Journal of Cases on Information Technology
Volume 24 • Issue 5

Application of Gesture Recognition Based on Spatiotemporal Graph Convolution Network in Virtual Reality Interaction

Ting Liao, Sichuan University of Arts and Science, China*

ABSTRACT

Aiming at the low recognition rate of traditional gesture, a gesture recognition algorithm based on spatiotemporal graph convolution network is proposed in this paper. Firstly, the dynamic gesture data were preprocessed, including removing invalid gesture frames, completing gesture frame data, and normalization of joint length. Then, the key frame of the gesture is extracted according to the given coordinate information of the hand joint. A connected graph is constructed according to the natural connection of time series information and gesture skeleton. A spatio-temporal convolutional network with multi-attention mechanism is used to learn spatio-temporal features to predict gestures. Finally, experiments are carried out on 14 types of gesture datasets in DHG-14 dynamic gesture dataset. Experimental results show that this method can recognize gestures accurately.

KEYWORDS

Convolutional, Gesture, Network, Reality, Recognition, Space-Time, Virtual

1 INTRODUCTION

With the development of related disciplines such as virtual reality and machine learning, the way people interact with computers is moving in a more natural and pervasive direction. There is an urgent need to use natural actions, rather than traditional dedicated input devices that send commands to control systems or interact with digital content in virtual environments. Human-computer interaction is changing from computer-centered to user-centered. Of all the body parts, the human hand plays an important role in interaction as a dexterous and effective executive organ. In daily life, people need to use their hands a lot to manipulate objects or communicate with others. The aim of gesture estimation is to recover the complete motion posture of hand in calculator system. Then, make the computer or other equipment can sense the spatial posture of the hand, so as to execute according to the instruction of the person. Accurate gesture estimation can not only build realistic virtual hand movements, but also enhance user experience in human-computer interaction (Chakraborty B K, Sarma D, Bhuyan M K, et al.2018). This helps computers better understand human behavior, which in turn makes interactions between humans and intelligent systems more intelligent.

As an important interactive way in computer graphics, virtual reality and human-computer interaction, gesture interaction provides a convenient, intuitive, simple and convenient interactive experience. Gesture interaction and recognition are of great significance to virtual reality interaction

DOI: 10.4018/JCIT.295246 *Corresponding Author

(De Smedt Q, Wannous H, Vandeborre J P.2016).3D motion sensing game (Duan H, Sun Y, Cheng W, et al.2021), assisted medical surgery (Gao X, Jin Y, Dou Q, et al.202) and other applications. However, due to the high degree of freedom of different gestures (Hussain S, Saxena R, Han X, et al.2017), the acquired gesture image data is usually characterized by low resolution, chaotic background, blocked hands, different finger shapes and sizes, and individual differences. This makes it difficult to accurately represent different gesture features, thus bringing difficulties and challenges to gesture recognition (Hou J, Wang G, Chen X, et al.2018).

Traditional gesture recognition is usually based on camera photos and 2D gestures for recognition and classification. Literature (Jiang D, Li G, Sun Y, et al.2019) analyzes the images of various resolutions in the image pyramid successively from low to high according to the changes in geometric dimensions of different parts of the hands obtained by segmentation. Literature (Li Y, He Z, Ye X, et al.2019)(Moin A, Zhou A, Rahimi A, et al.2021)proposed an effective Distance measure FEMD (finger-earth Mover's Distance) by using gesture shapes. This measure compares the shape differences of different gestures. Literature (Nasri N, Orts-Escolano S, Cazorla M.2020)proposed a gesture recognition method based on gesture main direction and Hausdorff-like distance template matching. This method has a high limitation on the main direction of gestures, which requires that the main direction of gestures obtained be consistent with the main direction of similar gestures in the training library, which limits the applicability of the method.

In recent years, many researchers have fully integrated data modeling and graph structure according to the characteristics of gesture sequence data, and proposed the idea of using Graph convolutional Networks (GCN)(Rudi F F Y, Yuniarno E M.) to predict actions. Because GCN can make full use of the spatial relationship of gestures, the performance of this method is greatly improved. However, a fixed topology is not the best choice for describing a diverse sample of actions, limiting the scope of messaging between nodes. Therefore, a graph structure that can be dynamically adjusted according to data samples is more suitable for modeling diverse gestures. In addition, the previous GCN ignored the importance of different channels. Often, the features produced by some channels are very important for motion recognition, while the features in some channels only play a minor role. In the process of feature extraction, we should pay more attention to those important channel features and ignore the unimportant channel information. In order to dynamically adjust the graph structure according to the data samples, a gesture recognition algorithm based on spatio-temporal graph convolution network is proposed in this paper.

2. THE PROPOSED MODEL IN THIS PAPER

2.1 Extraction of Dynamic Gesture Key Frames

2.1.1 Dynamic Gesture Data Preprocessing

A dynamic gesture is a sequence of gestures that change continuously over a period of time. The shape and position of the hand change over time. Dynamic gesture data sets are usually acquired by depth cameras or data gloves, and the acquired dynamic gestures usually have the problem of how to define the start frame and end frame. In the data set sequence used in this paper, participants were required to hold their entire hand fully open in front of the camera for a few seconds before each sequence began. This operation is mainly used to initialize the gesture estimation algorithm. Therefore, each gesture sequence has some gesture invalid frames independent of the gesture category. In order to avoid interference of invalid frame to gesture classification, it is necessary to delete invalid frame.

Gesture datasets usually need to be collected by different participants and keep gestures universal. However, different participants had different hand sizes and lengths between the joints. In order to eliminate the individual differences of hands, this paper normalized the joint length of hands to the same length, that is, changing the joint length without changing the Angle between the joints. For example, an abnormal movement of the fingertips across the plane of the palm may occur during a

clenched fist gesture. Literature (Sun Y, Weng Y, Luo B, et al.2020)normalized the joint length of hand into the average length of data set, but increased the amount of calculation. In this paper, the joint length of the hand is normalized based on the standard finger length.

Take a frame as an example to sketch the normalization process, using W; Represents the position of the x-th node in frame y. For convenience, the subscript y is omitted in the normalization process, that is, it is expressed as *M*. Among them, x = 0,,, 21. Vector is used to represent the joint pair formed by 22 joint nodes, i.e

$$V_x = \begin{cases} M_x - M_{x-1}, & 1 \leqslant x \leqslant 21 \text{ and } x \neq 6,10,14,18 \\ M_x - M_5, & x = 6,10,14,18 \end{cases} \tag{1}$$

The normalization process is

$$\bar{V}_x = L_x \frac{V_x}{V_x} \tag{2}$$

$$\bar{W}_x = \begin{cases} M_0, & x = 0 \\ \bar{V}_x + M_{x-1}, & 1 \leqslant x \leqslant 21 \text{ and } x \neq 6,10,14,18 \\ \bar{V}_x + M_5, & x = 6,10,14,18 \end{cases} \tag{3}$$

It should be pointed out that the normalization of the hand joint length in this paper is based on a standard finger length. Where, L is the corresponding standard length of joint segment *x*.

2.1.2 Dynamic Gesture Feature Representation

Based on the geometric characteristics of gesture and the Angle of time-space continuity, this paper proposes four characteristic representations of dynamic gesture. The process of dynamic gesture movement includes the global movement of the whole hand in space (i.e. translation and rotation of the hand in space) and the local movement of the fingers in the hand (i.e. translation and rotation of the fingers in the hand). The specific expressions are as follows:

(1) Hand translation movement in space. The movement process of hands in space is described by the distance of the center point of hands (node 1) in two frames before and after, i.e

$$T_y - T_{y-1} = \| \bar{W}_{1,y} - \bar{W}_{1,y-1} \| \tag{4}$$

(2) Hand rotation in space,

The hand flipping information in space is described by the distance of the main direction vector of the hand between the two frames. In this paper, the main direction of the hand is defined as $\bar{M}_1 - \bar{M}_0$, and the flip information is expressed as

$$P_y - P_{y-1} = \bar{M}_{1,y} - \bar{M}_{0,y} - \bar{M}_{1,y-1} - \bar{M}_{0,y-1} \tag{5}$$

(3) The translation movement of fingers inside the hand,

The translation movement of fingers is characterized by the relative distance of finger tips. In order to avoid the phenomenon of rotation error accumulation due to the rotation Angle information between joint segments as features, the distance between adjacent fingertips and the distance between fingertips and wrist were extracted as finger translation features. Specifically represented by the distance between adjacent finger tips

$$\begin{aligned} D_0 = \bar{M}_9 - \bar{M}_4, D_1 = \bar{M}_{13} - \bar{M}_9, \\ D_2 = \bar{M}_{17} - \bar{M}_{13}, D_3 = \bar{M}_{21} - \bar{M}_{17}, \end{aligned} \tag{6}$$

And the distance between the tip of the finger and the wrist

$$\begin{aligned} D_4 = \bar{M}_4 - \bar{M}_0, D_5 = \bar{M}_9 - \bar{M}_0, D_6 = \bar{M}_{13} - \bar{M}_0, \\ D_7 = \bar{M}_{17} - \bar{M}_0, D_8 = \bar{M}_{21} - \bar{M}_0. \end{aligned} \tag{7}$$

(4) The rotation movement of the fingers inside the hand.

The bending change of the hand are depicted by the rotation quaternion between the joints of the hand.

2.1.3 Dynamic Gesture Key Frame Extraction

In order to extract the key frame of dynamic gesture effectively, a gesture distance function is proposed by integrating four feature representations of gesture global motion and finger local motion. By sorting gesture distance, gesture frames with significant feature changes in dynamic gestures are selected as key frames. The frame generating motion mutation is used as the gesture key frame, and the distance between two frames before and after a dynamic gesture is defined as

$$\begin{aligned} L_y = \lambda_1 \sum_{x=0}^{13} \left(Q_{x,y} - Q_{x,y-1} \right) + \lambda_2 \sum_{k=0}^{8} \left(D_{k,y} - D_{k,y-1} \right) + \\ \lambda_3 \left(P_y - P_{y-1} \right) + \lambda_4 \left(T_y - T_{y-1} \right), y = S+1, \cdots, E \end{aligned} \tag{8}$$

The sequence number of the start frame of dynamic gesture is S, and the sequence number of the end frame is E.

In the segmental extraction of dynamic gesture key frames, it is assumed that the start frame of gesture is F_S and the end frame is F_E. The whole valid gesture can be expressed as { F_S, …, F_E }. If key frames of K frame are extracted, the whole gesture can be evenly divided into K segments, and the gesture segment I after segmentation is

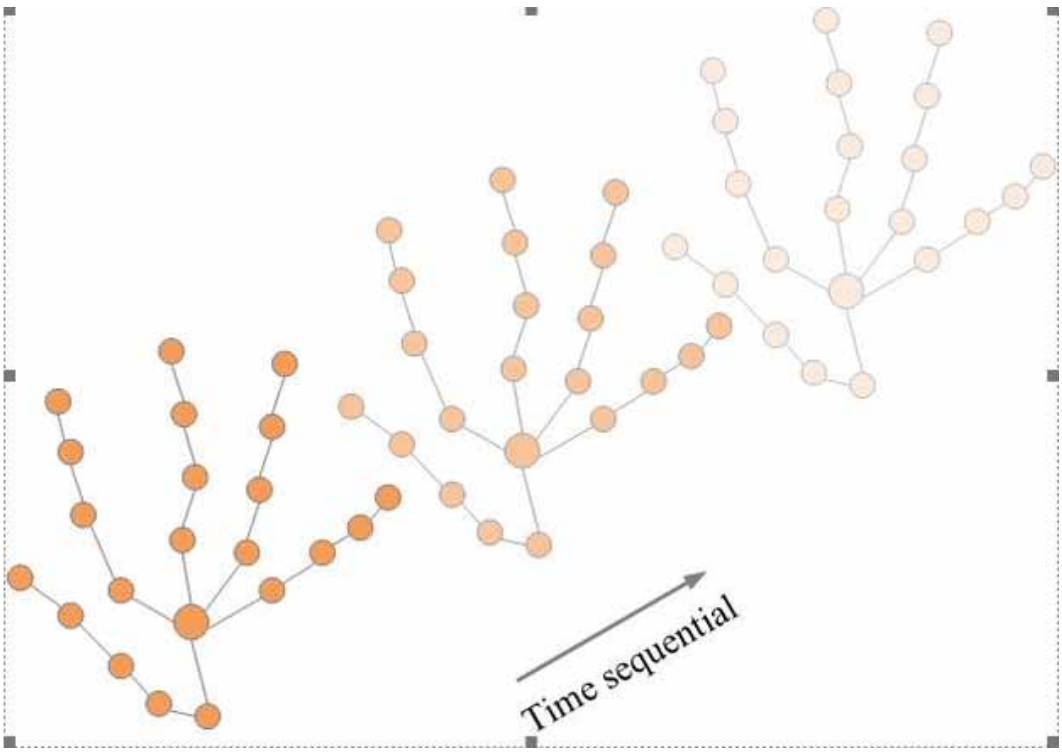
$$\boldsymbol{I} = \left\{ \left\{ F_S, \cdots, F_{S+d-1} \right\}, \cdots, \left\{ F_{S+(k-1)\cdot d}, \cdots, F_E \right\} \right\} \tag{9}$$

Among them, $d = \left(E - S + 1\right) / k$. Then, the frame with the maximum distance function (2) is selected as the key frame in each gesture segment.

2.2 ST-GCN Based on Multi-Attention Mechanism

This section first describes how the original ST-GCN (spatial-temporal graph convolutional networks) constructs graphs, and then describes how graph convolution operations are implemented in both spatial and temporal dimensions. The original gesture sequence is composed of a set of coordinate data, which can be represented by the two-dimensional or three-dimensional coordinates of all the human nodes in each frame. Considering that gesture itself is a topological structure, the temporal and spatial maps are used to model the key nodes in gesture sequence respectively in time and space dimensions. Specifically, for a gesture sequence containing N nodes and T frames, an undirected graph G=(V,E) is constructed on the gesture sequence. Among them, $V = \left\{v_{tx} \mid t = 1,2,\cdots,T, x = 1,2,\cdots,N\right\}$ represents the node set, it contains the sequence of all the properties of the node to node v. It is an eigenvector composed of the spatial coordinates of the point (x,y,z). The edge set E consists of parts E1 and E2. E represents the natural connection of gesture joints on the same frame. It's an intra-frame connection. E represents the connection of the same node on adjacent frames, which belongs to inter-frame connection. Figure 1 shows a constructed space-time diagram, in which blue dots represent gesture nodes. The blue lines represent the natural physical connections of the human body. The green line represents the time connection of the same node on adjacent frames.

Figure 1. The Space-time diagram



According to the defined graph G, the graph convolution operation is defined as follows on the spatial dimension. For vertex $v_{\tau x}$ on τ frame, it can be expressed as

$$g_{\text{out}}\left(v_{\tau x}\right)=\sum_{v_{\tau y}\in S_x}\frac{1}{T_{xy}}g_{\text{xn}}\left(v_{\tau y}\right)f\left(l_x\left(v_{\tau y}\right)\right) \tag{10}$$

Where, v represents vertices on the space-time graph. g stands for feature mapping. S_x is the sampling region of the convolution of the target vertex $v_{\tau x}$, and $v_{\tau x}$ here is the set. The weight function f is used to provide the weight vector. Since the number of neighbors of each node is different, the number in S is changing, while the number of weight vectors in f is constant. By transforming Eq. (10), the equation for realizing graph convolution in spatial dimension is

$$g_{\text{out}}=\sum_{k}^{K_v}f_k\left(g_{\text{xn}}\left(\widetilde{A_k}\odot M_k\right)\right) \tag{11}$$

Where, k represents the size of the convolution kernel. $\widetilde{A_k}$ is the normalized form of the adjacency matrix A. M is a learnable weight matrix. The sign $\odot$ stands for dot product. Equation (11) can be used to realize graph convolution operation in spatial dimension.

2.2.1 *Multi-Attention Mechanism*

This section mainly introduces the multi-attention mechanism introduced based on the original ST-GCN, including the proposed graph attention module and channel attention module respectively.

The convolution layer of spatial graph is modified and the graph attention module is introduced. This allows the model not only to learn the parameters of the network, but also to optimize the connected graph to get a graph structure more suitable for describing the action. In order to better predict the movement. Specifically, after adding the attention module of the human graph, the convolution equation of the spatial graph can be expressed as

$$g_{\text{out}}=\sum_{k}^{K_v}f_k\left(g_{\text{xn}}\left(A_k^{'}+B_k\right)\right) \tag{12}$$

Compared with Equation (2), it can be seen that the graph attention module includes two parts. $A_k^{'}$ is data-driven graph matrix. It can update the weight of the edge, so as to achieve the effect of replacing two matrices with one matrix. In addition, since A' is learned completely according to training data, the model can adapt to various action samples. In addition, A' is unique in different convolutional layers, so it is personalized and semantic in each layer.

The second part of this module is the graph attention matrix B, which can help the model to better model the actions for each sample and increase the personalization of the model. Specifically, for an input characteristic $g\left(v_{tx}\right)$. $g\left(v_{tx}\right)$ is divided into two convolution layers respectively. It maps to the vectors K and Q, namely

$$\begin{cases} K_{tx} = F_K g\left(v_{tx}\right) \\ Q_{tx} = F_Q g\left(v_{tx}\right) \end{cases} \tag{13}$$

Where, F_K and F_Q are the weight matrices corresponding to the two convolution layers respectively. Then, in order to limit the range of u to 0~1, it is normalized using the Softmax function, i.e

$$\alpha_{(t,x)\to(t,y)} = \frac{\exp\left(u_{(t,x)\to(t,y)}\right)}{\sum_{n=1}^{N} exp\left(u_{(t,x)\to(t,n)}\right)} \tag{14}$$

α is the similarity after normalization of the inner product u. It can be seen that α is also completely learned from different action samples, and it can effectively learn the weights of any two body nodes in different actions. This data-driven approach increases the flexibility and versatility of the model and enables the model to predict actions effectively in the face of diverse data.

By adding the graph attention module, the network can optimize the graph structure in the training process. It adapts to the changes of various samples and forms the most suitable topology to describe the action. Finally, the performance of the model is improved and the result of motion prediction is more accurate.

2.2.3 Structure of Multi-Attention ST-GCN

The network consists of nine basic units (B1~B9), as shown in Figure 2. The number of channels for data input is 3, the number of output channels for the first three basic units is 64, and the step size is 1. The number of output channels of the three basic units in the middle is 128. The last three units all have 256 output channels, and each base unit uses a residual mechanism. In cells B4 and B7, the step size is set to 2. Before sending skeleton data to the base unit, the data is normalized to make it more formal and processable, and this is done at the batch standardization layer. After 9 basic units, the output feature map is sent to the pooling layer for global average pooling to obtain a fixed size feature vector. At the end of the network is a Softmax classifier that classifies actions to predict the final outcome.

3 EXPERIMENT

3.1 Experimental Data Set

The data set adopted in this paper is DHG-14 dynamic gesture data set (Wei W, Wong Y, Du Y, et al.2019). This dataset contains 14 dynamic gesture categories, as shown in Table 1, and gestures are performed in two ways. The one-finger way and the whole hand way. Each gesture was completed by 20 participants in the above two ways, 5 times in each execution mode, a total of 2800 dynamic gesture sequences. Among the 14 gestures, 5 are Fine gestures and 9 are Coarse gestures. Meanwhile, the dataset contains not only the depth image of dynamic gesture video frame, but also the coordinate of 22 hand joints in 2D depth image and 3D space. The resolution of the depth image was 640×480, and the depth map and the hand skeleton were captured at the speed of 30 frames /s.

3.2 Determination of Frame Number by Gesture Key

In dynamic gesture key frame extraction, the number of gesture key frames must be determined first. Selecting the appropriate number of gesture key frames will affect the recognition accuracy of

Figure 2. The structure of multi-attention ST-GCN

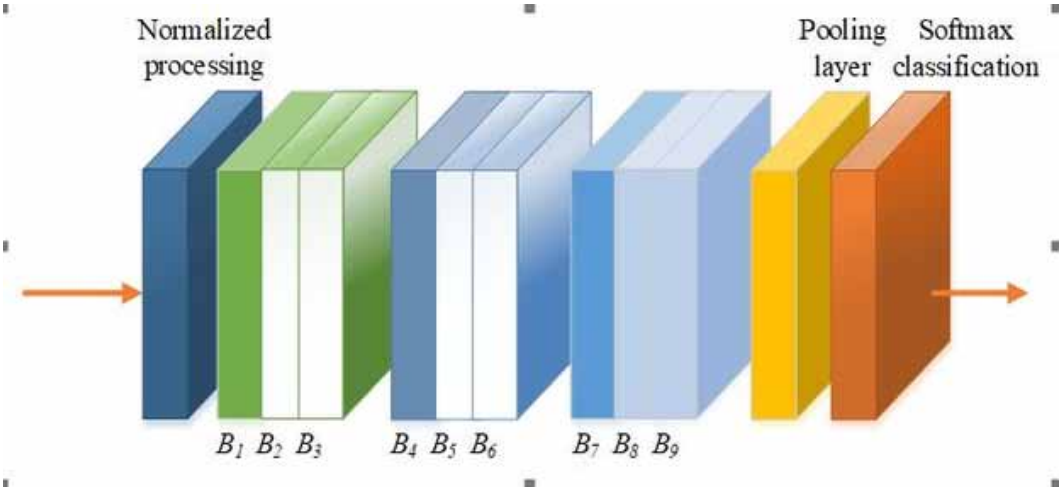


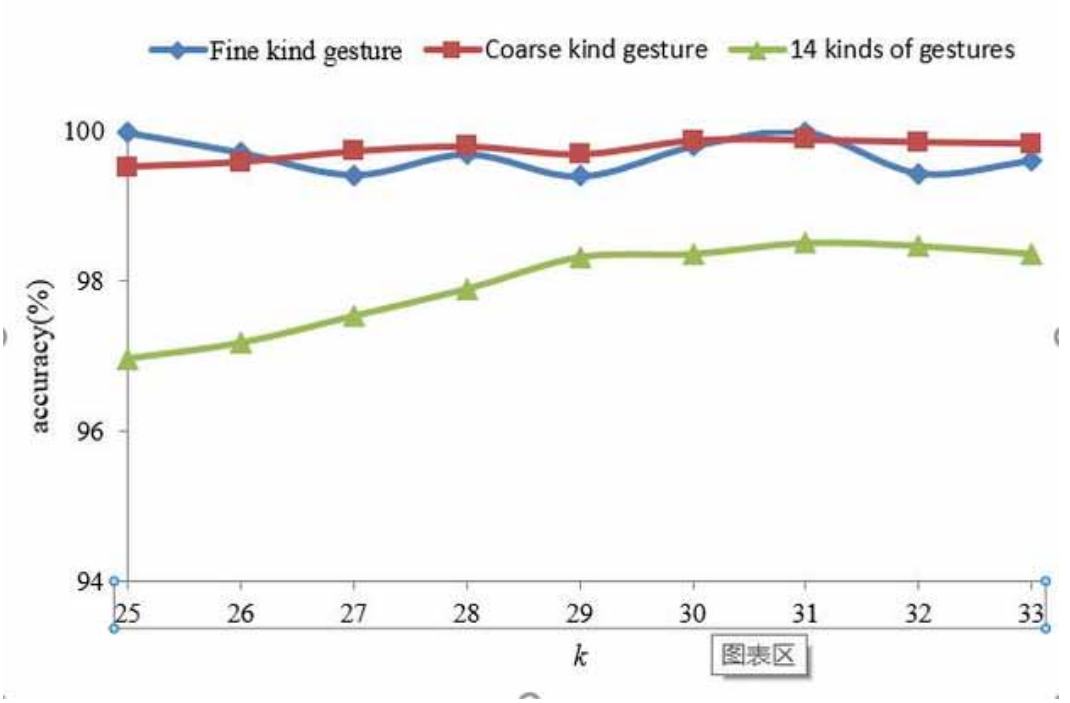
Table 1. Gesture categories in the dataset

Number	Gestures	Category
1	Grab	Fine
2	Expand	Fine
3	Pinch	Fine
4	Rotation CW	Fine
5	Rotation CCW	Fine
6	Tap	Coarse
7	Swipe right	Coarse
8	Swipe left	Coarse
9	Swipe up	Coarse
10	Swipe down	Coarse
11	Swipe X	Coarse
12	Swipe V	Coarse
13	Swipe +	Coarse
14	Shake	Coarse

gesture. This paper compares the gesture recognition accuracy and analyzes the k value of different key frames. As shown in Figure 3, with the increase of the number of key frames, gesture recognition accuracy increases and tends to be stable. The accuracy of gesture recognition tends to decrease when the number of frames is larger than 31. The experiment shows that the key frame selection of the same gesture may differ greatly if the number of key frames is small. This results in low accuracy of gesture recognition. Therefore, in order to improve gesture recognition accuracy, a relatively large number of gesture key frames should be reserved to avoid the above problems. At the same time, even

if some key frames of the same gesture differ greatly, there are still enough remaining key frames to shorten the distance between the same type of gesture. This serves to widen the gap between different gesture categories. In this experiment, k=31 key frames were selected for each dynamic gesture video. According to the key frame, the global motion of hand in space and the local motion feature of fingers inside the hand are extracted, and the dynamic gesture recognition and classification are realized based on gesture feature representation.

Figure 3. Gesture recognition accuracy



3.3 Comparison with Different Identification Methods

To illustrate the effectiveness of the proposed method, it is compared with 4 existing gesture recognition methods (Wang Y, Jung C, Yun I, et al.2019),(Wang K, Zhao R, Ji Q.2018),(Zhu Ji-Yu, Wang Xi-Ying, Wang Wei-Xin, et al.2006),(Mu Zhou, Yanmeng Wang, Zengshan Tian,et al.2019). The same experimental setup was followed in the comparison, and the leave-one-method cross-validation was adopted for the experiment. 19 subjects were trained on their dynamic gesture data, and recognition tests were performed on the remaining 1 subject's gesture data. The experiment was repeated 20 times using gesture data of different test subjects. In this paper, 5 kinds of gestures are verified by experiments.

For 14 gestures, the proposed method is compared with other dynamic gesture recognition methods. As can be seen from column 3 of Table 2, the highest recognition accuracy of existing methods is 86.76%. On this basis, the proposed method improves 12.75% and the gesture recognition accuracy reaches 99.51%. At the same time, experiments are carried out on Fine data and Coarse data respectively. This indicates that the proposed method pays more attention to hand detail changes

Table 2. Comparison of recognition accuracy of 14 gestures (%)

Method	Category		
	Fine	**Coarse**	**Both**
Paper(Wang Y, Jung C, Yun I, et al.2019)	74.61	89.31	84.08
Paper(Wang K, Zhao R, Ji Q.2018)	79.01	90.81	86.61
Paper(Zhu Ji-Yu, Wang Xi-Ying, Wang Wei-Xin, et al.2006)	77.01	91.73	86.47
Paper(Mu Zhou, Yanmeng Wang, Zengshan Tian,et al.2019)	76.61	92.40	86.76
Proposed	99.98	99.70	99.51

in dynamic gesture feature representation and extraction. The gesture feature integrates the global motion feature of hand in space and the local motion feature of fingers inside the hand. In addition, in the case of gesture data synthesis, the recognition effect of the proposed method is lower than that of Fine or Coarse data alone. This is because it is essentially a convex optimization problem.

3 CONCLUSION

In the traditional gesture recognition algorithm, the high degree of freedom of different gestures and the low resolution of gesture image, messy background, blocked hands, different shapes and sizes of fingers, individual differences and other problems lead to the low accuracy of gesture recognition. To solve this problem, this paper proposes a gesture recognition algorithm based on spatiotemporal graph convolution network. In order to remove redundant frames and complete gesture frame data, gesture key frames were extracted from continuous video frames. Then, the gesture data are fed into the spatio-temporal convolutional neural network to extract the spatio-temporal features and predict the gesture actions. Finally, the experimental results of dynamic gesture data set are used to verify the effectiveness of the proposed algorithm.

REFERENCES

Chakraborty, B. K., Sarma, D., Bhuyan, M. K., & MacDorman, K. F. (2018). Review of constraints on vision-based gesture recognition for human–computer interaction. *IET Computer Vision*, *12*(1), 3–15. doi:10.1049/iet-cvi.2017.0052

De Smedt, Q., Wannous, H., & Vandeborre, J. P. (2016). Skeleton-based dynamic hand gesture recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 1-9.

Duan, H., Sun, Y., Cheng, W., Jiang, D., Yun, J., Liu, Y., Liu, Y., & Zhou, D. (2021). Gesture recognition based on multi-modal feature weight. *Concurrency and Computation*, *33*(5), e5991. doi:10.1002/cpe.5991

Gao, X., Jin, Y., & Dou, Q. (2020). Automatic gesture recognition in robot-assisted surgery with reinforcement learning and tree search. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE.

Hou, J., Wang, G., & Chen, X. (2018). Spatial-temporal attention res-TCN for skeleton-based dynamic hand gesture recognition. *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*.

Hussain, S., Saxena, R., & Han, X. (2017). Hand gesture recognition using deep learning. In 2017 International SoC Design Conference (ISOCC). IEEE.

Jiang, D., Li, G., Sun, Y., Kong, J., & Tao, B. (2019). Gesture recognition based on skeletonization algorithm and CNN with ASL database. *Multimedia Tools and Applications*, *78*(21), 29953–29970. doi:10.1007/s11042-018-6748-0

Li, Y., He, Z., Ye, X., He, Z., & Han, K. (2019). Spatial temporal graph convolutional networks for skeleton-based dynamic hand gesture recognition. *EURASIP Journal on Image and Video Processing*, *2019*(1), 1–7. doi:10.1186/s13640-019-0476-x

Moin, A., Zhou, A., Rahimi, A., Menon, A., Benatti, S., Alexandrov, G., Tamakloe, S., Ting, J., Yamamoto, N., Khan, Y., Burghardt, F., Benini, L., Arias, A. C., & Rabaey, J. M. (2021). A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition. *Nature Electronics*, *4*(1), 54–63. doi:10.1038/s41928-020-00510-8

Nasri, N., Orts-Escolano, S., & Cazorla, M. (2020). An semg-controlled 3d game for rehabilitation therapies: Real-time time hand gesture recognition using deep learning techniques. *Sensors (Basel)*, *20*(22), 6451. doi:10.3390/s20226451 PMID:33198083

Rudi, F. F. Y., & Yuniarno, E. M. (2019). Contour to Centroid Distance Graph as Feature in Hand Gesture Recognition. *IOP Conference Series: Materials Science and Engineering, 536*(1), 012150.

Sun, Y., Weng, Y., Luo, B., Li, G., Tao, B., Jiang, D., & Chen, D. (2020). Gesture recognition algorithm based on multi-scale feature fusion in RGB-D images. *IET Image Processing*, *14*(15), 3662–3668. doi:10.1049/iet-ipr.2020.0148

Wang, K., Zhao, R., & Ji, Q. (2018). Human computer interaction with head pose, eye gaze and body gestures. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018).* IEEE.

Wang, Y., Jung, C., & Yun, I. (2019). SPFEMD: Super-pixel based finger earth mover's distance for hand gesture recognition. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE.

Wei, W., Wong, Y., Du, Y., Hu, Y., Kankanhalli, M., & Geng, W. (2019). A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface. *Pattern Recognition Letters*, *119*, 131–138. doi:10.1016/j.patrec.2017.12.005

Zhou, M., Wang, Y., Tian, Z., Lian, Y., Wang, Y., & Wang, B. (2019). Calibrated data simplification for energy-efficient location sensing in internet of things. *IEEE Internet of Things Journal*, *6*(4), 6125–6133. doi:10.1109/JIOT.2018.2869671

Zhu, J.-Y., Wang, X.-Y., & Wang, W.-X. (2006). Hand gesture recognition based on structure analysis. *ChineseJournal of Computers*, *29*(12), 2130–2137.

Liao Ting obtained a Bachelor of Engineering in Computer Science and Technology from Sichuan Normal University in 2001 and a Master of Education Technology from Sichuan Normal University in 2007. She is now working in the school of intelligent manufacturing of Sichuan University of Arts and Sciences. She is a professional teacher and associate professor. Her main research interests are virtual reality technology and computer graphics and image processing technology.