, we have

$$\begin{bmatrix} k(i) \\ 1 \end{bmatrix} = H_m^{(j,i)} \begin{bmatrix} k(j) \\ 1 \end{bmatrix} = \begin{bmatrix} A_m^{(j,i)} & b_m^{(j,i)} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k(j) \\ 1 \end{bmatrix}, \quad (23)$$

where $H_m^{(j,i)} \in SE(3)$ and $A_m^{(j,i)}$ and $b_m^{(j,i)}$ are the rotation and translation, which can transform $e_m(j)$ to the position and orientation of $e_m(i)$, respectively, as shown in Figure 4(c). $d_m^{(i,j)} = \left(H_m^{(j,i)}, H_m^{(i,j)}\right)$ is used to represent the motion of rigid body e_m between skeletons i and j. Then, the rigid-body motions between skeletons i and j can be represented by

$$f_{rg}(i,j) = \left\{d_1^{(i,j)}, d_2^{(i,j)}, \ldots, d_M^{(i,j)}\right\} \in SE(3) \times SE(3) \times \cdots \times SE(3), \quad (24)$$

where M is the total number of the rigid bodies in the skeleton.

In this study, our approach only considers the rigid-body motions between skeletons t and t-1 and those between skeletons t and 1. According to formula (24), the rigid-body motions between skeletons t and t-1 can be represented by $f_{rg}(t, t-1)$. Similarly, the rigid-body motions between skeletons t and 1 can be represented by $f_{rg}(t, 1)$. Then, the skeletal **dynamic feature** is defined as a set of the rigid-body motions in skeleton t as follows:

$$f_{mf}(t) = \left(f_{rg}(t, t-1), f_{rg}(t, 1)\right). \quad (25)$$

Skeleton t is represented by the static feature and the dynamic feature as follows:

$$f(t) = \left\{f_o(t), f_{mf}(t)\right\}, \quad (26)$$

where $f_o(t)$ represents the static feature of the skeleton and $f_{mf}(t)$ represents its dynamic feature.

3.3. Skeleton Sequence Representation

3.3.1. Lie Group Representation of Skeleton Sequence. Using the proposed skeletal feature, a skeleton sequence or an action is represented by

$$LG = \{f(t), t \in [1, T]\}, \quad (27)$$

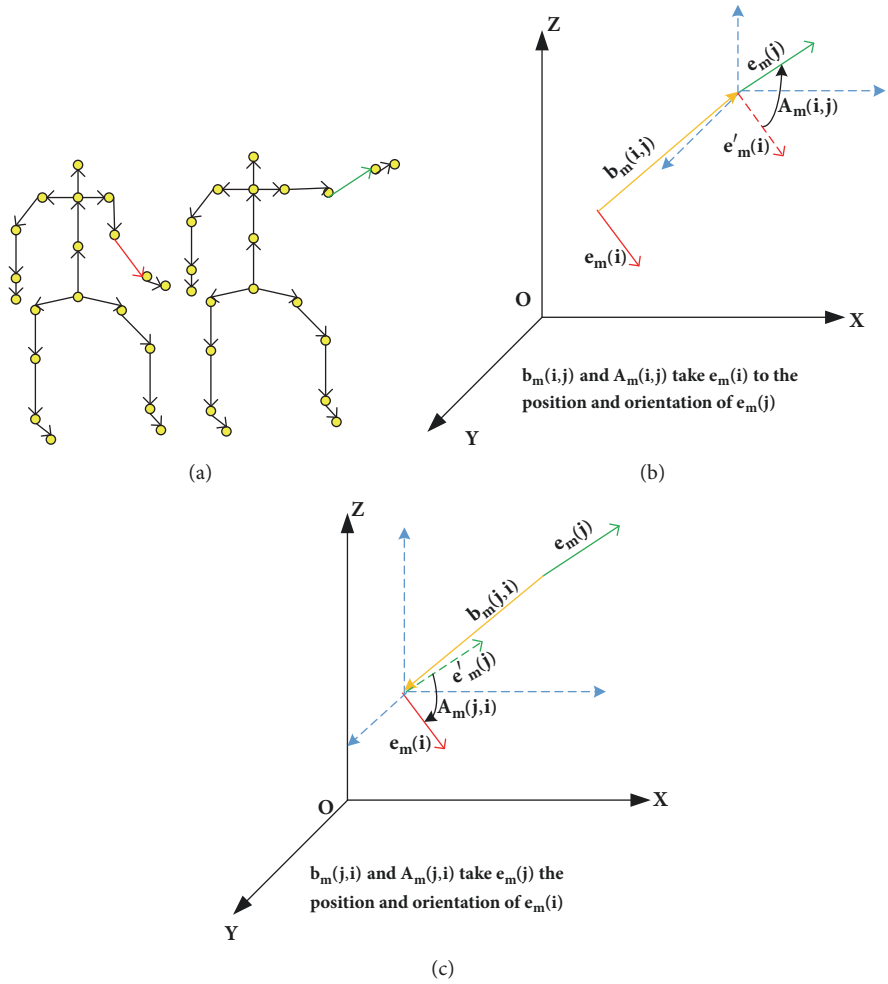

FIGURE 4: Representation of the relative geometries of rigid body e_m at time instance t.

where T is the total number of frames in the sequence and $f(t) \in$ Lie group $\{SO(3) \times \cdots \times SO(3), SE(3) \times \cdots \times SE(3)\}$.

3.3.2. Lie Algebra Representation of Skeleton Sequence. Since the most classification methods (such as SVM) cannot be directly applied to manifolds, to overcome these difficulties, $f(t)$ is mapped to its Lie algebra $(\mathfrak{so}(3) \times \cdots \times \mathfrak{so}(3), \mathfrak{se}(3) \times \cdots \times \mathfrak{se}(3))$.The Lie algebra of $f_o(t)$ is given by

$$g_o(t) = \left[vec\left(log\left(o_1(t)\right)\right), vec\left(log\left(o_2(t)\right)\right), \ldots, vec\left(log\left(o_M(t)\right)\right)\right] \in \mathfrak{so}(3) \times \mathfrak{so}(3) \times \cdots \times \mathfrak{so}(3) \quad (28)$$

and the Lie algebra of $f_{rg}(i, j)$ is given by

$$g_{rg}(i,j) = \left[vec\left(log\left(d_1^{(i,j)}\right)\right), vec\left(log\left(d_2^{(i,j)}\right)\right), \ldots, vec\left(log\left(d_M^{(i,j)}\right)\right)\right] \in \mathfrak{se}(3) \times \mathfrak{se}(3) \times \cdots \times \mathfrak{se}(3). \quad (29)$$

The Lie algebra of $f(t)$ is given by $g(t) = \{g_o(t), g_{mf}(t)\} = \{g_o(t), g_{rg}(t, t-1), g_{rg}(t, 1)\}$. A human action can be represented by the following Lie algebra structure:

$$LA = \{g(t), t \in [1, T]\} \quad (30)$$

where T is the total number of frames in the sequence.

M is the total number of rigid bodies in skeleton t. Given a rigid body i in skeleton t, $vec(log(o_i(t))$ is the Lie algebra's representation of the orientation of the rigid body, which is 18-dimensional vector($i \in [1, M]$). $g_o(t) = \{vec(log(o_1(t)), vec(log(o_2(t)), \ldots, vec(log(o_M(t))\}$ is the Lie algebra's representation of the static feature of skeleton t, which is a $(18 \times M)$-dimensional vector. $vec(log(d_i^{(p,q)})$ is the Lie algebra's representation of the motions of rigid body i between skeletons p and q, which is 12-dimensional vector. $g_{rg}(t, t-1) = \{vec(log(d_1^{(t,t-1)}), vec(log(d_2^{(t,t-1)}), \ldots, vec(log(d_M^{(t,t-1)})\}$ is the Lie algebra's representation of the rigid-body motions between skeletons t and t-1, which is a $(12 \times M)$-dimensional vector. $g_{rg}(t, 1) = \{vec(log(d_1^{(t,1)}), vec(log(d_2^{(t,1)}), \ldots, vec(log(d_M^{(t,1)})\}$ is the Lie algebra's representation of the rigid-body motions between

skeletons t and 1, which is a $(12 \times M)$-dimensional vector. $g_{mf}(t) = \{g_{rg}(t, t-1), g_{rg}(t, 1)\}$ is the Lie algebra's representation of the dynamic feature of skeleton t, which is a $(24M = 12M + 12M)$-dimensional vector. $g(t) = \{g_o(t), g_{mf}(t)\}$ is the Lie algebra's representation of skeleton t, which is a $(42M = 18M + 24M)$-dimensional vector. Hence, a human action can be seen as temporal evolutions of a $42M$-dimensional vector.

3.4. Key-Skeleton-Pattern Mining. In the previous subsections, a skeleton sequence is represented as the Lie algebra structure $LA = \{g(t), t \in [1, T]\}$, where T is the total number of frames in a skeleton sequence. However, a skeleton sequence can include many noisy skeletons, which reduce the accuracy and efficiency of the action recognition. In this subsection, the key-skeleton-patterns are used to eliminate noisy skeletons in a skeleton sequence in order to capture the most informative skeleton sequences.

3.4.1. Formal Definitions. To mine the key-skeleton-patterns, classic k-means is used to quantize all skeletons represented by the Lie algebra to K symbol. Let $S = \{s_1, \ldots, s_K\}$ be a set containing K symbol and $C = \{c_1, \ldots, c_K\}$ be a set of centroid. Then, a skeleton sequence can be represented as a symbol sequence $L = [l_1, \ldots, l_T]$ $(l_i \in S, i \in [1, T])$. Since different skeletons may be quantized as the same symbol, to minimize the effect of quantization errors, each skeleton $g(i)$ in a sequence is represented by probability p_i, which is used to measure the distance between skeleton $g(i) \in LA$ and centroid $c_i \in C$ as follows:

$$p_i = \frac{dist\left(c_i, g\left(i\right)\right)^{-1}}{\sum_{j=1}^{K} dist\left(c_j, g\left(i\right)\right)^{-1}} \quad \left(0 \leqslant p_i \leqslant 1\right), \tag{31}$$

where $c_j \in C$ correspond to symbol $s_j \in S$. Equation (31) shows the distance inversely proportional to p_i. Now, a skeleton sequence also can be represented by a probability sequence $P = [p_1, \ldots, p_T]$.

Definitions. Some terms are defined in this paper as follows (Important notations are in Table 2.).

Definition 1 (pattern). $\alpha = \left(\alpha^{(1)}, \ldots, \alpha^{(m)}\right)$ is a sequence that contains m symbols chosen from the dictionary, i.e., $\alpha^{(i)} \in S$.

Definition 2 (mining sequence). $d = \{P, L, LA\}$ is a mining sequence applied to mine the Key-skeleton sequence. P is a probability sequence, which represents a skeletons sequence. L is the symbol sequence, which corresponds to P. LA is a skeleton sequence represented by the Lie algebraic structure.

Definition 3 (projected dataset). Given a pattern α and a mining sequence dataset D $(d \in D)$ of an action class, the α-projected dataset $D|_\alpha$ is defined by the *set* $\{d | d \in D \wedge \alpha \subseteq d.L\}$.

Definition 4 (support). For a pattern α and a symbol sequence $d.L \in D|_\alpha$ (where d is an element of $D|_\alpha$), let $F(\alpha, d.L)$ be an indicator variable with value 1 if α is a subsequence of the symbol sequence $d.L$, and 0 otherwise. For any pattern α, its support in $D|_\alpha$ is denoted by $SUP(\alpha, D|_\alpha) = \sum_{i=1}^{|D|_\alpha|} F(\alpha, d.L)$

Table 2: Notations involving data mining.

Notation	Description
α	Skeleton pattern
d	Mining sequence
$D\|_\alpha$	Projected dataset
$SUP(\alpha, D\|_\alpha)$	Support of skeleton pattern α
$ES(\alpha, D\|_\alpha)$	Expected support of skeleton pattern α

Definition 5 (expected support). Given a pattern α and a symbol sequence $d.L$ (where d is an element of $D|_\alpha$), let $pos = Find(\alpha, d.L)$ be the positions where the pattern α takes up in $d.L$ and let $d.P(\alpha) = \prod_{j=1}^{|pos|} d.P[pos[j]]$ be the product of the probability. For any pattern α, its expected support in $D|_\alpha$ is denoted by $ES(\alpha, D|_\alpha) = (\sum_{i=1}^{|D|_\alpha|} d.P(\alpha))/|D|_\alpha|$.

Definition 6 (key-skeleton pattern). Given a pattern α and a mining sequence dataset D of an action class, if $SUP(\alpha, D|_\alpha)$ is larger than a threshold τ_{sup} and $ES(\alpha, D|_\alpha)$ is larger than a threshold τ_{prob}, then the pattern α is called a key-skeleton-pattern of that action class. A key-skeleton-pattern of length i is called an i-pattern.

3.4.2. MinP-PrefixSpan Algorithm. In this subsection, a new pattern-growth algorithm, called MinP-PrefixSpan, is proposed to mine the key-skeleton-patterns of an action class by searching over the enormous space of the symbol sequences and probability sequences of the action class. The algorithm is shown in Algorithm 1. In Lines 2-8, the dataset $D|\alpha$ is employed to construct a new projected dataset $D|\alpha e$. In Lines 9-11, if αe is the key-skeleton-pattern, pattern αe is appended to K and symbol table $T|_{\alpha e}$ is constructed by $Trim(T|_\alpha, D|_{\alpha e})$. In Line 12, MinP-PrefixSpan is recursively called to grow the key-skeleton-pattern until all key-sequence patterns are found.

In Line 10, the trim algorithm is used to improve the efficiency of the MinP-PrefixSpan algorithm by eliminating nonkey-skeleton-patterns. Algorithm 2 shows the implementation details of the trim algorithm. The trim algorithm mainly consists of the following two parts:

Triming rules. Given a mining sequence dataset of an action class D and a pattern α (according to Definition 3, $D|_\alpha$ is the α−projected dataset), two rules are proposed to trim nonkey-skeleton-patterns as follows:

(1) If $SUP(\alpha, D|_\alpha) \leq \tau_{sup}$, then pattern α is a non-key-skeleton-pattern.

(2) If $ES(\alpha, D|_\alpha) \leq \tau_{es}$, then pattern α is a non-key-skeleton-pattern.

Pattern growth. Referring to the pattern-growth method of Prefxispan, one symbol e is used to grow key-skeleton-pattern α and check the support and expected support of the pattern αe. A symbol table $T|_\alpha$ is used to store each symbol e in order to reduce the number of new skeleton patterns in each growth step. An important property is found between symbol tables.

Input: α–projected dataset D_α, symbol table $T|_\alpha$, key-skeleton-pattern dataset K
(1) **for** each symbol $e \in T|_\alpha$ **do**
(2) $D|\alpha e \longleftarrow \emptyset$
(3) **for** each $d \in D|_\alpha$ **do**
(4) $pos = Find(\alpha, d.L)$
(5) **if** $e \in d.L[pos[|\alpha|] + 1, \cdots, |d.L|]$ **then**
(6) Append d to $D|\alpha e$
(7) **end if**
(8) **end for**
(9) **if** $SUP(\alpha e, D|_{\alpha e}) \geq \tau_{sup}$ and $ES(\alpha e, D|_{\alpha e}) \geq \tau_{es}$ **then**
(10) $T|\alpha e \longleftarrow Trim(T|_\alpha, D|_{\alpha e})$
(11) Append αe to K
(12) MinP-PrefixSpan($T|_{\alpha e}, D|_{\alpha e}$,$K$)
(13) **end if**
(14) Free $D|_{\alpha e}$ and $T|_{\alpha e}$
(15) **end for**

Algorithm 1: MinP-PrefixSpan($T|_\alpha, D|_\alpha, K$).

Input: symbol table $T|_\alpha$, αe–projected dataset $D_{\alpha e}$
Output: symbol table $T|_{\alpha e}$
(1) $T|_{\alpha e} \longleftarrow \emptyset$
(2) **for** each symbol $m \in T|_\alpha$ **do**
(3) **if** $SUP(\alpha e m, D|_{\alpha e}) \geq \tau_{sup}$ and $ES(\alpha e m, D|_{\alpha e}) \geq \tau_{es}$ **then**
(4) $T|_{\alpha e} \longleftarrow T|_{\alpha e} \cup m$
(5) **end if**
(6) **end for**

Algorithm 2: Trim($T|_\alpha, D|_{\alpha e}$).

Property 7 (symbol table). If a key-skeleton-pattern γ grows from α, then $T|_\gamma \subseteq T|_\alpha$.

Proof. Let us denote $\gamma = \alpha e^{(1)}$ as the key-skeleton-pattern and $e^{(1)} \in T|\alpha$. Suppose $e^{(2)} \in T|_\gamma$ and $\beta = \alpha e^{(1)} e^{(2)}$ also is a key-skeleton-pattern on $D|_\alpha$. Since $SUP(\alpha e^{(2)}, D|_\alpha) \geq SUP(\beta, D|_\alpha)$ according to Definition 4, $SUP(\alpha e^{(2)}, D|_\alpha) \geq \tau_{sup}$. Since $ES(\beta, D|_\alpha) - ES(\alpha e^{(2)}, D|_\alpha) \geq 0$ according to Definition 5, $ES(\alpha e^{(2)}, D|_\alpha) \geq \tau_{es}$. We conclude that $e^{(2)}$ belongs to $T|_\alpha$,which implies that $T|_\gamma \subseteq T|_\alpha$. □

3.5. Discovering the Most Informative Skeleton Sequence. The task of Algorithm 3 is to discover the most informative skeleton sequences for all actions. Let J be a mining sequence dataset of all actions. K is a dataset used to store the key-skeleton-patterns of all action classes, and Q is a dataset used to store the most informative skeleton sequences of all actions. In Lines 2-11, the key-skeleton-patterns of each class action are mined from training dataset U and appended to dataset K. In Lines 12-22, the key-skeleton-patterns in dataset K are employed to discover the most informative skeleton sequence of the actions in dataset J, and all most informative skeleton sequences are stored in dataset Q(refer to Figure 5).

Dynamic Time Warping (DTW)[33] has excellent performance in searching for an **optimal alignment** between time sequences. Therefore, for each action class, our model uses the action standardization algorithm proposed by the author of [5] to compute a nominal action and employs DTW to warp all the training or testing actions into this nominal action. SVMs are extensively used in computer vision to achieve excellent performances in image and video classifications. To achieve better classification results, a linear SVM is used to classify the most informative skeleton sequences.

3.6. Datasets. In this study, three standard 3D human action datasets are employed to study the effectiveness of the proposed method.

MSRAction3D dataset [34] can be captured using a depth camera similar to the Kinect device. This dataset consists of 20 actions of 10 subjects, with each action having two or three repetitions. In total, there are 557 action sequences. The dataset provides 3D locations of 20 joints. The horizontal and vertical locations of each skeleton joint are stored in the screen coordinates, and the skeleton's depth position is stored in the global coordinates. Human actions in this dataset capture various types of motions, which are related to arms, legs, torso, and their combinations. Experiments on this dataset are challenging, but the dataset is widely applied

Input: key-skeleton-pattern dataset K, mining sequence dataset of all actions J
Output: The dataset of the most informative skeleton sequences Q
(1) Obtaining training dataset U from J
(2) $K \longleftarrow \emptyset; Q \longleftarrow \emptyset$
(3) **for** the dataset of each action class $D \subset$ U **do**
(4) T is the table that includes all symbols in D
(5) **for** each symbol $\alpha \in T$ **do**
(6) **if** $SUP(\alpha, D|_\alpha) \geq \tau_{sup}$ and $ES(\alpha, D|_\alpha) \geq \tau_{es}$ **then**
(7) $T|_\alpha \longleftarrow Trim(T, D|_\alpha)$
(8) append α to K
(9) MinP-PrefixSpan$(T|_\alpha, D|_\alpha, K)$
(10) **end if**
(11) **end for**
(12) **end for**
(13) **for** each element $d \in J$ **do**
(14) **for** each key-skeleton-pattern $m \in K$ **do**
(15) **if** m is a subsequence of d.L **then**
(16) $patternpos = zero(len(d.L), 1)$
(17) $position = Discoverlocation(m, d.L)$
(18) **for** $i \in position$ **do**
(19) $patternpos[i] = 1$
(20) **end for**
(21) **end if**
(22) **end for**
(23) **for** z=1 to Len(d.L) **do**
(24) **if** patternpos[i]==1 **then**
(25) append d.LA[z] to s
(26) **end if**
(27) **end for**
(28) Append s to Q
(29) **end for**
(30) return Q

Algorithm 3: Informative skeleton sequence (K, J).

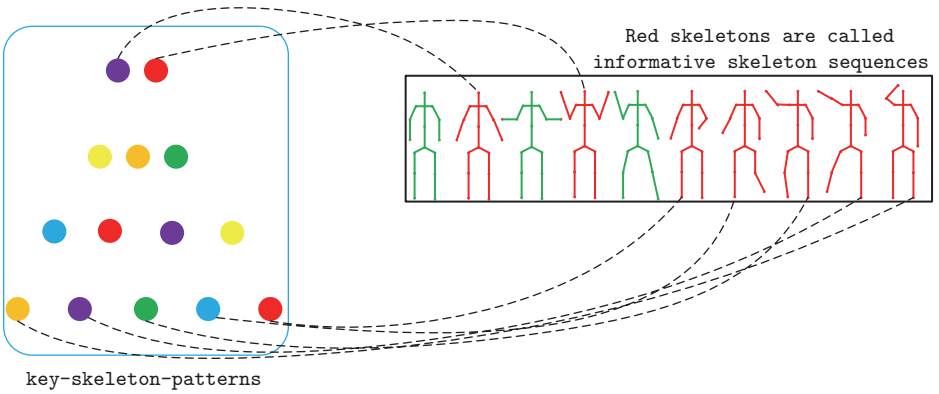

Figure 5: Discovering most informative skeleton sequence.

to test the accuracy and robustness of recognition methods for various actions.

UCKinect-Action dataset [19] is captured using a stationary Kinect sensor. It consists of 10 human actions obtained from daily life: walking, sitting down, standing up, picking up, carrying, throwing, pushing, pulling, waving, and clapping hands. Each human action is performed by 10 different subjects (nine males and one female) twice or thrice. In total, there are 199 action sequences. This dataset is very challenging. First, for some action sequences, parts of the human body are invisible because the body parts are out of the field of view. Second, subjects performed the same action using different limbs, such as waving the left hand and waving the right hand. Third, it is very difficult to capture the action sequences with invariance to the view point.

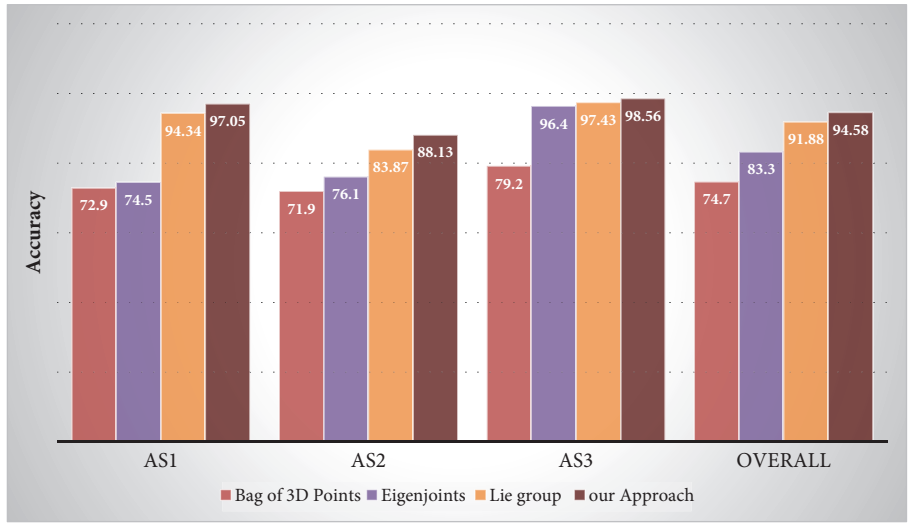

Figure 6: Recognition rate on the MSRAction3D dataset based on AS1, AS2, and AS3.

G3D dataset[23] consists of 663 sequences of 20 gaming actions captured by Kinect. Each actor performed each gaming action more than two twice. Although the dataset can provide synchronized video, depth, and skeleton data, skeleton data is only chosen in our experiment. The dataset is challenging because of the following two aspects: (1) if the body parts are occluded, the Kinect device gives inferred results, which may reduce the accuracy of the action recognition. (2) if two different actions have very small interclass variations, the two actions may easily interference with each other in the action recognition.

4. Experimental Results

The skeleton preprocess is as follows: a human action is composed of a continuous evolution of a series of skeletons. To make each skeleton view-invariant, all 3D joint coordinates in the skeleton are transformed to the coordinate system, which places the hip center at the origin. The entire skeleton will stop rotation until the global x-axis is aligned with the ground plane projection of the vector from the left hip to the right hip (refer to Figure 2(b)).

4.1. Experiments on the MSRAction3D Dataset. Following the experimental protocol of [4], 20 actions in the MSRAction3D dataset are divided into three subsets $AS1$, $AS2$, and $AS3$, each including eight actions. AS1 and AS2 include actions with similar movement. AS3 groups include more complex actions. A half of the subjects are chosen for training, and the remaining subjects for testing. The experiment is run on ten different combinations of training and testing sets, and the mean performance is reported. Figure 6 shows that our approach outperforms various other representations. Our approach achieves a mean accuracy of 94.58% on the MSRAction3D dataset, outperforming other action recognition approaches, including Bag of 3D Points [4], Eigenjoints [29], and Lie group [5], which achieved accuracies of 74.7, 83.3, and 91.88%, respectively. Our approach performs better than the others both in distinguishing similar actions and in recognizing complex actions. This is mainly because the informative skeleton sequences, represented by the Lie group, are used to train SVM classifiers.

Following the experimental protocol of [16], the dataset containing all actions is tested. The experimental setting is more challenging than that of [4]. Our approach achieves an accuracy of 97.4%, outperforming other relative action recognition approaches, including Grassmann manifold [16], graph-based model [20], and Lie group [5], which achieved accuracies of 91.21, 92.2, and 92.46%, respectively, as shown in Table 3.

Figure 7 shows the classification confusion matrix on the whole MSR-Action3D dataset. Most actions on the dataset can be correctly recognized by our approach, but classification errors occurred if two actions were extremely similar, such as *draw tick* and *draw X*.

Matlab is used to run the experiments on a 3.60GHz Intel Core i7-4790 CPU machine. The average testing time of one action sequence in the dataset only costs 35.1ms, which is lower than that of Lie group(72.5ms).The reason is that the skeleton feature dimension of our approach(798-dimension) is lower than that of Lie Group(2052-dimension). However, since the authors of Grassmann Manifold and Graph-based model do not open the source code of their approaches, the average testing time of their approaches cannot be obtained.

4.2. Experiments on the UCKinect-Action Dataset. The recognition rate of our approach on the UCKinect Dataset is 98.87%. Our method outperforms the Lie group[5], Grassmann manifold[16], Eigenjoints[29], and learning feature combination[30], which achieved recognition rates of 97.08, 97.91,97.1, and 98.00%, respectively, as shown in Table 4.

The average testing time of one action sequence in the dataset costs 33.6ms, which is lower than that of Lie group (69.2ms) but higher than that of Learning features combination (13.7ms). The reason is that the skeleton feature dimension of our approach (798-dimension) is lower than

Table 3: Recognition rate on the MSRAction3D dataset based on the experimental protocol of [16].

Approach	Grassmann manifold[16]	Graph-based model[20]	Lie Group[5]	Our approach
Accuracy	91.21	92.2	92.46	97.4

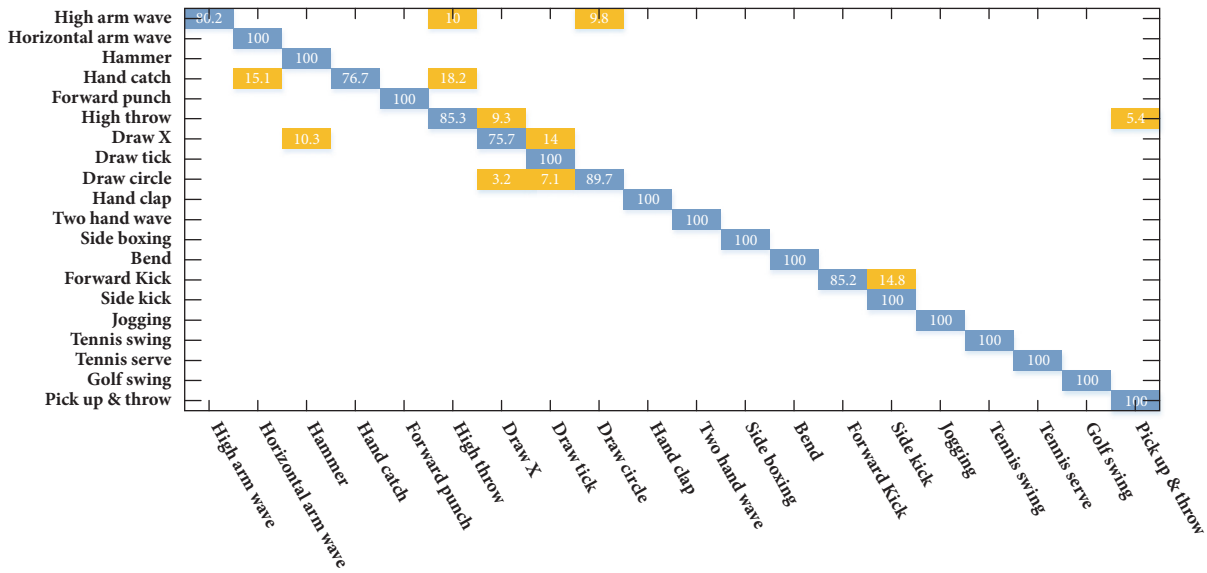

Figure 7: The confusion matrix for action classification in entire the MSR-Action3D dataset.

Table 4: Recognition rate on the UCKinect-Action dataset.

Approach	Recognition accuracy
Lie group[5]	97.08
Grassmann manifold[16]	97.91
Eigenjoints[29]	97.1
Learning feature combination[30]	98.00
Our approach	98.87

Table 5: Recognition rate on the Florence3D-Action dataset.

Approach	Recognition accuracy
SO(3)[14]	87.95
Lie group[5]	91.09
GB-RBM+HMM[31]	86.40
LieNet[22]	89.10
tLDS[32]	90.60
Our approach	95.01

that of Lie Group (2052-dimension) but higher than that Learning features combination (256-dimension). Unfortunately, the average testing time of Grassmann Manifold and Eigenjoints cannot be obtained without the source code of the approaches.

4.3. Experiments on the G3D-Gaming Dataset. The cross-subject test setting, in which half of subjects were used for training and the remaining subjects were used for testing, is used to perform recognition on the data. Table 5 compares our approach with other approaches on the dataset (GB-RBM+HMM [21] and LieNet [17] used a deep-learning method to recognize human action). Our approach achieves a higher recognition rate.

The average testing time of one action sequence in the dataset only costs 34.8ms, which is lower than that of Lie group(71.2ms) and $SO(3)$(58.9ms). The reason is that the skeleton feature dimension of our approach (798-dimension) is lower than that of Lie Group(2052-dimension) and $SO(3)$(1026-dimension). However, the authors of tLDS do not open the source code of their approach, the average testing time of their approach cannot be obtained. Since the deep learning-based approaches usually use GPU to accelerate their models while the nondeep learning-based approaches usually use CPU to perform their experiments, it is hard to implement a fair comparison between these two classes of approaches (our approach belongs to the nondeep learning-based approaches, and LieNet and GB-RBM+HMM belong to the deep learning-based approaches).

5. Conclusion and Future Work

This paper proposes a new skeleton-based action representation, which consists of static and dynamic features. First, the orientation of a rigid body is regarded as six rotation matrices, and each rotation matrix is represented as a point in $SO(3)$. The rigid-body orientations in a skeleton are used to construct the static feature in order to avoid dealing with skeletal scale variations. Second, the motions of rigid bodies, represented by $SE(3)$, are used to construct the dynamic features in order to capture the temporal information of

the skeleton. Finally, based on the proposed representation, the key-skeleton-patterns are employed to discover the most informative skeleton sequences. The experiment results show that our approach achieves better performance than other state-of-the-art skeleton-based action recognition approaches. Further research should combine the Lie group with a linear dynamical system to model human actions as a tensor time series.

Data Availability

In this article, we performed our experiments in the three public Datasets as follows. (1)MSR-Action3D Dataset is an action Dataset of depth sequences captured by a depth camera. The Dataset can be found in http://research.microsoft.com/en-us/um/people/zliu/actionrecorsrc/. (2)UTKinect-Action3D Dataset was collected as part of research work on action recognition from depth sequences. The Dataset can be found in http://cvrc.ece.utexas.edu/KinectDatasets/HOJ3D.html. (3) G3D: Dataset contained synchronised video, depth and skeleton data. The Dataset can be found in http://dipersec.king.ac.uk/G3D/ or search the three action recognition dataset https://github.com/liguang1980/Action-recognition-Datasets.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant no. 61571345, the Fundamental Research Funds for the Central Universities under Grant no. K5051203005, the National Natural Science Foundation of China under Grant no. 91538101, the National Natural Science Foundation of China under Grant no. 61850410523, Huawei Innovation Research Program under Grant no. 2017050310, the Fundamental Research Funds for Xidian University no. XJS18041, and the Natural Science Foundation of the Anhui Higher Education Institutions of China no. KJ2017A376.

References

[1] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: a review," *ACM Computing Surveys*, vol. 43, no. 3, article 16, 2011.

[2] K. M. Knutzen, "Kinematics of human motion," *American Journal of Human Biology*, vol. 10, no. 6, pp. 808-809, 1998.

[3] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 90–126, 2006.

[4] W. Li, Z. Zhang, and Z. Liu, "Action recognition based on a bag of 3D points," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '10)*, pp. 9–14, San Francisco, Calif, USA, June 2010.

[5] R. Vemulapalli, F. Arrate, and R. Chellappa, "Human action recognition by representing 3D skeletons as points in a lie group," in *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 588–595, Columbus, Ohio, USA, June 2014.

[6] C. Wang, Y. Wang, and A. L. Yuille, "An approach to pose-based action recognition," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 915–922, Portland, Ore, USA, June 2013.

[7] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 1290–1297, Providence, RI, USA, June 2012.

[8] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception & Psychophysics*, vol. 14, no. 2, pp. 201–211, 1973.

[9] W. Yang, Y. Wang, and G. Mori, "Recognizing human actions from still images with latent poses," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2030–2037, IEEE, June 2010.

[10] R. M. Murray and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*, CRC Press, 1994.

[11] W. M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975.

[12] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.

[13] W. Ding, K. Liu, X. Fu, and F. Cheng, "Profile HMMs for skeleton-based human action recognition," *Signal Processing: Image Communication*, vol. 42, pp. 109–119, 2016.

[14] R. Vemulapalli and R. Chellappa, "Rolling rotations for recognizing human actions from 3D skeletal data," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, pp. 4471–4479, USA, July 2016.

[15] J. Pei, J. Han, B. Mortazavi-Asl et al., "PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth," in *Proceedings of the 17th International Conference on Data Engineering*, pp. 215–224, April 2001.

[16] R. Slama, H. Wannous, M. Daoudi, and A. Srivastava, "Accurate 3D action recognition using learning on the Grassmann manifold," *Pattern Recognition*, vol. 48, no. 2, pp. 556–567, 2015.

[17] J. Liu, N. Akhtar, and A. Mian, *Skepxels: Spatio-Temporal Image Representation of Human Skeleton Joints for Action Recognition*, 2017.

[18] R. Chaudhry, F. Ofli, G. Kurillo, R. Bajcsy, and R. Vidal, "Bio-inspired dynamic 3D discriminative skeletal features for human action recognition," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2013*, pp. 471–478, USA, June 2013.

[19] L. Xia, C.-C. Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3D joints," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '12)*, pp. 20–27, Providence, RI, USA, June 2012.

[20] M. Li and H. Leung, "Graph-based approach for 3D human skeletal action recognition," *Pattern Recognition Letters*, vol. 87, pp. 195–202, 2017.

[21] G. Evangelidis, G. Singh, and R. Horaud, "Skeletal quads: Human action recognition using joint quadruples," in *Proceedings of the 22nd International Conference on Pattern Recognition, ICPR 2014*, pp. 4513–4518, Sweden, August 2014.

[22] Z. Huang, C. Wan, T. Probst, and L. V. Gool, "Deep Learning on Lie Groups for Skeleton-Based Action Recognition," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1243–1252, Honolulu, HI, July 2017.

[23] V. Bloom, D. Makris, and V. Argyriou, "G3D: A gaming action dataset and real time action recognition evaluation framework," in *Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2012*, pp. 7–12, June 2012.

[24] M. J. Zaki, "SPADE: an efficient algorithm for mining frequent sequences," *Machine Learning*, vol. 42, no. 1-2, pp. 31–60, 2001.

[25] J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M.-C. Hsu, "FreeSpan: Frequent pattern-projected sequential pattern mining," in *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery in Databases*, pp. 355–359, Boston, MA, USA, August 2000.

[26] R. Srikant and R. Agrawal, "Mining sequential patterns: generalizations and performance improvements," in *Advances in Database Technology—EDBT '96*, vol. 1057 of *Lecture Notes in Computer Science*, pp. 1–17, Springer, Berlin, Germany, 1996.

[27] M. Muzammal and R. Raman, "Mining sequential patterns from probabilistic databases," *Knowledge and Information Systems*, vol. 44, no. 2, pp. 325–358, 2015.

[28] J. Yang, W. Wang, P. S. Yu, and J. Han, "Mining long sequential patterns in a noisy environment," in *Proceedings of the ACM SIGMOD 2002 Proceedings of the ACM SIGMOD International Conference on Managment of Data*, pp. 406–417, USA, June 2002.

[29] X. Yang and Y. Tian, "Effective 3D action recognition using Eigen Joints," *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 2–11, 2014.

[30] D. Carbonera Luvizon, H. Tabia, and D. Picard, "Learning features combination for human action recognition from skeleton sequences," *Pattern Recognition Letters*, vol. 99, pp. 13–20, 2017.

[31] S. Nie and Q. Ji, "Capturing global and local dynamics for human action recognition," in *Proceedings of the 22nd International Conference on Pattern Recognition, ICPR 2014*, pp. 1946–1951, Sweden, August 2014.

[32] W. Ding, K. Liu, E. Belyaev, and F. Cheng, "Tensor-based linear dynamical systems for action recognition from 3D skeletons," *Pattern Recognition*, pp. 75–86, 2017.

[33] F. Moerchen, "Algorithms for time series knowledge mining," in *Proceedings of the the 12th ACM SIGKDD international conference*, p. 668, Philadelphia, PA, USA, August 2006.

[34] J. K. Aggarwal and L. Xia, "Human activity recognition from 3D data: a review," *Pattern Recognition Letters*, vol. 48, pp. 70–80, 2014.