

Research Article

Study on Volleyball-Movement Pose Recognition Based on Joint Point Sequence

Xi Li 

Physical Education Department, Taihu University of Wuxi, Wuxi 214000, Jiangsu, China

Correspondence should be addressed to Xi Li; 000058@wxu.edu.cn

Received 22 July 2022; Revised 15 August 2022; Accepted 20 August 2022; Published 17 February 2023

Academic Editor: N. Rajesh

Copyright © 2023 Xi Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the high-speed operation of society and the increasing development of modern science, people's quality of life continues to improve. Contemporary people are increasingly concerned about their quality of life, pay attention to body management, and strengthen physical exercise. Volleyball is a sport that is loved by many people. Studying volleyball postures and recognizing and detecting them can provide theoretical guidance and suggestions for people. Besides, when it is applied to competitions, it can also help the judges to make fair and reasonable decisions. At present, pose recognition in ball sports is challenging in action complexity and research data. Meanwhile, the research also has an important application value. Therefore, this article studies human volleyball pose recognition by combining the analysis and summary of the existing human pose recognition studies based on joint point sequences and long short-term memory (LSTM). This article proposes a data preprocessing method based on the angle and relative distance feature enhancement and a ball-motion pose recognition model based on LSTM-Attention. The experimental results show that the data preprocessing method proposed here can further improve the accuracy of gesture recognition. For example, the joint point coordinate information of the coordinate system transformation significantly improves the recognition accuracy of the five ball-motion poses by at least 0.01. In addition, it is concluded that the LSTM-attention recognition model is not only scientific in structure design but also has considerable competitiveness in gesture recognition performance.

1. Introduction

The posture of the human body is one of the important biological characteristics of the human body. It has many application scenarios, such as gait analysis, video surveillance, augmented reality, human-computer interaction, finance, mobile payment, entertainment and games, and sports science. Gesture recognition allows computers to know what a person is doing and who they are. Especially in the field of monitoring, it is a good solution when the resolution of the face image obtained by the camera is too small. It can also be used as an important auxiliary verification method in the target identification system to reduce the effect of misidentification. Human gesture recognition includes action recognition and identity recognition, and the key lies in human feature extraction. The human body feature extraction mainly completes action feature

extraction and identity feature extraction. In volleyball, detecting and identifying relevant poses in motion sequences can not only provide coaches and players with data-based guidance and advice but also help referees make fair and reasonable decisions in various games. However, most of the existing human gesture recognition methods are aimed at the recognition of daily simple actions [1]. Data collection is simple and further research is needed.

Most of the traditional human pose recognition research is done on sequences of video frames [2]. Although some research results have been achieved, it is difficult to break through the bottleneck of human pose recognition research using video due to changes in light intensity, interference from complex backgrounds, and the self-occlusion of target users. In the recent years, with the rapid development of video capture technology, such as Kinect, researchers can easily obtain coordinated information on the image, depth

image, and skeletal joint points [3, 4]. The information provided by depth images [5] can reflect the three-dimensional structural information and the geometry of target objects well compared with the images. Moreover, it has strong robustness to the influence of factors such as light intensity and scale changes. The posture of ball sports is more complex compared with the simple daily posture of the human body, and it also has requirements for research data. Besides, the existing recognition methods cannot effectively judge the ball movement gestures due to the change in the difficulty of gesture recognition. Therefore, further research on pose data and recognition methods is needed to improve the accuracy of ball-motion pose prediction classes.

In this article, volleyball is represented. The problem of volleyball-gesture recognition is studied combined with the analysis and summary of the existing research on human gesture recognition based on the joint point sequence and the long short-term memory (LSTM) network [6]. In addition, this article proposes a data preprocessing method based on the angle and relative distance feature enhancement and a volleyball-motion pose recognition model based on LSTM-attention. Experimental results show that the data preprocessing method reported here can further improve the accuracy of gesture recognition. Besides, the LSTM-attention recognition model is not only scientific in structure design but also has considerable competitiveness in gesture recognition performance, which can provide a basis for the research on gesture recognition in volleyball. Although the LSTM-attention recognition model is not only scientifically designed in structure but also quite competitive in gesture recognition performance, this approach is not yet universal and should be refined in future studies.

2. Materials and Methods

2.1. Recurrent Neural Network (RNN). The RNN is obtained by simulating the human neural transmission system [7]. When humans are thinking, such as reading an article, they may not be able to understand the meaning of the article only by relying on the information currently read. It is often necessary to combine the previous content to understand the essence of the article. This shows that humans are not in a blank brain when they think about problems. The brain will not discard the content of the articles that have been read before but will understand and analyse based on the previous readings. This reveals that human thinking is a continuous process [8]. However, traditional neural networks cannot achieve this, so there is an RNN. The RNN is a neural network with a short-term memory and continuously transmits information by adding loops to the network. It is suitable for processing sequential data. Figure 1 shows the general structure of an RNN. In Figure 1, h_t is the value of the hidden layer at time t , x_t is the input of the network at time t , and h_t is its output.

Each block of the neural network in Figure 1 is represented by A. Figure 2 shows the extended structure of the RNN, which is also the internal structure of the RNN transformed according to the time dimension [9]. In the RNN, there is a signal transfer between all the hidden layer

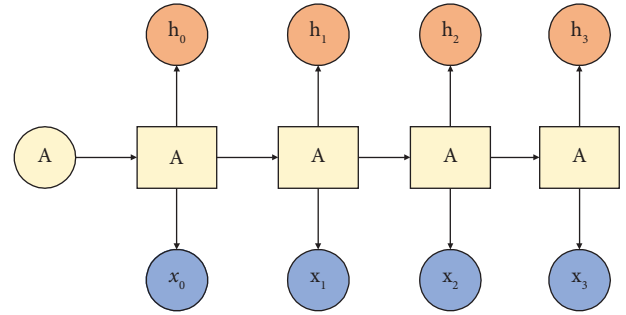


FIGURE 1: RNN structure diagram.

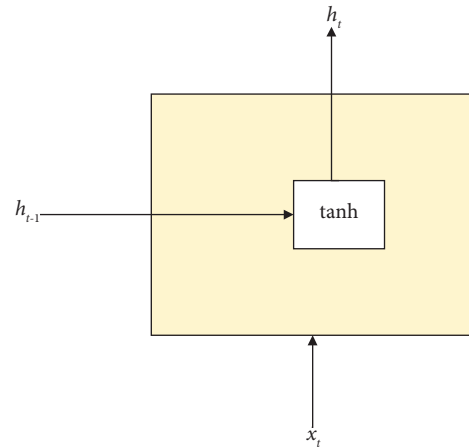


FIGURE 2: RNN internal structure diagram.

nodes [10]. The output of the hidden layer of the RNN at time t is fed back to the RNN at time $t + 1$. Then, the input of the RNN at time $t + 1$ and the output of the RNN at time t will act on the output of the RNN at time $t + 1$. The chain structure of the RNN essentially determines its strong processing capability for the sequence data. In the recent years, the RNN has performed very well in language modeling, speech recognition, image captioning, and machine translation.

The number of units in the input layer of a neural network is fixed, so inputs of variable length must be processed in a loop or recursion. RNN implements the former. It works by dividing an input of variable length into small chunks of equal length. Then, they are sequentially input into the network, realizing the processing of the variable length input by the neural network. The RNN can encode a tree/graph structure information as a vector, mapping the information into a semantic vector space. This semantic vector space satisfies a class of properties. For example, semantically similar vectors are close together. However, a big limitation of the RNN is the vanishing gradient problem [11]. The RNN is a short-term memory neural network that can only memorize short-distance information sequences. When the time interval becomes large, the RNN will gradually lose its ability to learn the information of the previous time nodes, which will make the learning of the RNN very difficult. This problem is also known as the “long sentence dependency problem.”

2.2. *LSTM*. The researchers make related improvements to its basic structure to solve the “long sentence dependency problem” of the RNN. The researchers design the cell state internally to record the historical state information [12] and introduce the gating unit to control the node information of the hidden layer. This variant of the RNN can solve the above problems very well. It can memorize long-term information related to the current recognition task. This variant is called the LSTM network.

LSTM can be regarded as a special type of the RNN [13], which can greatly enhance the network’s ability to store information within long time intervals. For LSTM and the RNN, the same is that they are both chain structures, and the difference is the structure inside their network. LSTM is the most effective sequence model in deep learning (DL) [14], which mainly consists of the forget gate, input gate, and output gate. The RNN model has the problem of missing gradients. LSTM effectively avoids this drawback and proposes a new cell structure, which can judge the retention or forgetting of data. Real-time data are processed from the far left to the far right [15]. Also, the data are processed from the input. Therefore, it is necessary to judge which information continues to run and which is abandoned in the endless input information. This process follows a switch control, which is $f(t)$.

The control function is as follows:

$$f^{(t)} = \sigma(w_f[h^{(t-1)}, x^t] + b_f). \quad (1)$$

In equation (1), w_f and b_f are the weight and bias of the forget gate, respectively. The previous information is input into the input gate. The task at this layer is to decide which information needs to be updated and how much to update. σ represents the activation function, b_f represents the bias value, $h^{(t-1)}$ represents the short-term memory, and x^t represents the current input.

$$i^t = \sigma(w_i[h^{(t-1)}, x^t] + b_i), \quad (2)$$

$$c^t = \sigma(w_c[h^{(t-1)}, x^t] + b_c), \quad (3)$$

$$C^t = i^t * c^t + f^{(t)} * C^{t-1}. \quad (4)$$

In equations (2)–(4), w_i and w_c represent the corresponding weights, b_i and b_c represent the corresponding biases, and C^t represents the current cell state value. After the screening of the first two gates is completed, the output gate determines which information needs to be output. There is a switch to control the output in the output gate.

$$o^t = \sigma(w_o[h^{(t-1)}, x^t] + b_o), \quad (5)$$

$$h^t = o^t * \tanh^{-1}(c^t). \quad (6)$$

In equations (5) and (6), w_o and b_o represent the weight and bias of the output gate. o^t represents the output gating unit, h^t is the output value of the current unit, and σ represents the activation function.

2.3. *Volleyball-Movement Pose Recognition Method Based on LSTM-Attention*. This section constructs a volleyball-movement posture recognition method based on LSTM-Attention to help LSTM effectively extract the feature information before and after the action [16]. This section first gives a brief overview of the model and analyses the various modules involved. The experimental results demonstrate the effectiveness of the LSTM-Attention method proposed here in ball-motion pose recognition. In the recent years, the research on human gesture recognition has gradually replaced the status of traditional methods with the continuous development of DL-related technologies. In human gesture recognition, although the RNN-based gesture recognition method has obvious advantages in short-term memory, it has great difficulties in dealing with some recognition scenarios that require long-term memory. As a special type of the RNN, LSTM can not only solve the problem of the disappearance of the RNN gradient but also enhance the network’s ability to memorize information for long time intervals. The latter is favored by many researchers compared with the former because it is good at obtaining feature information between long-term sequences. Many networks optimized based on this have been produced with the in-depth study of the LSTM neural network. They are widely used in text, speech, and image recognition. In the research on human posture recognition, different human skeletal structures will produce great differences when they are playing volleyball. This can lead to indistinguishable target users in the same ball game pose. For this problem, this article proposes a feature enhancement preprocessing method based on the angle and the relative distance. Besides, contextual information between action sequences plays a crucial role in gesture recognition. This section constructs a ball-motion pose recognition method based on the LSTM-Attention model to effectively extract the long-sequence feature information of ball motion poses. The overall process of ball sports gesture recognition is shown in Figure 3.

First, Kinect is used to acquire the joint point coordinate data of the human skeleton performing ball motion [17]. Then, the scale-invariant angle and relative distance features are extracted from the joint point information. Finally, the LSTM-attention network is used to mine the deep timing information in the skeleton sequence of the human body when ball sports are performed. Furthermore, this information is combined with spatial features to recognize the ball motion pose of the human body [18]. The model learns the correlation between the time series data of ball motion poses autonomously and effectively by combining LSTM with the attention mechanism to improve the accuracy of the model pose recognition.

2.4. *Angle Feature Extraction*. In the process of the human body performing ball sports, the features selected based on joint point information should have general behavior [19]. Features do not vary greatly due to differences in the human skeletal structure, and they do not shift because the target user and the Kinect depth sensor are in different

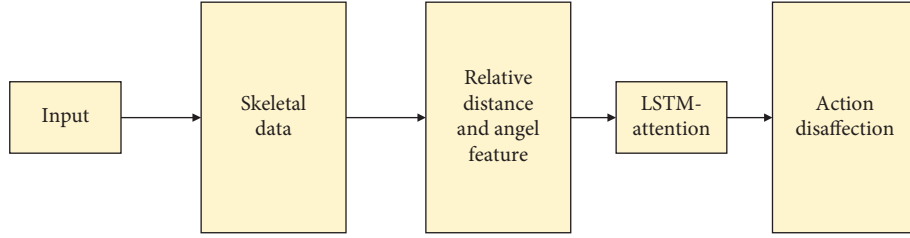


FIGURE 3: Flowchart of volleyball gesture recognition.

positions. Therefore, the angle features extracted based on the joint point information are used to predict the pose category of the human body during ball sports [20]. When the human body performs different ball sports, there will be different angular relationships between the joint points of the human bones in space. Especially for the joint points on the arm, the included angle between the joint points involved in the corresponding action of the ball sports will have a relatively fixed range of variation. This can intuitively describe the ball game posture. Therefore, the coordinate information of the eight joint points is decomposed into five parts according to the human body structure, including the trunk and the limbs. Then, the angle feature extraction is performed on the joint-point coordinate information of these five parts.

When the angle of the joint point information of the human skeleton is calculated, the limbs of each part in the human skeleton model need to be regarded as a vector. The correspondence between the angles between the joint points and each component vector is shown in Table 1. r_{ij} represents the vector that forms the angle of a joint point.

In Figure 4, the right arm model of the human body is taken as an example [21], and the joint angle r_5 consists of two vectors, $r_{4,5}$ and $r_{5,6}$, respectively.

2.5. Relative Distance Feature Extraction. The relative distance feature between the human skeleton joint points is a kind of information that can express different ball-motion pose data in the spatial dimension [22]. When the human body performs a ball action, the spatial position information of the joint points of each part of the human skeleton will also change. Moreover, the relative distance between some joint points that change with the movement will also form a change rule for an action posture [23]. For example, when people perform a badminton swing, the relative distance between the user's hand and the base of the spine is a very expressive information. Therefore, the relative distance feature extracted based on joint point information is used here to analyse the pose category of the human body when they perform ball sports.

It is found that the joint points at the base of the spine have stability in the process of the human body movement in expressing the human body ball sports posture through analysis of the joint point data when the human body performs ball sports posture. Therefore, this article regards the joint point at the base of the spine as the center point and

TABLE 1: Correspondence between the angle between the joint points and each component vector.

Joint angle number	Composition vector
1	$r_{1,2}, r_{1,3}$
2	$r_{2,1}, r_{2,3}$
3	$r_{3,1}, r_{3,2}$
4	$r_{4,5}, r_{4,6}$
5	$r_{5,4}, r_{5,6}$
6	$r_{6,4}, r_{6,5}$
7	$r_{7,8}, r_{7,9}$
8	$r_{8,7}, r_{8,9}$

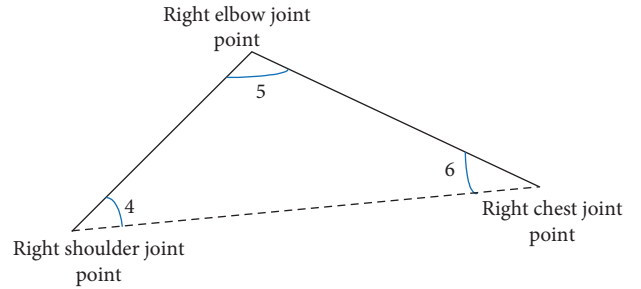


FIGURE 4: Schematic diagram of the characteristics of the right arm.

analyses the ball sports posture based on the relative distance between the center point and other joint points.

2.6. Ball-Movement Pose Recognition Method Based on LSTM-Attention. Here, Kinect is used to collect the joint point data of the human body during ball sports [24]. Based on this, two kinds of geometric features are artificially designed to describe the pose of the ball. The appropriate relevant parameters in the model are obtained after the LSTM-attention network model constructed here is trained on the ball sports training set. Besides, the recognition and classification of ball sports poses are carried out on the test set. Its network structure is demonstrated in Figure 5.

In this model, a multilevel LSTM structure is designed to improve the learning ability of the recognition network [25] to handle complex feature representations in ball sports poses. The number of network layers of LSTM is designed to be three layers. An attention mechanism is added to the model. This design enables the feature vector to spontaneously perceive the network weights that significantly impact the recognition results of ball motion gestures. Some important feature information gets attention. This can also

perform further feature enhancement on the feature data extracted from the previous network layers. In addition, a dropout layer is added between the LSTM structures. This can reduce the occurrence of overfitting of the gesture recognition model when the number of experimental samples is limited. To sum up, feature learning through the multilevel LSTM network is combined with feature enhancement of the attention mechanism. This enables the network model to fully and effectively learn the correlation between the time series data in the ball motion poses, thereby improving the pose recognition ability of the entire network model.

Figure 5 displays a diagram of the LSTM-attention network model that is expanded according to time. The model can form an action sequence according to the time sequence of the features of the human body during ball sports, and it is used as the input of the ball-sports pose recognition network. The input includes an angle feature and a relative distance feature. The ball motion features of each human body have become a 38-dimensional data through the previously mentioned joint evaluation, repair, and feature extraction. Also, the length of each action sequence will be affected by the different frame numbers of different actions. It is indispensable to perform isometric operations on the data in the dataset before the feature data are input into the gesture recognition network model. According to the longest frame value in each ball motion sequence, the other motion sequences are zeroed.

The multidimensional feature sequence is input into the LSTM-attention network, which is processed by LSTM, dropout, and the attention mechanism. The intermediate value is sent to the output layer. The function used in this layer is the softmax function. The function can judge the corresponding ball sports posture and output the probability value of five different ball-sports posture labels. Finally, the maximum value of the probability evaluation values is selected as the output category of the final ball-motion pose.

When the human body performs different ball motion poses, all the joint motion data contained in the human skeleton are not equally important. For example, the changes in the joint point data of the human bones are mainly concentrated in the right arm part in the process of the human body completing the badminton swing. The joint point data of other parts of the body have little effect on the final gesture recognition effect. Therefore, the attention mechanism is introduced into the improved human ball-motion pose recognition model. In the process of movement, the important data of human limbs and joints can be marked and much attention can be given to them. In the research data here, the coordinate data of fifteen joint points of the human body during ball sports are collected, and the human skeleton model is established based on this. In most cases, the joint point information that a ball motion pose can be associated with is fixed. These fixed-joint point information will be converted into feature vectors through LSTM. The essence of the attention mechanism is to perform a weighted summation of these feature vectors to find out the

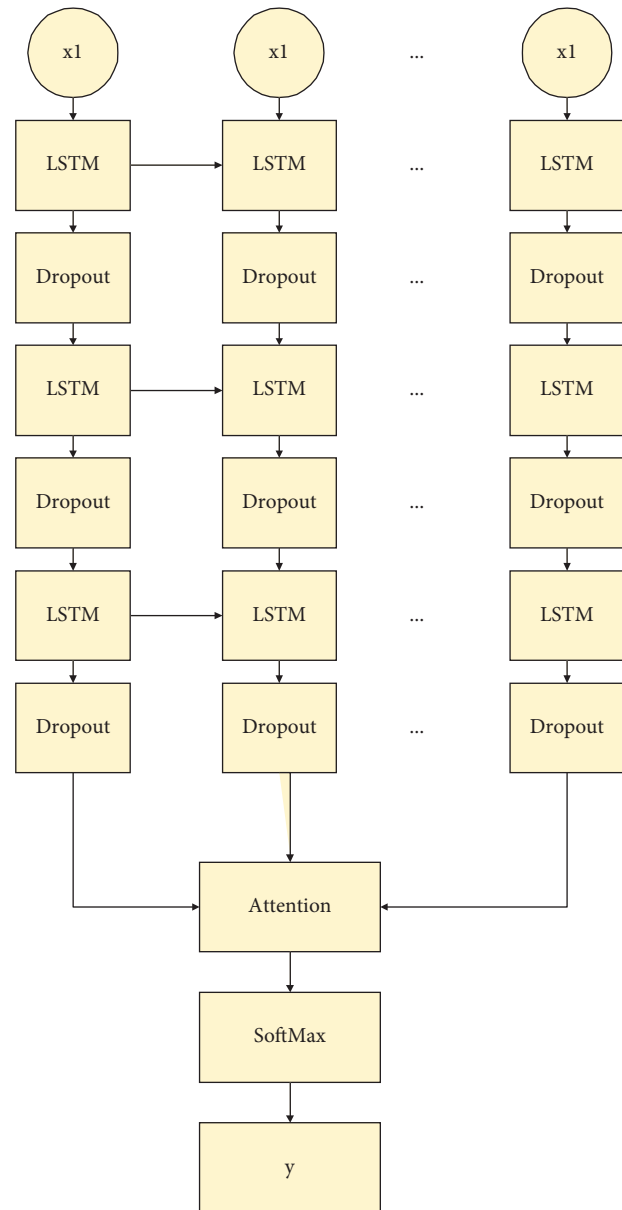


FIGURE 5: Network structure of volleyball pose recognition.

joint point information that importantly impacts the recognition of ball motion poses.

3. Results and Discussion

The comparative experiments before and after the coordinate system transformation of joint point information verify the improvement effect of the joint-point preprocessing method based on the coordinate system transformation on the accuracy of gesture recognition. This section conducts experiments. First, this article conducts experiments on the joint point data before and after the coordinate system transformation through the traditional LSTM-pose recognition network model. MATLAB software is used to simulate and simulate the gesture recognition process of the network model. The results are shown in Figure 6.

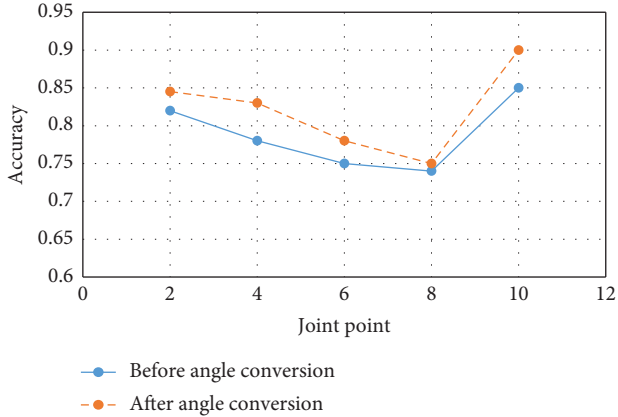


FIGURE 6: LSTM prediction results before and after angle conversion.

Figure 6 shows the comparative experimental results before and after the coordinate information conversion preprocessing of the joint point data through the LSTM neural network. The experimental results indicate that the joint-point coordinate information of the coordinate system transformation significantly improves the recognition accuracy of the five ball motion poses by at least 0.01. However, the improvement in gesture recognition accuracy for smash actions is lower compared to the improvements in gesture recognition accuracy for serve, lift, high clear, and backhand. This also means that gesture recognition accuracy for actions like smash is less affected by angular changes during data collection than other actions.

In addition to LSTM, this section also conducts comparative experiments on gesture recognition methods such as BiLSTM, linear SVM, and multilayer LSTM. The experimental data and settings remain the same as in the above LSTM experiments. BiLSTM is the abbreviation of bidirectional long short-term memory, which means a bidirectional long and short-term neural network. It is composed of forward LSTM and backward LSTM. Both are often used to model contextual information in natural language processing tasks. Figure 7 shows the overall prediction results of each gesture recognition method for the joint-point coordinate information before and after the coordinate system conversion.

The experimental results show that the joint-point coordinate information after the coordinate system transformation can effectively improve the overall accuracy of the ball-motion gesture recognition network model. After converting the coordinate system, the accuracy is improved by at least 0.01, and the accuracy of the method proposed here is higher than the other methods. The optimal hierarchical experiment of the LSTM multilayer structure is expected to integrate the feature information of the long-term sequence on a global scale and realize the high-level abstraction of the input human skeleton joint point data. Therefore, this article constructs a multilevel LSTM structure based on a classification model. However, if the number of layers in the LSTM multilayer structure is large, the model will take a lot of time to converge, which will complicate the

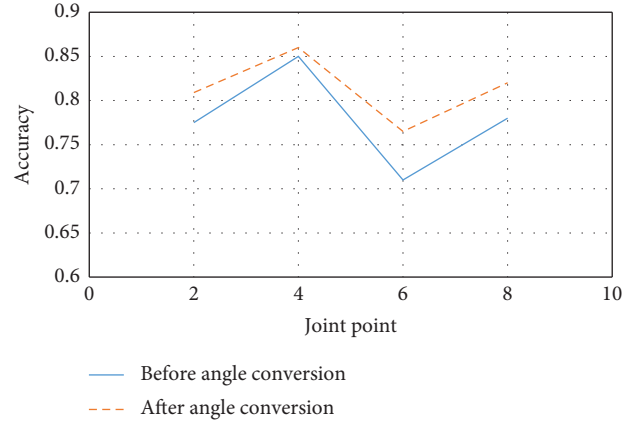


FIGURE 7: Overall prediction results before and after angle conversion.

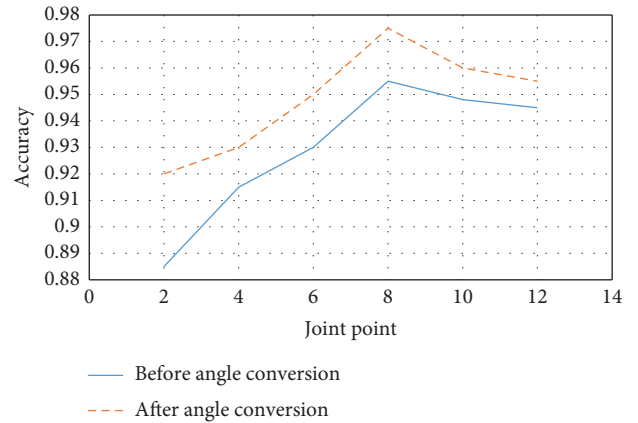


FIGURE 8: Comparative experiments of different LSTM layer structures in action recognition.

model. Therefore, this article conducts comparative experiments on different layers of LSTM structures on the BadmintonData and MSRAction3D datasets to verify the scientificity and effectiveness of the ball-motion gesture recognition model based on the three-layer LSTM structure designed here. The BadmintonData dataset contains various volleyball poses. The MSRAction3D dataset records 20 actions and ten subjects. Each subject performs each action two to three times. There are a total of 567 depth map sequences with a resolution of $640 * 240$. Data are recorded with a depth sensor similar to the Kinect unit.

The number of layers of the LSTM multilayer structure is set as one, two, three, four, and five, respectively. LSTM-Attention_{*n*} is a model that fuses the *n*-layer LSTM structure and the attention mechanism, respectively. Besides, LSTM is a model that only contains a single-layer LSTM structure without an attention mechanism. The results are revealed in Figure 8.

The purpose of the comparative experiment before and after the restoration of joint point information is to verify the improvement effect of the joint-point processing method based on the bone length and motion continuity on the performance of the pose recognition method. Its essence is

TABLE 2: Comparative experiments before and after the data recovery of joint points.

Experimental data	LSTM _	B I-LSTM	Linear SVM	Multilayer LSTM
Badminton data	0.81	0.87	0.76	0.82
BadmintonData_Repairing	0.83	0.88	0.83	0.85
MSRAAction3D_	0.90	0.95	0.92	0.95
MSRAAction3D_Repairing	0.92	0.96	0.93	0.95

to evaluate the influence of the joint point data with errors on the accuracy of human pose recognition.

This section uses the abovementioned gesture recognition methods to conduct comparative experiments on the BadmintonData and MSRAAction3D datasets. In the experiment, other experimental configurations are the same except for the preprocessing process of joint-point data repair. The two datasets after joint data repair processing are recorded as BadmintonData_Repairing and MSRAAction3D_Repairing, respectively. Then, model generation and prediction are performed on the BadmintonData, MSRAAction3D, BadmintonData_Repairing, and MSRAAction3D_Repairing datasets using the above four pose recognition methods, respectively. The predicted results are shown in Table 2.

From the experiment, the same action poses in the BadmintonData_Repairing and MSRAAction3D_Repairing datasets are recognized, and the pose recognition results after the joint repair operation will be much higher than the data without joint repair. It is proved that the processing method based on the bone length and motion continuity proposed in this article can improve the final accuracy of the gesture recognition network. Therefore, it is crucial to evaluate the reliability of the joint-point coordinate information obtained by Kinect and restore the joint point information with errors before the joint-point coordinate information is used for action and pose recognition.

4. Conclusion

Studying, recognizing, and detecting volleyball postures can provide theoretical guidance and suggestions for people, which can be applied in competitions. It also helps event judges make sound decisions. At present, the pose recognition of ball sports is very challenging in action complexity and research data, and this research also has an important application value. This article takes volleyball as the representative to study the problem of volleyball movement pose recognition combined with the analysis and summary of the existing human pose recognition research based on the joint point sequence and the LSTM network. This article proposes a data preprocessing method based on the angle and relative distance feature enhancement and a volleyball motion pose recognition model based on LSTM-attention. The experimental results imply that the data preprocessing method reported here can further improve the accuracy of gesture recognition. The joint-point coordinate information of the coordinate system transformation significantly improves the recognition accuracy of the five ball motion poses by at least 0.01. Moreover, it is concluded

that the LSTM-attention recognition model is not only scientific in structure design but also has considerable competitiveness in gesture recognition performance. However, this method is not yet universal and should be refined in future research.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Q. Dang, J. Yin, B. Wang, and W. Zheng, "Deep learning based 2d human pose estimation: a survey," *Tsinghua Science and Technology*, vol. 24, no. 6, pp. 663–676, 2019.
- [2] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, 2019.
- [3] J. Huang, C. Chi, W. Wang, and H. Huang, "A sequence-scheduled and query-based mac protocol for underwater acoustic networks with a mobile node," *Journal of Communications and Information Networks*, vol. 5, no. 2, pp. 150–159, 2020.
- [4] D. Bodur and M. Demiralp, "Influence of a simple pole on the convergence of separate node ascending derivatives expansion (SNADE) on a sequence of nodes alternating between 2 values," *Mathematical Methods in the Applied Sciences*, vol. 41, no. 17, pp. 7333–7350, 2018.
- [5] K. Smagulova and A. P. James, "A survey on LSTM memristive neural network architectures and applications," *The European Physical Journal - Special Topics*, vol. 228, no. 10, pp. 2313–2324, 2019.
- [6] Y. Huang, M. Kaufmann, E. Aksan, M. J. Black, O. Hilliges, and G. Pons-Moll, "Deep inertial poser: learning to reconstruct human pose from sparse inertial measurements in real time," *ACM Transactions on Graphics*, vol. 37, no. 6, pp. 1–15, 2018.
- [7] Y. Xiao and J. Wu, "Data transmission and management based on node communication in opportunistic social networks," *Symmetry*, vol. 12, no. 8, p. 1288, 2020.
- [8] T. Jamal and S. A. Butt, "Malicious node analysis in MAN-ETS," *International Journal of Information Technology*, vol. 11, no. 4, pp. 859–867, 2019.
- [9] J. U. Seong, B. H. Lee, and D. Yang, "A multi-node channel rendezvous algorithm in cognitive radio ad-hoc networks," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 23, no. 4, pp. 453–461, 2019.

- [10] R. Wang, Y. Liu, and J. Chen, "Network representation learning algorithm combined with node text information," *Journal of Physics: Conference Series*, vol. 1769, no. 1, Article ID 012054, 2021.
- [11] D. Cai and W. Lam, "Graph transformer for graph-to-sequence learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 5, pp. 7464–7471, 2020.
- [12] R. Ma, X. Zheng, P. Wang, H. Liu, and C. Zhang, "The prediction and analysis of COVID-19 epidemic trend by combining LSTM and Markov method," *Scientific Reports*, vol. 11, no. 1, Article ID 17421, 2021.
- [13] C. Liu, Y. Zhang, J. Sun, Z. Cui, and K. Wang, "Stacked bidirectional LSTM RNN to evaluate the remaining useful life of supercapacitor," *International Journal of Energy Research*, vol. 46, no. 3, pp. 3034–3043, 2022.
- [14] H. Lu, L. Jin, X. Luo, B. Liao, D. Guo, and L. Xiao, "RNN for solving perturbed time-varying underdetermined linear system with double bound limits on residual errors and state variables," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 11, pp. 5931–5942, 2019.
- [15] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: a hybrid CNN-RNN based deep learning approach," *International Journal of Information Management Data Insights*, vol. 1, no. 1, Article ID 100007, 2021.
- [16] A. Shewalkar, D. Nyavanandi, and S. A. Ludwig, "Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 9, no. 4, pp. 235–245, 2019.
- [17] N. Majumder, S. Poria, D. Hazarika, R. Mihalcea, A. Gelbukh, and E. Cambria, "Dialoguernn: an attentive rnn for emotion detection in conversations," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 6818–6825, 2019.
- [18] A. H. Khan, S. Li, and X. Luo, "Obstacle avoidance and tracking control of redundant robotic manipulator: an RNN-based metaheuristic approach," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4670–4680, 2020.
- [19] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting," *Energies*, vol. 13, no. 2, p. 391, 2020.
- [20] J. Du, C. M. Vong, and C. L. P. Chen, "Novel efficient RNN and LSTM-like architectures: recurrent and gated broad learning systems and their applications for text classification," *IEEE Transactions on Cybernetics*, vol. 51, no. 3, pp. 1586–1597, 2021.
- [21] B. Zhao, X. Li, and X. Lu, "TTH-RNN: tensor-train hierarchical recurrent neural network for video summarization," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 4, pp. 3629–3637, 2021.
- [22] D. Kollias and S. Zafeiriou, "Exploiting multi-cnn features in cnn-rnn based dimensional emotion recognition on the omg in-the-wild dataset," *IEEE Transactions on Affective Computing*, vol. 12, no. 3, pp. 595–606, 2021.
- [23] H. K. Ahn and N. Park, "Deep RNN-based photovoltaic power short-term forecast using power IoT sensors," *Energies*, vol. 14, no. 2, p. 436, 2021.
- [24] B. Zhao, X. Li, and X. Lu, "CAM-RNN: Co-attention model based RNN for video captioning," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5552–5565, 2019.
- [25] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Sequence of the most informative joints (SMI): a new representation for human skeletal action recognition," *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 24–38, 2014.