Skeleton-Based Hand Gesture Recognition by Using Multi-input Fusion Lightweight Network

Qihao Hu[1], Qing Gao[2,3](✉), Hongwei Gao[1], and Zhaojie Ju[4](✉)

[1] School of Automation and Electrical Engineering, Shenyang Ligong University,
Shenyang 110159, China
ghw1978@sohu.com

[2] Institute of Robotics and Intelligent Manufacturing and School of Science and Engineering,
The Chinese University of Hong Kong, Shenzhen 518172, China
gaoqing@cuhk.edu.cn

[3] Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen 518129, China

[4] School of Computing, University of Portsmouth, Portsmouth PO13HE, UK
Zhaojie.Ju@port.ac.uk

Abstract. Skeleton-based hand gesture recognition has achieved great success in recent years. However, most of the existing methods cannot extract spatiotemporal features well due to the skeleton noise. In real applications, some large models also suffer from a huge number of parameters and low execution speed. This paper presents a lightweight skeleton-based hand gesture recognition network by using multi-input fusion to address those issues. We convey two joint-oriented features: Center Joint Distances (CJD) feature and Center Joint Angles (CJA) feature as the static branch. Besides, the motion branch consists of Global Linear Velocities (GLV) feature and Local Angular Velocities (LAV) feature. Fusing static and motion branches, a robust input can be generated and fed into a lightweight CNN-based network to recognize hand gestures. Our method achieves 95.8% and 92.5% hand gesture recognition accuracy with only 2.24M parameters on the 14 gestures and 28 gestures of the SHREC'17 dataset. Experimental results show that the proposed method outperforms state-of-the-art (SOAT) methods.

Keywords: Skeleton-based hand gesture recognition · Multi-input fusion · Joint-oriented feature Second Keyword

1 Introduction

Recently, thanks to the development of machine learning and computer vision, dynamic hand gesture recognition becomes a popular research topic in many fields, e.g., human-computer interaction (HRI), sign language interpretation and medical assistive applications. Over the past decade, with the widespread use of depth cameras and great developing of hand-pose estimation, skeletal data of high accuracy can be generated easily. Skeletal data is a time sequence of 3D coordinates of multiple hand joints. Compared with RGB and RGB-D inputs, skeletal data is more robust to background changes

H. Liu et al. (Eds.): ICIRA 2022, LNAI 13455, pp. 24–34, 2022.
https://doi.org/10.1007/978-3-031-13844-7_3

and illumination variations. Skeleton-based gesture recognition has shown powerful classification effect in many applications.

One essential problem in dynamic hand gesture recognition is how to extract rich features to fully describe the variations of spatial configurations and temporal dynamics in gestures. Skeleton-based gesture recognition algorithms are developing rapidly, and there are mainly three deep learning methods, namely Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Graph Convolutional Networks (GCN). The above three methods transform the raw skeletal data into pseudo graph, time series and graph structure for feature extraction, respectively. CNN-based method is of frequently used as a backbone model of real-time gesture detection and recognition because of its compact structure and fast processing speed.

In real applications, a desirable gesture recognition model should be adaptable to the influence caused by the variation of the viewpoints and achieves high recognition accuracy. It also should run efficiently by using a few parameters. To meet those requirements, we propose a multi-input fusion lightweight network, which is a CNN model equipped with a static features branch and a motion feature branch. The proposed model takes into account both the recognition accuracy and the execution speed. Extensive experiments are conducted on public dataset to demonstrate the effectiveness of our proposed method.

Specifically, our research is implemented based on the unique properties of skeletal data. To tackle the issue of viewpoint rotation, we propose a simplified joint distances feature. Meanwhile, to alleviate magnitude changes caused by the distance variations between observer and hand, we introduce the feature of center joint angles. As shown in Fig. 1, joint distances feature and joint angles feature can cope with the variations of input data caused by viewpoints changes. To make full use of the rich spatio-temporal information of skeleton data, motion features generated by joint coordinates and center joint angles are extracted as input features. We adopt a fast slow frame generation method, which is applied to motion branch. Different frame generation method can distinguish the influence of the speed of gestures. At the network structure level, we employ 1D convolutional neural network to embed the above features, and then utilize 2D convolutional neural network to process the fused features. The network structure not only provides small parameter scale and fast running speed, but also can extract spatio-temporal information well. Compared with other similar CNN method, our proposed method has achieved better performance through experimental verification.

The contributions of this paper are as follows:

1. Two geometric features with translation, rotation and scaling invariance are compounded to constitute the static feature module. Besides, motion features are introduced to improve the sensitivity of the model to different temporal and spatial scales, and improve the classification effectiveness.
2. The network architecture combining 1D CNN and 2D CNN is adopted to extract the rich spatio-temporal features, and avoids unnecessary parameters and slow processing speed.
3. Comparative experiment proves that the accuracy of the model is ahead of other advanced CNN-based networks.

The rest of the paper are arranged as follows: We review the related works in Sect. 2. Section 3 introduces the methodology of our model. The fourth section makes ablation studies and comparative experiments to demonstrate the effectiveness of our model. The last section concludes our paper and the future works.

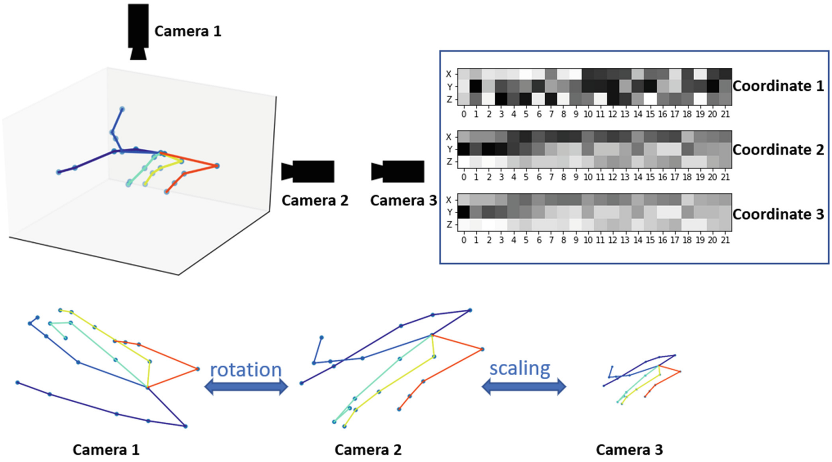


Fig. 1. Variations of Cartesian coordinates caused by viewpoint changes. Camera 1 and Camera 2 have different observation directions, which makes the skeleton rotation. Camera 2 and Camera 3 have different observation distances, which makes the skeleton scaling.

2 Related Works

2.1 Static and Motion Features

In many prior works, static features are frequently applied in recognizing action and gesture tasks, e.g., position, angle, distance, velocity and acceleration. Li et al. [9] propose 2D and 3D joint distance map (JDM) features. Zhang et al. [10] provide a variety of distance and angle features of lines and planes. Liao et al. [4] utilize a set of joint-oriented features for human action recognition. Song et al. [8] propose an early fused Multiple Input Branches (MIB) architecture to capture structure features from skeleton sequences.

Motion features contain rich dynamic information. Chen et al. [11] extract finger articulated features from the hand skeleton by a variational autoencoder (VAE). Choutas et al. [12] introduce a fixed-sized representation that encodes pose motion. Feichtenhofer et al. [5] propose two scales motion features difference of slow and fast motions. Different from these works, our work obtains static and motion features by a center joint-oriented method, which can reduce the noise of skeletal data and consume a small amount of computing resources.

2.2 Skeleton-Based Gesture Recognition

Skeleton-based action recognition has been studied for decades. Yang et al. [1] propose DD-Net solely based on 1D CNNs for easy computation and training, while taking into consideration the integration of location-viewpoint invariant feature Joint Collection Distances (JCD) and two-scale global motion features. Ding et al. [13] encoded five spatial skeleton features into images and then fed those features to a CNN structure. Ke et al. [14] created texture arrays from 3D coordinates of body joints using 4 key body joints as a reference to form the center of a coordinate system by which the 3D positions of body joints are shifted before conversion into cylindrical coordinates. Twelve maps were generated which are fed to 12 CNN streams. Guo et al. [18] propose a normalized edge convolution operation to recognize hand gestures. In [15], a skeleton sequence representation was proposed in the form of a matrix that concatenates the joint coordinates in each instant and arranged those vector representations in a chronological order. Vemulapalli et al. [6] utilize rotations and translations to represent the 3D geometric relationships of body parts in Lie group. Some methods cost huge computing resources [4, 10–12] or contain redundant input [9]. Inspired by [1], we design our method on two aspects: introduce new features for skeleton sequences and propose novel neural network architectures.

3 Methodology

This section will describe the implementation process of the model. The framework of our networks is shown in Fig. 2. Our network takes a hand skeleton sequence as input and predicts the class label of dynamic hand gesture. It consists of two main branches, which process static features and motion features, respectively. In the following, we explain our motivation for designing input features and network structure of the model.

3.1 Modeling Static Feature by Center Joint Oriented Method

Raw skeleton data is a set of 3D Cartesian coordinates of hand joints. For one frame, the n$^{\text{th}}$ joint can be donated by$J_n = \left(J_x, J_y, J_z\right)$, where n ∈ {0, 1, 2, …, (N − 1)} and N is the number of the hand joints. However, the Cartesian coordinate is variant to locations and viewpoints. As Fig. 1 shows, when the position of observer changes, skeletons may rotate or zoom. The Cartesian coordinate will be changed significantly. However, the geometric feature (e.g., distances and angles) is location-viewpoint invariant, and thereby we adopt it as the static feature input of the network. To reduce the computation and decrease the noise interference of bone data, we adopt a joint oriented method to extract static features.

First, a center joint J_0 is selected as original point. For each frame, the Euclidean distance D_n between joints J_0 and J_n can be denoted as.

$$D_n = \|J_n - J_0\|_2, n \in \{1, 2, \ldots, (N-1)\}. \tag{1}$$

The cosine $A_{i,k}$ of the joint angle $J_i - J_0 - J_k$ is denoted as

$$A_{i,k} = cos\ < \vec{J_i}, \vec{J_k} >, i, k \in \{1, 2, \ldots, (N-1)\}, i \neq k, \tag{2}$$

where $\vec{J_i}$ is the vector from $\vec{J_0}$ to $\vec{J_i}$, and $< \vec{J_i}, \vec{J_k} >$ is the angle of vector $\vec{J_i}$ and $\vec{J_k}$.

Except for the center joint J_0, the other $(N-1)$ joints can generate $(N-1)$ joint oriented distances by formula (1). The collection of those distances is named Center Joint Distances (CJD). The dimension of $CJD : \left[D_1 D_2 \ldots D_{N-1}\right]$ is $(N-1)$. Similarly, the collection of all joint angles is named Center Joint Angles (CJA). The CJA feature can be denoted as

$$CJA = \begin{bmatrix} A_{2,1} & & \\ \vdots & \ddots & \\ A_{(N-1),1} & \cdots & A_{(N-1),(N-2)} \end{bmatrix}. \tag{3}$$

In our processing, the CJA is flattened to be a one-dimensional matrix and the dimension of the flattened CJA is

$$d_{CJA} = C_{N-1}^2 = \frac{(N-1)(N-2)}{2}. \tag{4}$$

3.2 Extracting Global and Local Motion Features by Different Frames

Since static features do not contain motion information, we introduce the motion features as another input. Two kinds of motion features can be extracted by calculating the temporal differences of the Cartesian coordinate feature and geometric feature. Inspired by [5], we adopt the slow-fast networks method to extract two scale of velocities:

$$s(t) = x(t+1) - x(t), t = 1, 2, 3, \ldots, T-1, \tag{5}$$

$$f(t) = x(t+2) - x(t), t = 1, 3, 5, \ldots, T-2, \tag{6}$$

where $s(t)$ and $f(t)$ are the slow and fast motion at frame t. $x(t)$ is the physical quantity at frame t. $x(t+1)$ and $x(t+2)$ represents the physical quantities 1 frame and 2 frames after frame t, respectively. The frame number of the temporal sequence is denoted as T.

To represent the motion, we introduce Global Linear Velocities (GLV) and Local Angular Velocities (LAV). GLV represent the movements of all hand joints' coordinates $J_n = \left(J_x, J_y, J_z\right)$ in the Euclidean space, while LAV represent the rates of change of Center Joint Angles (CJA).

3.3 Dimension Adjustment and Feature Fusion by CNN Embedding

After extracting static and motion features, we adopt embedding method similar to [1]. 1D convolutions are used to transform the features into four embeddings, which are concatenated together to feed in spatiotemporal 2D representation layers. The embedding method can automatically learn the correlation between joint points and reduce the noise interference of skeletal data. In order to fuse different features and eliminate the inconsistency of different time dimensions, we adopt zero padding method and two kinds of different embedding methods.

Specially, dim of static features is $d_{static}{*}T$ for it is extracted per frame. While dim of slow-motion feature is $d_{motion}{*}(T-1)$. We employ a zero padding in slow motion feature so that it can match with the frame number of the static features. Same zero padding is employed in fast motion feature as well so that its dimension can be resized to $T/2$. We introduce two embedding operations for features of different dimension.

More formally, let embedding representations of static feature, slow motion features and fast motion features to be e_{static}, e_{slow} and e_{fast}, respectively. The embedding operation is as follows:

$$e_{static} = Embed_s[CJD \oplus CJA], \tag{7}$$

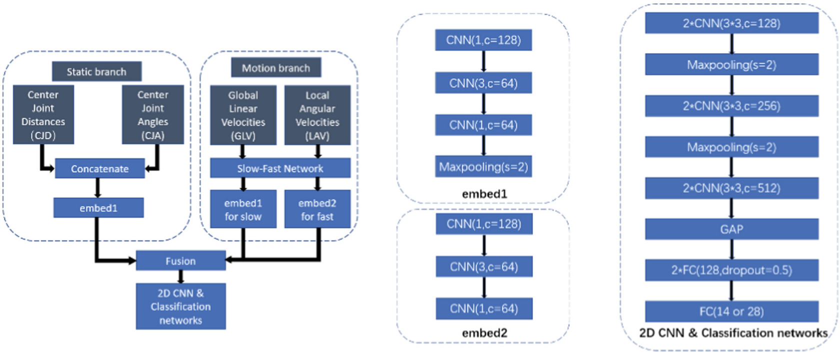
$$e_{slow} = Embed_s[s(t)], \tag{8}$$

$$e_{fast} = Embed_f\left[f(t)\right], \tag{9}$$

where $\oplus$ is the concatenation operation. Our network further fuses those embedding features to a representation e by concatenation:

$$e = e_{static} \oplus e_{slow} \oplus e_{fast}. \tag{10}$$

Rich spatial features are extracted by embedding and feature fusion. Then we use 2D convolutional neural network to extract spatiotemporal features and classification. Our



(a) Overview

(b) Details of embedding operators and networks

Fig. 2. The network architecture of our network. "2 * CNN (3 * 3, c = 128)" denotes two 2D Conv-Net layers (kernel size = 3 * 3, channels = 128), and "CNN (1, c = 128) represents a 1D ConvNet layer with a 1-dimension kernel. Other CNN layers are defined in the same way. GAP denotes Global Average Pooling. "Maxpooling(s = 2)" denotes a Maxpooling with 2 strides. FC denotes Fully Connected Layers (Dense Layers in our experiments).

feature fusion network embeds the static features (CJD, CJA) and the two-scale motion features into latent vectors at each frame. Through the embedding, the correlation of joints can be automatically learned. Also, joint-oriented method and embedding process can reduce the effect of skeleton noise. The overall process is shown in Fig. 2.

4 Experiments

4.1 Dataset

The performance of our method is evaluated on SHREC'17 Track dataset[2], which is a challenging gesture dataset with skeletal data. In this subsection, we introduce the experimental dataset in detail.

The SHREC'17 Track dataset [2] use Intel RealSense short range depth camera to collect hand gesture data. The depth images and hand skeletons were captured at 30 frames per second. Each sample gesture has 20 to 50 frames. Each frame of sequences contains a depth image, the coordinates of 22 joints both in the 2D depth image space and in the 3D world space forming a full hand skeleton. We take only 3D hand skeletons sequences as the raw data for all experiments.

The dataset contains sequences of 14 hand gestures performed in two ways: using one finger and the whole hand. Each gesture is performed between 1 and 10 times by 28 participants in 2 ways, resulting in 2800 sequences. Those 2800 sequences are divided into 1960 sequences (70% of the dataset) for training and 840 sequences (30% of the dataset) for testing. We adopt the same evaluation metric.

4.2 Training Details

The project was completed on a computer equipped with Intel Xeon E-2136 CPU and NVIDIA Quadro P5000 GPU. The environment of deep learning is Python3.7, tensorflow2.4.0, CUDA11.0.

To show the generalization of our methods, we use the same configuration for all experiments. Skeleton sequences are normalized into 32 frames which is as same as the settings in [1]. Besides, the learning rate is set to 0.001 for faster convergence. We use the Adam as the optimizer and the cross-entropy as the loss function. Training for 400 epochs with 128 batches, we achieve the following experimental results.

4.3 Ablation Studies

In this experiment, we explore how each feature component contributes to the hand gesture recognition performance by removing one or more component while remaining others unchanged. We conduct experiments on SHREC-28 dataset. Except for the explored parts, other details are set the same for fair comparison.

Table 1 shows the necessity of each input branch. With the increase of branches, the model performance is improved. This phenomenon further confirms the effectiveness of the data preprocessing module. More specifically, similar to video recognition, motion features play an important role in dynamic hand gesture recognition. Without motion feature, a network with solely static feature input only achieves 70.95% accuracy. Besides, we cannot ignore the contributions of the static geometric feature. The CJD feature provides our multi-input network rotation invariability property, while the CJA feature provides scaling invariability property. The ablation studies prove that all of the input branch in our method make the input robust.

Table 1. Contributions of different components

Ablations	CJD	CJA	Motion	Accuracy
	√	√	×	70.95%
	×	×	√	86.43%
	√	×	√	91.31%
	×	√	√	90.83%
Ours	√	√	√	**92.50%**

4.4 Comparison with Previous Methods

The hand gesture classification results of SHREC'17 Track dataset are presented in Table 2 and more details are listed in their confusion matrices. The confusion matrices of 14 gestures and 28 gestures are shown in Fig. 3(a) and (b), respectively.

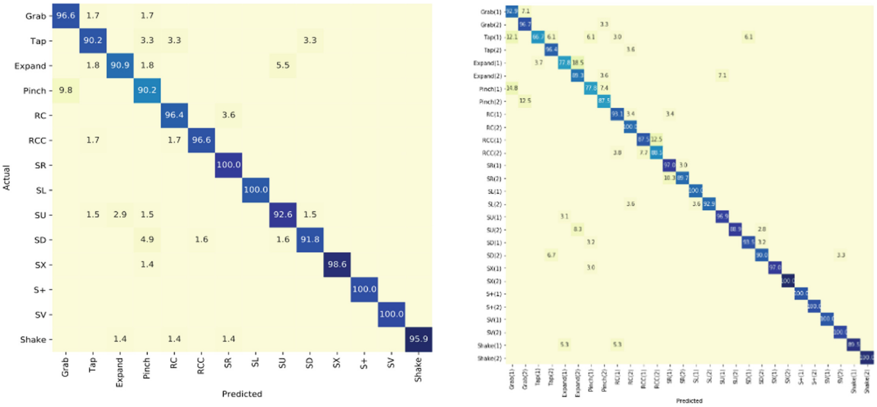
As shown in Table 2, our network achieves the accuracy of 95.8% for the 14 gestures setting and 92.5% for the 28 gestures setting. The effect of our model outperforms the state-of-the-art models'. This shows that our method has a satisfactory effect on hand gesture recognition. Our model brings 1.2% and 0.6% improvements for 14 gestures and 28 gestures setting compared with the state-of-the-arts. Due to the simple CNN-based structure, our model contains only 2.24M parameters, which is smaller than many

Table 2. Accuracy of SHREC dataset

Method	Parameters	14 Gestures	28 Gestures
Dynamic hand [3]	–	88.2%	81.9%
Key-frame CNN [2]	7.92M	82.9%	71.9%
CNN + LSTM [21]	8–9M	89.8%	86.3%
Parallel CNN [20]	13.83M	91.3%	84.4%
STA-Res-TCN [9]	5–6M	93.6%	90.7%
MFA-Net [11]	–	91.3%	86.6%
NormEdgeConv [18]	–	92.9%	91.1%
DD-Net [1]	**1.82M**	94.6%	91.9%
Our method	2.24M	**95.8%**	**92.5%**

other methods and only 0.42M more than DD-Net [1]. Compared with other methods, the proposed model utilizes multi-features as input and a lightweight network structure, which leads to high classification effect and fast execute speed. Thus, our method is hardware-friendly.

As shown in Fig. 3(a), our network achieves recognition rate higher than 95.0% in 9 of the 14 gestures, and achieves 100.0% recognition rate in 4 of the 14 gestures. All 14 gestures can be classified with more than 90.0% accuracy. Figure 3(b) shows the confusion matrix of 28 gestures setting. The proposed model achieves recognition rate higher than 90.0% in 18 of 28 gestures and recognition rate higher than 95.0% in 13 of 28 gestures. Our model shows high classification accuracy for many different hand gesture categories.



(a) 14 hand gestures

(b) 28 hand gestures

Fig. 3. Confusion matrices of SHREC dataset (14 hand gestures & 28 hand gestures)

5 Conclusion

This paper proposed a pipeline for skeleton-based hand gesture recognition. First, we introduced new static and motion features as robust input for our network. To satisfy calculation speed of some real-time hand detection and recognition applications, a lightweight CNN structure was proposed. Compared with other methods with numerous parameters, our network has simple structure and requires less memory and processing power. Our network showed great accuracy and speed advantages over similar networks on our experimental dataset.

To improve the effectiveness of the algorithm and make it better adapt to different environments, the following aspects can be considered for future work:

- We have verified the effectiveness of the network on the SHREC'17 dataset. Even though the model achieved satisfactory results, it needs to be tested on other benchmark datasets for robustness and generalization;

- More new features can be proposed and fused to the input branch. Besides, new fusion methods can be utilized instead of simple concatenation;
- Other powerful convolutional neural networks, e.g., 3D-CNN, can be used to explore rich spatiotemporal information.

Acknowledgements. The authors would like to acknowledge the support from the AiBle project co-financed by the European Regional Development Fund, National Natural Science Foundation of China (grant No. 52075530, 62006204), Guangdong Basic and Applied Basic Research Foundation (2022A1515011431), and Shenzhen Science and Technology Program (RCBS20210609104516043).

References

1. Yang, F., Wu, Y., Sakti, S., Nakamura, S.: Make skeleton-based action recognition model smaller, faster and better. In: Proceedings of the ACM Multimedia asia, pp. 1–6 (2019)
2. De Smedt, Q., Wannous, H., Vandeborre, J.P., Guerry, J., Le Saux, B., Filliat, D.: Shrec 2017 track: 3D hand gesture recognition using a depth and skeletal dataset. In: 3DOR-10th Eurographics Workshop on 3D Object Retrieval, pp. 1–6 (2017)
3. De Smedt, Q., Wannous, H., Vandeborre, J.P.: Skeleton-based dynamic hand gesture recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–9 (2016)
4. Liao, L.C., Yang, Y.H., Fu, L.C.: Joint-oriented features for skeleton-based action recognition. In: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), pp. 1154–1159. IEEE (2019)
5. Feichtenhofer, C., Fan, H., Malik, J., He, K.: SlowFast networks for video recognition. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6202–6211. IEEE (2019)
6. Vemulapalli, R., Arrate, F., Chellappa, R.: Human action recognition by representing 3D skeletons as points in a lie group. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 588–595. IEEE Computer Society (2014)
7. Gao, Q., Liu, J., Ju, Z., Zhang, X.: Dual-hand detection for human–robot interaction by a parallel network based on hand detection and body pose estimation. IEEE Trans. Industr. Electron. **66**(12), 9663–9672 (2019)
8. Song, Y.F., Zhang, Z., Shan, C., Wang, L.: Constructing stronger and faster baselines for skeleton-based action recognition. IEEE Trans. Pattern Anal. Mach. Intell. (2022)
9. Li, C., Hou, Y., Wang, P., Li, W.: Joint distance maps based action recognition with convolutional neural networks. IEEE Signal Process. Lett. **24**(5), 624–628 (2017)
10. Zhang, S., et al.: Fusing geometric features for skeleton-based action recognition using multilayer LSTM networks. IEEE Trans. Multimedia **20**(9), 2330–2343 (2018)
11. Chen, X., Wang, G., Guo, H., Zhang, C., Wang, H., Zhang, L.: Mfa-net: motion feature augmented network for dynamic hand gesture recognition from skeletal data. Sensors, **19**(2), 239 (2019)
12. Choutas, V., Weinzaepfel, P., Revaud, J., Schmid, C.: PoTion: pose motion representation for action recognition. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7024–7033. IEEE (2018)
13. Ding, Z., Wang, P., Ogunbona, P.O., Li, W.: Investigation of Different Skeleton Features for CNN-based 3D Action Recognition. IEEE Computer Society, IEEE Computer Society (2017)

14. Ke, Q., Bennamoun, M., An, S., Sohel, F., Boussaid, F.: A new representation of skeleton sequences for 3D action recognition. In: CVPR 2017, pp. 3288–3297. IEEE Computer Society (2017)
15. Li, C., Zhong, Q., Xie, D., Pu, S.: Skeleton-based action recognition with convolutional neural networks. In: 2017 IEEE International Conference on Multimedia \& Expo Workshops (ICMEW), pp. 597–600. IEEE Computer Society (2017)
16. Huang, Z., Wan, C., Probst, T., Van Gool, L.: Deep learning on lie groups for skeleton-based action recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6099–6108. IEEE Computer Society (2016)
17. Paulo, J.R., Garrote, L., Peixoto, P., Nunes, U.J.: Spatiotemporal 2D skeleton-based image for dynamic gesture recognition using convolutional neural networks. In: 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), pp. 1138–1144. IEEE (2021)
18. Guo, F., He, Z., Zhang, S., Zhao, X., Tan, J.: Normalized edge convolutional networks for skeleton-based hand gesture recognition. Pattern Recogn. **118**(6), 108044 (2021)
19. Sabater, A., Alonso, I., Montesano, L., Murillo, A.C.: Domain and view-point agnostic hand action recognition. IEEE Robot. Autom. Lett. **6**(4), 7823–7830 (2021)
20. Devineau, G., Xi, W., Moutarde, F., Yang, J.: Convolutional neural networks for multivariate time series classification using both inter-and intra-channel parallel convolutions. In: Reconnaissance des Formes, Image, Apprentissage et Perception (RFIAP'2018), June 2018
21. Nunez, J.C., Cabido, R., Pantrigo, J., et al.: Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition. Pattern Recogn. J. Pattern Recogn. Soc. **76**, 80–94 (2018)
22. Gao, Q., Liu, J., Ju, Z.: Robust real-time hand detection and localization for space human–robot interaction based on deep learning. Neurocomputing **390**, 198–206 (2020)
23. Song, Y.F., Zhang, Z., Shan, C., Wang, L.: Stronger, faster and more explainable: a graph convolutional baseline for skeleton-based action recognition. In: MM 2020: The 28th ACM International Conference on Multimedia pp. 1625–1633. ACM (2020)