Signal, Image and Video Processing (2024) 18:1343–1354
https://doi.org/10.1007/s11760-023-02829-z

ORIGINAL PAPER

Computer vision-based approach for skeleton-based action recognition, SAHC

M. Shujah Islam[1]

Received: 30 March 2023 / Revised: 22 September 2023 / Accepted: 4 October 2023 / Published online: 11 November 2023

Abstract
Given their small size and low weight, skeleton sequences are a great option for joint-based action detection. Recent skeleton-based action recognition techniques use feature extraction from 3D joint coordinates as per spatial–temporal signals, fusing these exemplifications in a motion context to improve identification accuracy. High accuracy has been achieved with the use of first- and second-order characteristics, such as spatial, angular, and hough representations. In contrast to the and hough transform, which are useful for encoding summarized independent joint coordinates motion, the spatial, and angular features all higher-order representations are discussed in this article for encoding the static and velocity domains of 3D joints. When used to represent relative motion between body parts in the human body, the encoding is effective and remains constant across a wide range of individual body sizes. However, many models still become confused when presented with activities that have a similar trajectory. Suggest addressing these problems by integrating spatial, angular, and hough encoding as relevant order elements into contemporary systems to more accurately reflect the interdependencies between components. By combining these widely-used spatial–temporal characteristics into a single framework SAHC, acquired state-of-the-art performance on four different benchmark datasets with fewer parameters and less batch processing.

Keywords Computer vision · Machine learning · Skeleton-based action recognition · Human action recognition · Artificial intelligence

1 Introduction

Human action recognition is more resilient to background information and simpler to process, gaining a growing number of scientific interest due to its many applications in the domains of health care [1], virtual reality [2], innovation technology [3], and defense security [4], etc. Recent advances in Skeleton-based action recognition have contributed to an increase in the accuracy of human action recognition. By utilizing motion feature detection networks, action recognizers extract topological information from skeletal sequences with greater precision. To apply motion features to skeleton-based action identification, skeletons are regarded as graphs, with each vertex representing a body joint and each edge representing a bone. Initially, only first-order features were used to express the joint coordinates [5]. Subsequently, [6] proposed a second-order characteristic: each bone is described as the vector difference between the coordinate of one joint and that of its closest neighbor in the direction of the body's center. Experiments indicate that these second-order properties enhance the accuracy of skeleton-based action recognizers.

Existing approaches, however, have a difficult time differentiating between activities that have very similar motion trajectories. Since the joint coordinates in each frame are comparable in these motions, determining the source of differences in coordinates may be difficult. Differences in body size, movement speed, or the nature of the movements being performed may all have a role. This research suggests using higher-order representations in the form of angles to accurately record the relative motions between body components while remaining invariant over a wide range of human body sizes. This new capability, which is called angular encoding, is intended to be used in the static and velocity domains of human body joints. For this reason, the suggested encoding improves the model's action recognition accuracy. Adding angular information to state-of-the-art

✉ M. Shujah Islam
msameem@kfu.edu.sa

[1] College of Computer Sciences and Information Technology, King Faisal University, Al Ahsa, Saudi Arabia

action recognition architectures like the SpatioTemporal Network [5] and Decoupling [7] improves the classification of complex action sequences, even when the actions have identical motion trajectories, as shown in experimental data. A wide variety of approaches, including skeletons, skeleton joints, deep learning, and silhouette frames, have been proposed in the literature to solve human action recognition. There are advantages and disadvantages to every available human action recognition method. For instance, techniques based on the skeleton and the skeleton joint's points [8] allow for the elimination of extraneous data such as clothing texture, lighting, and backdrop while simultaneously having to deal with less data extraction. Further, the computational expense and financial commitment involved in implementing deep learning algorithms [9] are significant. In contrast, silhouette images provide a challenging environment for object detection [10]. Keeping these drawbacks of state-of-the-art methods in mind, presented an action descriptor called SAHC that is based on skeletal joint locations and forms connections between joints and frames. Since multiple joints are coupled to one another and frames rely on one another when completing any action sequence, both linkages are useful for action recognition. Previously, joints recognized using straight-line inter of spatial, angular, and hough features SAH were used to extract spatial information of features, whereas connection-frames now provide temporal information for the suggested descriptor. Based on the connections between the joint and the frame, the proposed system extracts the skeletal joint-based action recognition features to the representative frame. As they suggest a hierarchical method using 3D skeletal joints, they then use spatial, angular, and hough features to calculate the spatial and temporal details to form a skeletal modeling method. The final SAHC feature vector, a classifier, is used for labeling the human activities after the combination of spatial, angular, and Hough features is taken into consideration for feature selection and maintaining the previous data. Four publicly accessible and frequently utilized action datasets [11–14] are used to assess the proposed SAHC. The proposed work contributes to the field by making SAHC a more effective approach than its predecessors by collecting both the spatial and temporal information between connected joints. Summarize proposed contributions as follows:

- The spatial, angular, and hough features, all of which are higher-order representations, are discussed in this article for encoding the static and velocity domains of joints. When applied to the human body, the encoding successfully represents relative motion between body components while being invariant over a wide range of individual body sizes.
- Integrating the joints and frame connections into preexisting action recognition systems is a straightforward way to further improve performance. The results demonstrate that these associations provide valuable further data to the already extant elements, such as the joint representations.
- To the best of my knowledge, the proposed descriptor is the first to combine several types of angular characteristics into state-of-the-art spatial–temporal SAHCs and results on a number of benchmarks are among the best available. Meanwhile, the suggested SAHC encoding may provide even a basic model with a significant boost in performance. Therefore, the suggested angular encoding enables edge devices to recognize actions in real time.

2 Related work

Introduce related work to the suggested SAHC action recognition framework in the related work section. Here, separate the relevant research into two categories: Skeleton-based Action Recognition and cutting-edge action recognition techniques.

Many of the earlier efforts at skeleton-based action detection recorded all joint coordinates of the human body in each frame as a feature vector for pattern learning [15]. These models seldom included the interdependencies between bodily joints, resulting in a dearth of action-related data. There have also been kernel-based algorithms developed for action recognition [16]. Subsequently, when deep learning became the norm for video processing [17] and comprehension [18], RGB-based videos began to address action identification. Nevertheless, they have difficulties in domain adaptation [19] due to their diverse topic backgrounds and textures. Conversely, skeletal data have substantially fewer domain adaption difficulties. The use of convolutional neural networks to skeleton-based action recognition resulted in an improvement [20]. CNN, however, are built for grid-based data and cannot utilize the topology of a graph, making them unsuitable for graph data. Recently, there has been an increase in interest in deep graph neural networks [21]. Graph neural networks have also begun to garner interest in skeletal recognition. A skeleton is represented as a graph in GCN-based models, with joints as nodes and bones as edges. Methodology proposed in [5] was an early application that used graph convolution to spatially aggregate joint data and to convolve successive frames along the temporal axis. Consequently, an algorithm proposed in [22] was developed to further enhance the spatial feature aggregation utilizing a learnable adjacency matrix in its place of a fixed graph skeleton. Methodology proposed in [23] acquired long-range temporal relationships using LSTM as its backbone, then altered every gate operation from the original fully connected layer to a graph convolution layer, therefore maximizing the use of the skeleton's topological knowledge. By applying a learnable residual mask to the adjacency matrix of the graph

convolution, as suggested in [6], the skeleton's topology was made more malleable, and a second-order feature, the difference in coordinates between adjacent joints, was proposed to serve as the bone evidence. A combination of two models, trained using joint and bone characteristics, significantly increased the cataloging accuracy. In skeleton-based action recognition, several graph convolution approaches, such as SGN [24] with self-attention and shift convolution, have been suggested. A technique presented in [25], recently obtained better results by introducing graph 3D convolutions to aggregate characteristics inside a window of successive frames. Nevertheless, 3D convolutions need a lengthy running time.

Although there are numerous works in downstream video understanding tasks based on cutting-edge action recognition techniques. Recent authors have presented self-attention models that optimize the graph structure dynamically [26]. The authors built a CNN architecture that captures topological information more effectively [27]. The authors investigate the skeleton-based action recognition one-shot issue [28]. In a collection of activity reference samples, they relate the metric learning scenery and transfer the issue to a nearest-neighbor search. The adversarial assault issue in skeleton-based action recognition was scrutinized [29]. They examined a perceptual deficit that renders an assault imperceptible. The black-box assault on skeleton-based action recognition was examined by the authors [30]. They presented an attack technique and demonstrated that adversarial attacks are a concern and that adversarial samples on manifolds are prevalent for skeletal movements. All present approaches suffer from a lack of discrimination accuracy between actions with comparable motion trajectories. This drives to pursue a novel encoding to enable the model to distinguish between two seemingly identical events. Nearly all paintings exhibit angle characteristics comparable to the local characteristics given in this article [31]. In contrast, as provided an assortment of angular encoding forms. Each category has more subcategories. Diverse types of angular encrypting are intended to represent motion characteristics of diverse kinematic body sections. Approaches such as motion-driven spatial and temporal adaptive using graph convolutional networks [32], multi-level spatial–temporal transformer for group activity recognition [33], single and two-person action recognition based on R-WAA [34], learning instance-level spatial–temporal patterns for person re-identification [35], and anomaly detection via motion exemplar guidance techniques are utilized for the temporal modeling of the human body for the action recognition [36].

3 Proposed methodology

The skeleton of the human body is represented as a collection of joints' 3D spatial coordinates. Thus, a unique SAHC can be used to represent the spatial relationship between skeletal joints. Adding joint node properties with $V = \{j_i\}$ i $= 1$ to the total number of skeletal joints, for instance, may maintain the original coordinate information. By way of dividing each i-th joint in the body into its three components, $j_i(x_i y_i z_i)$. By way of creating an action feature matrix that is then utilized to extract features. Equally, define the signal of each joint for bone data along with its 3D spatial coordinates. Briefly go over some key methods employed in the proposed SAHC action descriptor in this part. First, the data preprocessing module is described and the suggested features are formulated. The intended action description is finally constructed as SAHC. By employing skeleton joint points and joint positions in the proposed method for human action recognition.

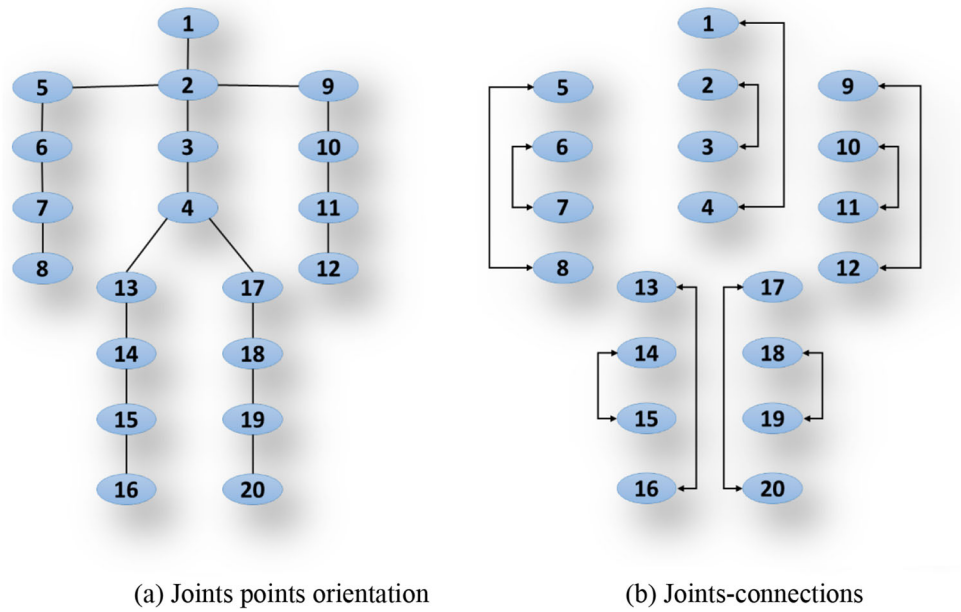
Table 1 Ten distinct connections are being made between the joints, and they are crucial in order to extract all necessary characteristics between the joints

Joints-connections	Among joints connections
c_1	Head joint and hip center joint
c_2	Neck joint and spine joint
c_3	The right shoulder joint and right hand joint
c_4	Right elbow joint and right wrist joint
c_5	The left shoulder joint and left hand joint
c_6	Left elbow joint and left wrist joint
c_7	Right hip joint and right foot joint
c_8	Right knee joint and right ankle joint
c_9	Left hip joint and left foot joint
c_{10}	Left knee joint and left ankle joint

The purpose of joints-connections is to link together the joints' spatial velocities in a single representation. Joints-connections help the model's procedures extract crucial motion information

3.1 Joints-connections

A joint is the connection between two bones in the skeleton, while an interrelationship is the connection between and influence among several joints. When a human being performs a series of movements, the body's many joints all work together. Research shows that the magnitude of a joint's moment is related to its degree. By establishing connections between the head, arms, legs, and neck, for instance, joints-connections may be produced. Connections between the torso and limbs, including the hip and the ankle. Each joints-connection consists of two joints: the first, denoted by the letters $p_i(x_p y_p z_p)$ and $q_i(x_q y_q z_q)$. Let's look at Table 1, as shown that c_i $(i = 1$ to $10)$ is the total number of joints-connections. Joints-connections is a tool for acquiring this spatial information; the connections between the various datasets are shown in Fig. 1. Even by using these simple ten



(a) Joints points orientation (b) Joints-connections

Fig. 1 The proposed joints' point orientation and detected joints-connections are represented. Joint connection is very important in extracting the proposed features of SAHC

connection, the action recognition is enhanced a lot even by using simple features.

Datasets [11, 12] demonstrate joint connections using 20 joint points, whereas dataset [13] uses 15 joints. In contrast, the dataset described in [14] has 30 joints split evenly between two people. Using joints, such as have computed the aforementioned connections, with the overall orientation of the joints' points shown in Fig. 1a and the created connections shown in Fig. 1b.

3.2 Spatial features

First extract spatial information in order to use cutting-edge SAHC features. Because they indicate where the joints are physically located, spatial characteristics are crucial for encoding the results of boundary analysis on joints [37]. The distance between two real-valued vectors is defined by the Euclidean distance. If the data rows include numbers (floating point or integer values), you will probably utilize Euclidean distance to determine how far apart they are. After a Joints-connections are made in Table 1, the distance between any two places in the joint may be determined. Using the 3D distance relation in Eq. 1, can determine the spatial characteristics (μ_f) between any i^{th} joints $p_i(x_p y_p z_p)$ and $q_i(x_q y_q z_q)$ by using c_i. The distance feature is crucial in developing the action descriptor since it reveals the separation between the joints. By concatenating joints-connections, a skeletal joint distance descriptor that describes relationships between distances of skeletal joints, spatial features are extracted. Ten different joint connections are tested, which aids in the extraction of features and the selection of input joints and bones.

$$\mu_f = \sqrt{\left((x_q - x_p)^2 + (y_q - y_p)^2 + (z_q - z_p)^2\right)} \tag{1}$$

3.3 Angular features

As input joint data, now sent the motion joint's angular data into the recognition algorithm [38]. One way to depict movement is by the use of "joint motion". All of an actor's trajectories may be seen plotted out on the 3D coordinates. It is possible to make changes to the uploaded trajectory and identify the distinctive features of the movements. 3D coordinates allow for the interpretation of the trajectory at p and q angles. Those angular measurements served as the basis for the recognition algorithm. By solving Eq. 5, can convert the input angle data into feature vectors that include both spatial and temporal information. Now by using connection joints, it may use them to extract angle characteristics. Angle features for SAHC are first computed using the cosine angle (θ_a), as indicated in Eq. 1. Using Table 1 c_i, may depict the relationship between two i^{th} joints by labeling their respective joint points $p_i(x_p y_p z_p)$ and $q_i(x_q y_q z_q)$, with the angle feature between them denoted by θ_a. The angle feature is computed so as to reveal the precise angular relationship between the two joints in a three-dimensional space. As shown by Eq. 2, this is a crucial metric that has to be computed. In order to learn an effective representation of complicated actions, a set of body joint connections are connected to calculate effective angular characteristics. Due to the 3D representation, the system may learn specific information that distinguishes

between nearby joints' angular values that correspond to various body part pairs.

$$\theta_a = \cos^{-1}\left(\frac{\left(x_p x_q + y_p y_q + y_p y_q\right)}{\sqrt{\left(\left(x_p^2 + y_p^2 + z_p^2\right)\left(x_q^2 + y_q^2 + z_q^2\right)\right)}}\right) \tag{2}$$

3.4 Hough features

The Hough transform can be used to detect lines, circles or other parametric curves [39]. The Hough transform is a feature extraction technique used in image analysis, computer vision, and digital image processing. The Hough transform is a technique that can be used to isolate features of a particular shape within an image. Joint's sine relation feature is the most important part of the proposed SAHC descriptor. In this part of the feature vector, as have two i^{th} joints p_i and q_i taken from Table 1, are individually transformed using the joint's sine relations. Each coordinate of joint $p_i(x_p y_p z_p)$and$q_i(x_q y_q z_q)$ is represented by a difference of $\frac{\pi}{2}$ using calculated θ_f, by means of c_i. The reason behind using a difference of $\frac{\pi}{2}$ is that each coordinate is placed in a particular quadrant. Hough features are used to represent joint p_i by ε_p, similarly, joints q_i by ε_q using Eq. 3 and Eq. 4, respectively. The appearance and patterns of the motions in video sequences serve as the basis for human action labeling in action recognition systems, but the majority of current research and most conventional methodologies either ignore or are unable to use the individual monitor of each joint motion that is placed in a separate phase difference. In a sine angle relation, a mapping between densely sampled feature patches and the votes assigned to them is learned using Hough features. The suggested system performs skeletal joints-based action recognition more successfully by utilizing low-level features like Hough features.

$$\begin{aligned}\varepsilon_p =& x_p \sin\left(\pi/4 - \theta_a\right) + y_p \sin\left(\pi/2 - \theta_a\right) \\ &+ z_p \sin(\pi - \theta_a)\end{aligned} \tag{3}$$

$$\begin{aligned}\varepsilon_q =& x_q \sin\left(\pi/4 - \theta_a\right) + y_q \sin\left(\pi/2 - \theta_a\right) \\ &+ z_q \sin(\pi - \theta_a)\end{aligned} \tag{4}$$

3.5 Frames-connections

Unlike joints-connections, which discussed the interrelationships between joints, frames-connections analyzed the interrelationships between the frames. Frames-connections capture the temporal information of a frame. It is constructed using the relationships between three frames: the specific current frame ($\boldsymbol{s_c}$), the initial frame ($\boldsymbol{s_1}$), and the subsequent frame ($\boldsymbol{s_{c+1}}$). Frames-connections include three connections: 1) First frame connection is present in $\boldsymbol{s_c}$, as are $\boldsymbol{p_i}$ and $\boldsymbol{q_i}$. The second frame connection is located between $\boldsymbol{s_c}$ and $\boldsymbol{s_1}$; $\boldsymbol{p_i}$ resides in $\boldsymbol{s_c}$ while $\boldsymbol{q_i}$ resides in $\boldsymbol{s_1}$. The third frame connection is the relationship between $\boldsymbol{s_c}$ and $\boldsymbol{s_{c+1}}$, where $\boldsymbol{p_i}$ resides in $\boldsymbol{s_c}$ and $\boldsymbol{q_i}$ resides in $\boldsymbol{s_{c+1}}$. In the event that $\boldsymbol{s_c}$ is the final frame, the succeeding frame becomes the first frame, so $\boldsymbol{s_{c+1}} = \boldsymbol{s_1}$. The first frames-connection is used to calculate object deviation relative to the first frame; the first frame is essential for descriptor generation [9]. Additionally, emphasize the significance of utilizing first frames-connection and third frames-connection when performing skeletal joints-based action recognition tasks. Modeling temporal data seek to identify methods of symbolically describing time-based situations for eventual computer depiction and replication. By linking together successive frames of a 3D skeleton, density rendering may include and make sense of temporal information. Multiple sequences may be used to build up a comprehensive picture of a video's temporal dynamics, allowing for things like efficient scalar feature field recognition. The time spent on making flow animations of motion joint fields may be drastically cut down by using the proposed frames-connections, which also provide the added benefit of amassing the efficient characteristics discovered. Without increasing their computational cost, frames-connections solve the issue of capturing temporal information for video classification in 3D networks. Existing approaches concentrate on transforming the architecture of 3D networks by using filters in the temporal dimension, or by using optical flow, etc., which raises the cost of computation. Instead, in order to capture quick frame-to-frame changes, the proposed skeletal-based action recognition system SAHC suggests a novel sampling strategy that involves rearranging the channels of the input video. The proposed sampling strategy enhances performance on numerous SAH-based architectures, as seen without any ornamentation (Fig. 2).

3.6 Action feature, SAHC

As per established connections between joints and between frames, and have tested the connectivity between joints in various frames. The feature vector may then be constructed using this accumulated data. Dimensions of space and angles between joints are calculated, revealing the relationship between two joints (μ_f). For Hough features, p_i and q_i are used to separately represent the two associated joints, while θ_a is used to represent the whole group. As propose a new action descriptor, which is called SAHC, that takes into account the unique characteristics of each human and their environment to offer both temporal and geographical details about the skeleton's joint points. The correlation between SAHC is seen in Eq. 2. Each frames-connections

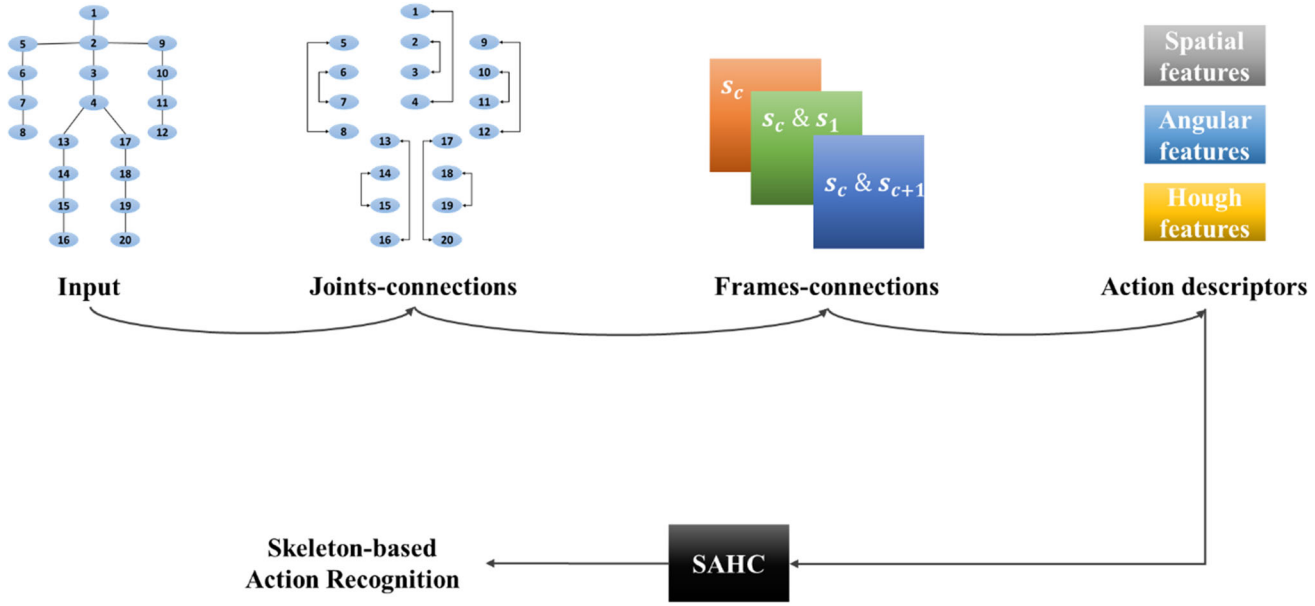


Fig. 2 Three spatial–temporal feature detection blocks, including a spatial multiscale, an angular multiscale, and a hough multiscale staio-temporal features unit, make up the backbone of proposed system. The suggested method makes use of the information contained in 3D structural skeletons, namely the extracted joints and frame connections. SAHC, a multiscale unit-based skeletal joint descriptor, is suggested and correlated with four functional categories. For more explanation of the suggested methodology, please refer to Sect. 3

Table 2 Proposed action descriptor, SAHC algorithm

Algorithm 1: SAHC Algorithm
Use 3D skeleton joint's
for frame = 1 to n do
Using current frame (s_c)
Spatial features
Angular features
Hough features
Using current and initial frame (s_c & s_1)
Spatial features
Angular features
Hough features
Using current and subsequent frame (s_c & s_{f+1})
Spatial features
Angular features
Hough features
SAHC Features
end for
Apply a classification
Return the best solution

SAHC is computed using joints-connections (Table 1), as seen in the system diagram (Fig. 5). In the instance of the provided descriptor, the SAHC characteristics are computed by drawing on the first, second, and third frames-connections. Table 2 and Eq. 5 depict the suggested SAHC action descriptor method. Since a single feature-based representation is insufficient to accurately describe the perspective and faction movement of joints, the fusion of numerous characteristics is crucial for understanding actions. As the suggested action descriptor SAHC incorporates spatial, angular, and hough information to compute the spatial and temporal details to construct a skeletal modeling approach, by leveraging the connections between the joint and the frame to extract taking into consideration actions of 3D joints. In this paper, propose a unique descriptor that may accurately characterize human activities and events by using just 3D skeleton joints. By statistically evaluating the motion patterns of the 3D joint locations of the human body, may extract the suggested SAHC descriptor ($\vartheta_{\mathrm{SAHC}}$), a low dimensional vector from each sequence. The suggested descriptor has been optimized for detecting events and activities that include humans. Recognition strategies based on 3D skeleton joints have gained popularity because of the proliferation of inexpensive action descriptors like the SAHC. Human skeletal information is stored as three-dimensional coordinates of landmarks on a person's body. The major benefit of this 3D data format is the size savings it offers over color/gray pictures or depth maps. When dealing with enormous datasets for the purposes of classifier training, this makes a tremendous impact. Although there has been substantial development in human action identification, current algorithms are far from ideal, particularly if any actor is doing activities. Thus, the suggested action descriptor incorporates all necessary data.

Table 3 An overview of spatial, angular, and Hough feature assembling evaluations

Approach	Accuracy (%)
Du et al. [42]	94.49
Chen et al. [43]	94.90
Xu et al. [44]	96.10
Jin et al. [40]	96.50
Luo et al. [41]	97.26
SAHC	99.8

Joints-connections, and Frames-connections, both of which are ensembles, are especially noteworthy since they achieve the best prediction accuracy. The suggested technique is shown to improve performance when compared to other methods, and these methods are all evaluated for their correctness. Analysis of SAHC in relation to the other methods used on the dataset [11]

Table 4 A overview of spatial, angular, and Hough feature assembling evaluations

Approach	Accuracy (%)
McNally et al. [47]	90.0
Islam et al. [48]	91.8
Chikhaoui et al. [49]	92.67
Mengyuan et al. [45]	94.51
Tasnim et al. [46]	95.1
SAHC	96.5

Joints-connections, and Frames-connections, both of which are ensembles, are especially noteworthy since they achieve the best prediction accuracy. The suggested technique is shown to improve performance when compared to other methods, and these methods are all evaluated for their correctness. Analysis of SAHC in relation to the other methods used on the dataset [12]

$$\vartheta_{\text{SAHC}} = \Big[[\mu_f \theta_a \varepsilon_p \varepsilon_q]_{s_f}, [\mu_f \theta_a \varepsilon_p \varepsilon_q]_{s_c \& s_1}, \times [\mu_f \theta_a \varepsilon_p \varepsilon_q]_{s_c \& s_{f+1}} \Big] \Big|_{f=0}^{f=F-1} \tag{5}$$

4 Results and discussion

In the following, by way of evaluate proposed models' accuracy in comparison to the current standard. Here, such as contrast SAHC with state-of-the-art methods and investigate the synergy between the two sets of representations by fusing them together later on, using the SAHC's suggested vectorized features to achieve state-of-the-art performance. Here, while compare the proposed SAHC's performance outcomes in terms of accuracy (%) utilizing Datasets [11–14] to those of other state-of-the-art techniques. Using low-level features known as the SAHC, which are made up of spatial, angular, and hough features, it is shown that SAHC not only achieves state-of-the-art performance on multiple datasets covering a wide range of action-recognition scenarios but also performs real-time processing because the proposed system works with minimal information.

Table 3 shows the top performance on the skeleton joint-based action recognition dataset [11]. Keep in mind that cross-validation was used to choose the model parameters for the test. These factors led to a % accuracy while linearizing a sequence compatibility kernel. There was a 99.8% improvement in accuracy with the recommended action description. With respect to performance, Jin et al. [40] were successful 96.50% of the time. However, Luo et al. [41] achieved somewhat better results, with 97.26% accuracy. Table 3 shows the complementary nature of the proposed technique SAHC, which combines [40, 41] to enhance accuracy by 2.54% above the state-of-the-art on this dataset.

The uppermost performance on the skeletal joint-based action recognition dataset [12] is shown in Table 4. Keep in mind that cross-validation was used to choose the appropriate values for the evaluation model's parameters. The linearization of a sequence compatibility kernel using these settings yielded a certain percentage of correctness. The proposed system achieved a 96.5% success rate with a planned action description. To compare, Mengyuan et al. [45] had a success rate of 94.51%. Tasnim et al. [46] achieved somewhat better results, with an accuracy of 95.1%. Table 4 displays the comparative strengths of the suggested methods SAHC providing a 1.4% accuracy gain above the state-of-the-art on this dataset.

The suggested technique was compared to five other methods using the accuracy index of the dataset [12] in Table 4. According to the data, the lowest accuracy was attained by Chen et al. [11] at 79.1%, while the highest accuracy was achieved by Chikhaoui et al. [25] with 92.67%. However, for dataset [12], suggested SAHC achieved 99.2%, which was an increase of 3.33% above the previous highest efficiency.

Table 5 summarizes the results of a comparison between the suggested SAHC technique and the five currently used approaches to assessing dataset [13] precision. Based on the provided values, the highest accuracy was reached by Papadopoulos et al. [11] with 96.3 & 97.41%, while the lowest accuracy was obtained by Gaglio et al. [9] with 84.8 & 84.5% across all methodologies tested. The efficiency of the suggested SAHC, on the other hand, was 97.6% for the dataset [13], which was a substantial improvement over the prior best. However, for the dataset [13], the recommended SAHC reached 97.6%, an improvement of 0.19% above the maximum efficiency previously reported.

With respect to the accuracy index of the dataset [14], Table 6 compares the suggested technique to the 10 current methodologies. Researchers in the approaches [53–57]

Table 5 An overview of spatial, angular, and Hough feature assembling evaluations

Approach	Accuracy (%)
Gaglio et al. [50]	84.8 & 84.5
Cippitelli et al. ($P = 7$) [51]	94.0 & 93.7
Cippitelli et al. ($P = 11$) [51]	95.1 & 95.0
Cippitelli et al. ($P = 15$) [51]	95.0 & 94.8
Papadopoulos et al. [52]	96.3 & 97.41
SAHC	97.6

Joints-connections, and Frames-connections, both of which are ensembles, are especially noteworthy since they achieve the best prediction accuracy. The suggested technique is shown to improve performance when compared to other methods, and these methods are all evaluated for their correctness. Analysis of SAHC in relation to the other methods used on the dataset [13]

Table 6 An overview of spatial, angular, and Hough feature assembling evaluations

Approach	Accuracy (%)
Qiuhong Ke et al. [53]	93.47
Qiuhong Ke et al. [54]	93.57
Jun Liu et al. [55]	94.1
Fabien Baradel et al. [56]	94.1
Jun Liu et al. [57]	94.9
SAHC	97.7

Joints-connections, and Frames-connections, both of which are ensembles, are especially noteworthy since they achieve the best prediction accuracy. The suggested technique is shown to improve performance when compared to other methods, and these methods are all evaluated for their correctness. Analysis of SAHC in relation to the other methods used on the dataset [11]

obtained accuracies of 93.47, 93.57, 94.1, 94.1, and 94.9% using the aforementioned values. However, for dataset [14], suggested SAHC achieved 97.7% accuracy, which is an enhancement above the previous best.

The effectiveness of a classification method may be summarized in a table called a confusion matrix. A confusion matrix provides a visual representation and summary of a classification algorithm's efficiency. In Figs. 3, 4, 5, 6, see a confusion matrix where action classification and misclassification are taken into account, and where various degrees of intensity are shown. The 20 classes in the dataset [11] are as follows: high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw cross, draw tick, draw circle, hand clap, two-hand wave, side-boxing, bend, forward kick, side-kick, jogging, tennis swing, tennis serve, golf swing, and pick-up and throw. Figure 3 displays the confusion matrix of the dataset [11] as a measure of individual class performance. An excellent class performance was attained with the suggested action description. Every action motion is recognized except for the side-kick class. While the performance of the sidekick class exhibits some bit errors, overall, the description has a net performance of 99.80%.

In order to determine the source of categorization inaccuracies, researchers often turn to a confusion matrix. The columns show what the expected results would have been in each class. The columns, meanwhile, show the forecasts. This table makes it clear which assumptions were incorrect. In Fig. 4, it can be seen the confusion matrix represents the performance of each class on the dataset [12]. As suggested action description outperformed the rest of the class significantly. Class "Baseball swing from the right" has the worst performance overall, while "Tennis serve" and "Two-hand push" do better. Using SAHC, the proposed model can get an interclass performance of 97.6% overall.

Confusion matrix for the dataset [13] is shown in Fig. 5 using a confusion chart, which includes eighteen distinct actions: horizontal arm wave, high arm wave, two-hand wave, high throw, draw X, draw tick, forward kick, side kick, bend, and hand clap, catch cap, toss paper, take umbrella, walk, phone call, drink, sit, and stand up. Performance is lowest in the "Walk" class and highest in the "Draw X" class. As by the use of SAHC to measure performance across classes, as find that it is 97.6% effective.

On the Dataset [14], which includes eight classes such as "approaching," "departing," "exchanging," "hugging," "kicking," "punching," and "pushing," DAP-JF achieved 99.6% [13]. This confusion matrix [58] was created using the SAHC descriptor and is shown in Fig. 6. Class "Departing" indicates the desired outcome, whereas class "Exchanging" reveals the undesirable outcome with slightly less wrong misclassification.

5 Conclusion

To extend the capacity of skeleton joints-based extraction in human body, by recognizing representations at a higher level using motion features, The suggested angular characteristics are resilient against subject fluctuations and represent relative motion between body components in a comprehensive way. Therefore, they propose a joints-connections, which creates issues for preexisting models but allows them to distinguish between difficult actions with similar motion trajectories. Experimental results show that the proposed SAHC features are complementary to existing features, i.e., the joint and bone representations. By incorporating frames-connections and accommodating all aspects of temporal details to achieve action recognition. The article's key contribution is the estimation of full-body human postures using an

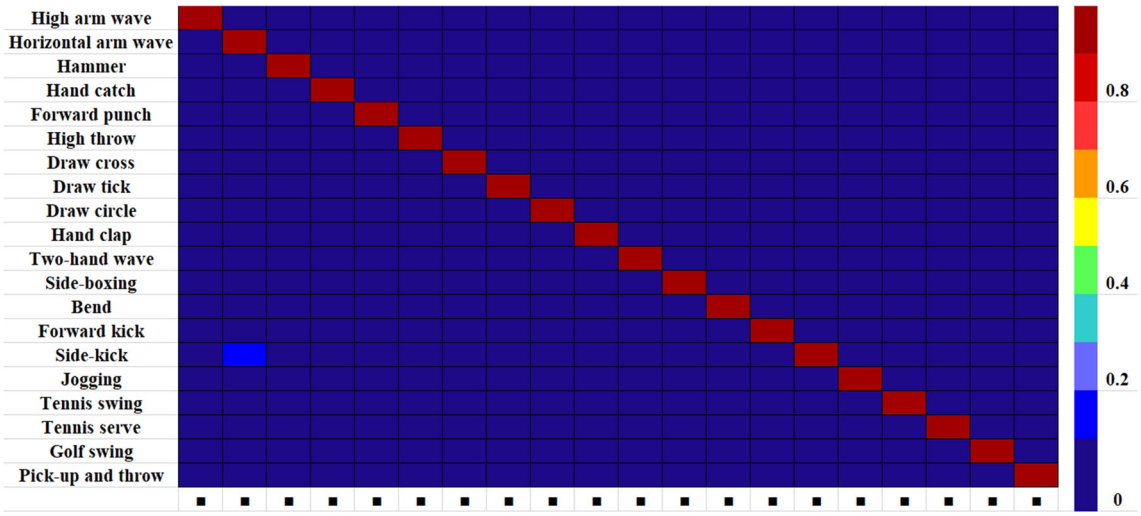


Fig. 3 The proposed SAHC's accuracy in recognizing skeleton-based activities is improved by the use of encoding several types of motion characteristics [11]. Joints' steady state and kinetic domains are analyzed, with the optimal accuracy for each domain marked along the diagonal

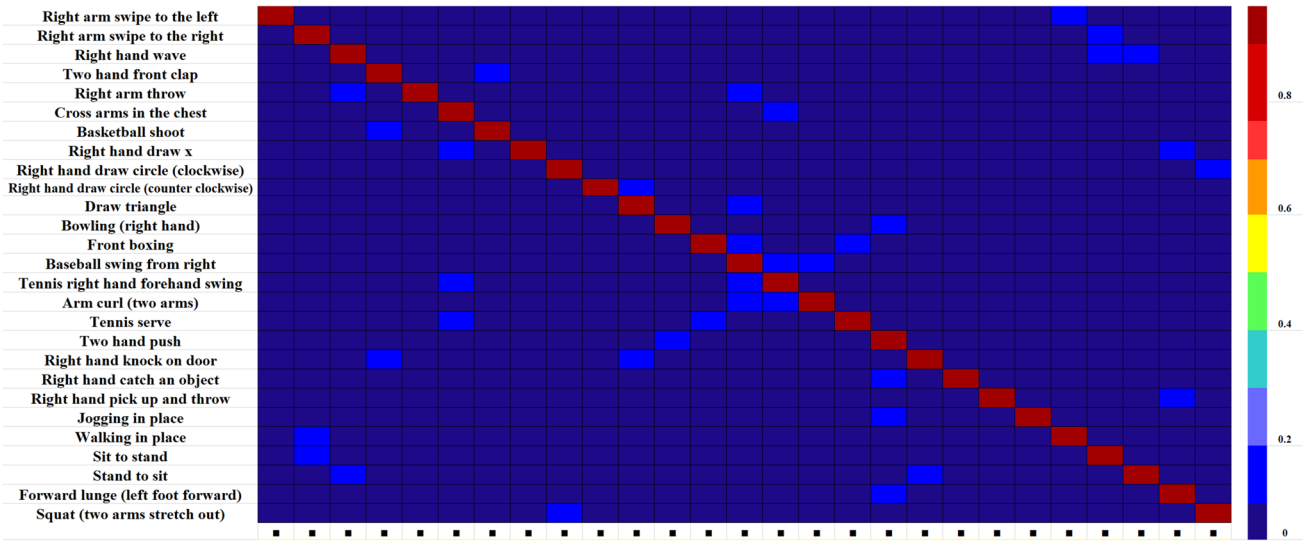


Fig. 4 The proposed SAHC's accuracy in recognizing skeleton-based activities is improved by the use of encoding several types of motion characteristics [12]. Joints' steady state and kinetic domains are analyzed, with the optimal accuracy for each domain marked along the diagonal

SAHC architecture modified for an action recognition regression issue. SAHC is made up of Spatial, Angular, and Hough characteristics that were retrieved utilizing Joint and frame connections in order to obtain skeleton-based action detection features. To enable real-time action detection on edge devices, though accomplish state-of-the-art accurateness on many benchmarks despite the fact keeping computational costs to a minimum. Future work will focus on enhancing the efficacy of the system itself and applying it to the field of improving health through human action recognition.

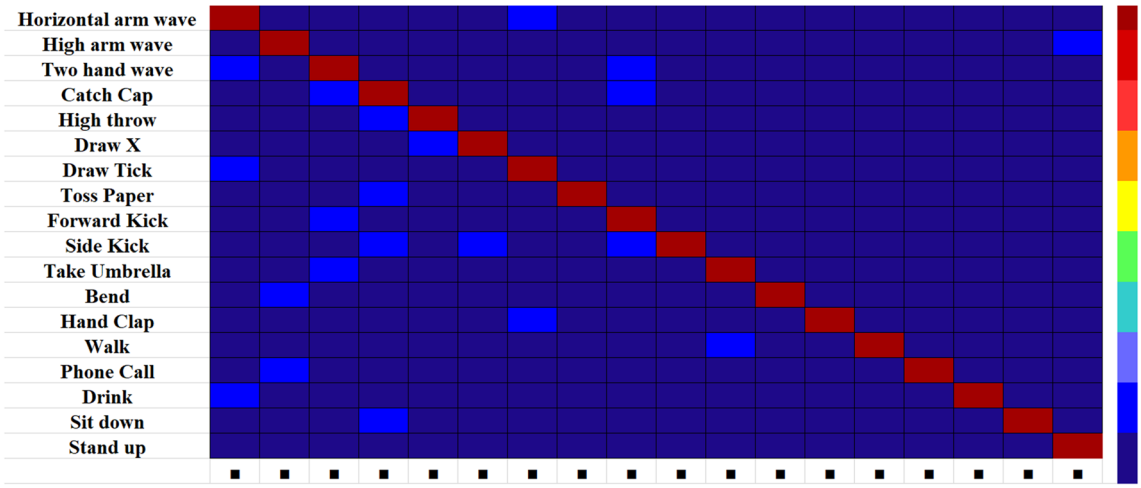
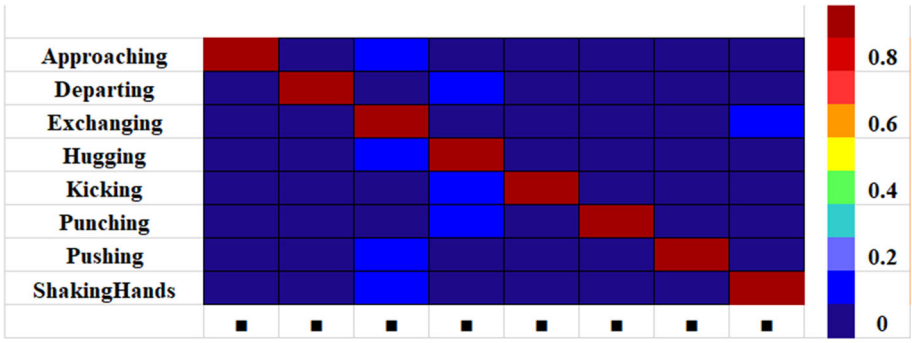


Fig. 5 The proposed SAHC's accuracy in recognizing skeleton-based activities is improved by the use of encoding several types of motion characteristics [13]. Joints' steady state and kinetic domains are analyzed, with the optimal accuracy for each domain marked along the diagonal

Fig. 6 The proposed SAHC's accuracy in recognizing skeleton-based activities is improved by the use of encoding several types of motion characteristics [14]. Joints' steady state and kinetic domains are analyzed, with the optimal accuracy for each domain marked along the diagonal



Acknowledgements The authors acknowledge the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, under project Grant No. 3161.

Author contributions In this paper, I propose SAHC, a new purely relevant action descriptor in this study. I for one create a novel SAHC predictor with a multiple endpoints action recognition backbone system, many diverse characteristics, and a simple prediction network that outputs the action class and the temporal distance between the start and finish of each location. The highlights of the article are given below for your kind perusal. Kindly consider and forward my article for further process. •The spatial, angular, and Hough features, all of which are higher-order representations, are discussed in this article for encoding the static and velocity domains of joints. When applied to the human body, the encoding successfully represents elative motion between body components while being invariant over a wide range of individual body sizes. •Integrating the joints and frames connections into preexisting action recognition systems is a straightforward way to further improve performance. Our results demonstrate that these associations provide valuable further data to the already extant elements, such as the joint representations. •To the best of my knowledge, proposed descriptor the first to combine several types of angular characteristics into state-of-the-art spatial-temporal SAHCs, and our results on a number of benchmarks are among the best available. Meanwhile, the suggested SAHC encoding may provide even a basic model a significant boost in performance. Therefore, the suggested angular encoding enables edge devices to recognize actions in real-time.

Availability of data and materials I have used publicly available datasets which are cited in the article, no further Availability of data and materials.

Declarations

Conflict of interest Proposed work is applicable and includes interests of a financial or personal nature.

Ethical approval Article have followed up: Ethical committees, Internal Review Boards and guidelines followed must be named. When applicable, additional headings with statements on consent to participate and consent to publish are also required.

References

1. Rahimi, S., Aghagolzadeh, A., Ezoji, M.: Human action recognition based on the Grassmann multi-graph embedding. SIViP **13**, 271–279 (2019)
2. Lee, J., Lee,Minhyeok., Lee, Dogyoon., and Lee, Sangyoon.: Hierarchically Decomposed Graph Convolutional Networks for

Skeleton-Based Action Recognition. arXiv preprint arXiv:2208.10741 (2022)
3. Bakhat, K., Kashif Kifayat, M., Islam, S., Mattah Islam, M.: Katz centrality based approach to perform human action recognition by using OMKZ. Signal, Image Video Process. **17**(4), 1677–1685 (2023)
4. Zeng, Ailing., Sun, Xiao., Yang, Lei., Zhao,Nanxuan., Liu, Minhao., and Xu, Q..: Learning skeletal graph neural networks for hard 3d pose estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 11436–11445. (2021)
5. Sijie,Y., Xiong,Yuanjun., and Lin,Dahua.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In Thirty-second DAHCI conference on artificial intelligence (2018)
6. Shi, Lei., Zhang,Yifan., Cheng, Jian., and Lu, Hanqing.: Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 12026–12035 (2019)
7. Cheng, Ke., Zhang,Yifan., Cao,Congqi., Shi, Lei., Cheng, Jian., and Lu, Hanqing.: Decoupling gcn with dropgraph module for skeleton-based action recognition. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16, pp. 536–553. Springer International Publishing (2020)
8. Islam, M.S., Bakhat, K., Khan, R., Mansoor Iqbal, M., Islam, M., Ye, Z.: Action recognition using interrelationships of 3D joints and frames based on angle sine relation and distance features using interrelationships. Appl. Intell. **51**, 6001–6013 (2021)
9. Islam, M.S., Bakhat, K., Iqbal, M., Khan, R., Ye, ZhongFu, Mattah Islam, M.: Representation for action recognition with motion vector termed as: SDQIO. Expert Syst. Appl. **212**, 118406 (2023)
10. Islam, S., Qasim, T., Yasir, M., Bhatti, N., Mahmood, H., Zia, M.: Single-and two-person action recognition based on silhouette shape and optical point descriptors. SIViP **12**, 853–860 (2018)
11. Li, W., Zhang, Z., Liu, Z.: Action recognition based on a bag of 3d points. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pp. 9–14. IEEE (2010)
12. Chen, C., Jafari, R., Kehtarnavaz, N.: Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In 2015 IEEE International conference on image processing (ICIP), pp. 168–172. IEEE (2015)
13. Gaglio, S., Re, G.L., Morana, M.: Human activity recognition process using 3-D posture data. IEEE Trans. Human Mach. Syst. **45**(5), 586–597 (2014)
14. Yun, K., Honorio, J., Chattopadhyay, D., Berg, T.L., Samaras, D.: Two-person interaction detection using body-pose features and multiple instance learning." In 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 28–35. IEEE (2012)
15. Wang, L., Huynh, Du.Q., Koniusz, P.: A comparative review of recent kinect-based action recognition algorithms. IEEE Trans. Image Process. **29**, 15–28 (2019)
16. Koniusz, P., Wang, L., Cherian, A.: Tensor representations for action recognition. IEEE Trans. Pattern Anal. Mach. Intell. **44**(2), 648–665 (2021)
17. Anwar, S., Barnes, N.: Densely residual laplacian super-resolution. IEEE Trans. Pattern Anal. Mach. Intell. **44**(3), 1192–1204 (2020)
18. Li, Dongxu., Yu, Xin., Xu,Chenchen., Petersson,Lars., and Li, Hongdong.: Transferring cross-domain knowledge for video sign language recognition." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6205–6214 (2020)
19. Zhang, Yiyang., Liu, Feng., Fang, Zhen., Yuan, Bo., Zhang, G., and Lu, J.: Clarinet: a one-step approach towards budget-friendly unsupervised domain adaptation. arXiv preprint arXiv:2007.14612 (2020)
20. Wang, Lei., Koniusz,Piotr., and Huynh, Du Q.: Hallucinating idt descriptors and i3d optical flow features for action recognition with cnns. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8698–8708 (2019)
21. Wang, Yu Guang., Li, Ming., Ma, Zheng., Montufar, Guido., Zhuang, Xiaosheng., and Fan, Yanan.: Haar graph pooling. In International conference on machine learning, pp. 9952–9962. PMLR (2020)
22. Li, Maosen., Chen, Siheng., Chen, Xu., Zhang, Ya., Wang, Yanfeng., and Tian, Qi.: Actional-structural graph convolutional networks for skeleton-based action recognition." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 3595–3603 (2019)
23. Si, Chenyang., Chen, Wentao., Wang, Wei., Wang, Liang., and Tan, Tieniu.: An attention enhanced graph convolutional lstm network for skeleton-based action recognition." In proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 1227–1236 (2019)
24. Zhang, Pengfei., Lan, Cuiling., Zeng, Wenjun., Xing, Junliang., Xue, Jianru., and Zheng, Nanning.: Semantics-guided neural networks for efficient skeleton-based human action recognition. In proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 1112–1121 (2020)
25. Liu, Ziyu., Zhang, Hongwen., Chen, Zhenghao., Wang, Zhiyong., and Ouyang, Wanli.: Disentangling and unifying graph convolutions for skeleton-based action recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 143–152 (2020)
26. Qin, X., Cai, R., Jiabin, Yu., He, C., Zhang, X.: An efficient self-attention network for skeleton-based action recognition. Sci. Rep. **12**(1), 4111 (2022)
27. Xu, Kailin., Ye, Fanfan., Zhong, Qiaoyong., and Xie, Di.: Topology-aware convolutional neural network for efficient skeleton-based action recognition. In Proceedings of the DAHCI Conference on Artificial Intelligence, vol. 36, no. 3, pp. 2866-2874 (2022)
28. Memmesheimer, Raphael., Häring, Simon., Theisen, Nick., and Paulus, Dietrich.: Skeleton-DML: deep metric learning for skeleton-based one-shot action recognition. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3702–3710 (2022)
29. Wang, He., He, Feixiang., Peng, Zhexi., Shao, Tianjia., Yang, Yong-Liang., Zhou, Kun., and Hogg, David.: Understanding the robustness of skeleton-based action recognition under adversarial attack. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14656–14665 (2021)
30. Diao, Yunfeng., Shao, Tianjia., Yang, Yong-Liang., Zhou, Kun., and Wang, He.: BASAR: black-box attack on skeletal action recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7597–7607 (2021)
31. Hu, K., Ding, Y., Jin, J., Weng, L., Xia, M.: Skeleton motion recognition based on multi-scale deep spatio-temporal features. Appl. Sci. **12**(3), 1028 (2022)
32. Huang, Z., Qin, Y., Lin, X., Liu, T., Feng, Z., Liu, Y.: Motion-driven spatial and temporal adaptive high-resolution graph convolutional networks for skeleton-based action recognition. IEEE Trans. Circuits Syst. Video Technol. **33**(4), 1868–1883 (2022)
33. Zhu, X., Zhou, Y., Wang, D., Ouyang, W., Rui, Su.: MLST-former: multi-level spatial-temporal transformer for group activity recognition. IEEE Trans. Circuits Syst. Video Technol. **33**, 3383 (2022)
34. Islam, M.S., Bakhat, K., Rashid Khan, M., Islam, M., Ye, ZhongFu: Single and two-person (s) pose estimation based on R-WAA. Multimedia Tools Appl **81**, 1–14 (2022)
35. Ren, Min., He,Lingxiao., Liao, Xingyu., Liu, Wu., Wang, Yunlong., and Tan, Tieniu.:Learning instance-level spatial-temporal patterns

for person re-identification. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 14930–14939 (2021)
36. Su, Y., Zhu, H., Tan, Y., An, S., Xing, M.: Prime: privacy-preserving video anomaly detection via motion exemplar guidance. Knowl.-Based Syst. **278**, 110872 (2023)
37. Azher, U.M., Lee, Y.-K.: Feature fusion of deep spatial features and handcrafted spatiotemporal features for human action recognition. Sensors **19**(7), 1599 (2019)
38. Ryu, J., Patil, A.K., Chakravarthi, B., Balasubramanyam, A., Park, S., Chai, Y.: Angular features-based human action recognition system for a real application with subtle unit actions. IEEE Access **10**, 9645–9657 (2022)
39. Liu, J., Li, Y.: The visual movement analysis of physical education teaching considering the generalized hough transform model. Comput. Intell. Neurosci. (2022). https://doi.org/10.1155/2022/3675319
40. Jin, Ke., Jiang, M., Kong, J., Huo, H., Wang, X.: Action recognition using vague division DMMs. J. Eng. **2017**(4), 77–84 (2017)
41. Luo, Jiajia., Wang, Wei., and Qi, Hairong.: Group sparsity and geometry constrained dictionary learning for action recognition from depth maps. In Proceedings of the IEEE international conference on computer vision, pp. 1809–1816 (2013)
42. Du, Yong., Wang, Wei., and Wang, Liang.: Hierarchical recurrent neural network for skeleton based action recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1110–1118 (2015)
43. Chen, Chen., Jafari, Roozbeh., and Kehtarnavaz, Nasser.: Action recognition from depth sequences using depth motion maps-based local binary patterns. In 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 1092–1099. IEEE, (2015)
44. Xu, Haining., Chen, Enqing., Liang, Chengwu., Qi, Lin., and Guan, Ling.: Spatio-Temporal Pyramid Model based on depth maps for action recognition. In 2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP), pp. 1–6. IEEE, (2015)
45. Liu, Mengyuan., and Yuan, Junsong.: Recognizing human actions as the evolution of pose estimation maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1159–1168 (2018)
46. Tasnim, N., Islam, M.M., Baek, J.-H.: Deep learning-based action recognition using 3D skeleton joints information. Inventions **5**(3), 49 (2020)
47. McNally, William., Wong,Alexander., and McPhee, John.: STAR-Net: Action Recognition using Spatio-Temporal Activation Reprojection." arXiv preprint arXiv:1902.10024 (2019)
48. Islam, M.S., Bakhat, K., Khan, R., Nuzhat Naqvi, M., Islam, M., Ye, Z.: Applied human action recognition network based on SNSP features. Neural. Process. Lett. **54**(3), 1481–1494 (2022)
49. Chikhaoui, Belkacem., and Gouineau,Frank.: Towards automatic feature extraction for activity recognition from wearable sensors: a deep learning approach. In 2017 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 693–702. IEEE, (2017)
50. Gaglio, S., Re, G.L., Morana, M.: Human activity recognition process using 3-D posture data. IEEE Trans. Human-Mach. Syst. **45**(5), 586–597 (2014)
51. Cippitelli, E., Gasparrini, S., Gambi, E., Spinsante, S.: A human activity recognition system using skeleton data from rgbd sensors. Comput. Intell. Neurosci. **2016**, 21 (2016)
52. Papadopoulos, Konstantinos., Antunes, Michel., Aouada, Djamila., and Ottersten, Björn.: Enhanced trajectory-based action recognition using human pose. In 2017 IEEE International Conference on Image Processing (ICIP), pp. 1807–1811. IEEE (2017)
53. Ke, Q., An, S., Bennamoun, M., Sohel, F., Boussaid, F.: Skeletonnet: mining deep part features for 3-d action recognition. IEEE Signal Process. Lett. **24**(6), 731–735 (2017)
54. Ke, Qiuhong., Bennamoun, M., An, S., Sohel, F., and Boussaid,Farid.: A new representation of skeleton sequences for 3d action recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3288–3297 (2017)
55. Liu, Jun., Wang, Gang., Hu, Ping., Duan, Ling-Yu., and Kot, Alex C.: Global context-aware attention LSTM networks for 3D action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1647–1656 (2017)
56. Baradel, Fabien., Wolf,Christian., and Mille, Julien.: Pose-conditioned spatio-temporal attention for human action recognition. arXiv preprint arXiv:1703.10106 (2017)
57. Liu, J., Wang, G., Duan, L.-Y., Abdiyeva, K., Kot, A.C.: Skeleton-based human action recognition with global context-aware attention LSTM networks. IEEE Trans. Image Process. **27**(4), 1586–1599 (2017)
58. Bakhat, K., Kashif Kifayat, M., Islam, S., Mattah Islam, M.: Human activity recognition based on an amalgamation of CEV & SGM features. J. Intell. Fuzzy Syst. Preprint **43**, 1–12 (2022)