Journal of Computers Vol. 32 No. 2, 2021, pp. 187-196
doi:10.3966/199115992021043202016

Sports Video Key Pose Data Incremental Mining Algorithm Based on Deep Learning

Libo Zhang*

Suihua University, Suihua 152061, China
zhanglibozhlb@163.com

Received 1 May 2020; Revised 2 June 2020; Accepted 4 June 2020

Abstract. In order to solve the problems of low classification performance, low statistical similarity and low mining accuracy of traditional data mining algorithms, an incremental mining algorithm for sports video key pose data based on depth learning is proposed. First, the training tag of depth learning is made by using analog signal matrix, and the implementation prospect of sports video key pose frame is extracted with Caffe (Convolutional Architecture for Fast Feature Embedding) open source framework. The interference region in key pose frame is removed by clustering algorithm, and the key pose region of sports video is obtained. Secondly, the SOFM (Self-Organizing Feature Map) network is used to cluster the data of the key pose area of sports video, and the incremental mining model of the key pose data of sports video is established, and the data acquisition operation is carried out. The incremental mining parameters of key pose data of sports video are obtained by using the combined paradigm, finally, the mining parameters are input into the mining model, and the incremental mining of data is realized by using bwmorph method. The experimental results show that the key pose classification performance of the algorithm is much higher than that of the traditional sports video key pose data mining algorithm, the statistical similarity is high, and the method has higher mining accuracy and is more suitable for the mining of the key gesture data of the sports video.

Keywords: Deep learning, Sports video, Key pose data, Incremental mining algorithm

1 Introduction

With the explosive growth of sports video data, a lot of junk data has emerged. Therefore, it is necessary to abandon junk data and obtain key posture data of sports videos in time to analyze sports events scientifically. For various sports videos, coaches, athletes, and audiences pay more attention to the key postures of athletes in the video [1]. For example, in football games, the audience often pays more attention to the moment of the goal or the highlights of the event; in diving competitions, coaches and athletes often pay more attention to the landing and take-off postures. And the action in sports videos is very technical and complex, so it is more difficult to capture key gestures than other videos. Therefore, in order to strengthen the mining of key posture data of sports videos, help various sports events to better judge action, enhance the watching effect of sports events, and assist coaches in guiding athletes' training actions, it is necessary to obtain key posture data in sports videos at present [2].

Aiming at the key pose data mining of sports video, foreign scholars have proposed a model-based key pose data mining algorithm for sports video [3]. First, perform edge detection on sports videos, and then use them for feature training after obtaining edge features; after obtaining pose features, use the voting method to perform key pose recognition on sports videos; finally, complete data mining through linear estimation algorithms; some domestic scholars have proposed a key pose data mining algorithm for sports videos based on pose estimation. The key poses of sports videos are represented by human silhouettes or human contours, the boundary shape of key poses of sports videos is represented, and the key poses of sports videos are described. This description has a certain proportion of non-deformation

---

* Corresponding Author

and rotation non-deformation, which can reduce the impact of noise points on data mining of key poses of sports videos. Finally, the key pose data of sports videos is mined in a non-linear way [4].

Literature [5] proposed an action video key frame extraction algorithm combining pose estimation and specific part tracking. First, the temporal continuity of non-determined human body parts is used to improve the accuracy of human body pose estimation in a single frame image based on a articulated human model of a flexible component, and the candidate key frame set is determined using the extreme value determination principle. Then, strategies like the ISODATA dynamic clustering algorithm are used to implement key frame collection enhancement to determine key frames. Literature [6] proposed a motion video key frame extraction technique combining flexible pose estimation and spatiotemporal feature embedding. First, the temporal continuity of human motion is used to establish an articulated human body model with flexible constraints and time constraints. Through the non-deterministic continuity of human body parts movement constrains, the N-best algorithm is used to estimate the human pose parameters in a single frame of image, and the relative positions and movement directions of human body parts are used to describe human body motion characteristics. The Laplacian score method was used to implement data dimensionality reduction to obtain discriminative human motion feature vectors with strong local topological structure expression capabilities. Iterative self-organizing data analysis technology algorithms were used to dynamically determine key frames. Literature [7] proposed a new method for mining large-scale high-resolution video image data from different perspectives. The Harris corner detection method was used to extract the spatiotemporal features of the high-resolution video image data to be mined. Based on the spatiotemporal characteristics of high-resolution video image data, recursive graphs under different perspectives of the same thing are established through the autocorrelation matrix. The recursive graph is regarded as an image, and the recursive feature descriptor is constructed by calculating the gradient vector of the pixel points. It is necessary to mine the correlation between different perspectives of the same thing, and aggregate high-resolution video image data with the same recursive graph gradient characteristics to achieve data mining.

The above method has a poor removal effect of the interference region in the key pose frame, and it is easy to reduce the mining accuracy. Therefore, an incremental mining algorithm for key pose data of sports videos based on deep learning is proposed, and the algorithm's key pose classification performance, statistical similarity, and mining accuracy are verified experimentally.

2 Design of Incremental Mining Algorithm for Sports Video Key Posture Data

2.1 Acquisition of Key Pose Areas in Sports Videos

Foreground extraction is performed on key pose frames of sports videos based on deep learning, and a clustering algorithm is used to remove interference areas in key pose frames of sports videos to obtain key pose areas.

Fig. 1 shows that before acquiring key frames, it is necessary to extract the foreground of key frames of sports videos based on deep learning to determine whether the cluster area is the largest. If it is the largest, remove the foreground interference area and obtain the key frame pose area. If the cluster area is not the largest, repeat the clustering step until the area is the largest.

If the clustering area does not reach the maximum, repeat the clustering steps until the area is the maximum; if the clustering area reaches the maximum, remove the interference area in the foreground information and obtain the key frame attitude area. Generally, there are a lot of interference information in the foreground of sports video key frame. The interference information is composed of interference information. Compared with the target area (key posture area) to be extracted, it has a negative impact on the key posture extraction process. Therefore, removing the interference area is conducive to improving the reliability of key posture extraction results.

First, it is necessary to extract the foreground of the key pose frame based on deep learning, that is, to segment the key pose frame of the sports video, so as to reduce the influence of the background on extracting the key pose area of the sports video. The specific approach is to label the key posture areas in sports videos, label athletes and sports equipment with red and green, and label background areas with blue [8-9]. Then process the key pose frames of sports video, label the background area as 0, label the athletes and sports equipment as 1, as the analog signal of deep learning. Taking table tennis as an

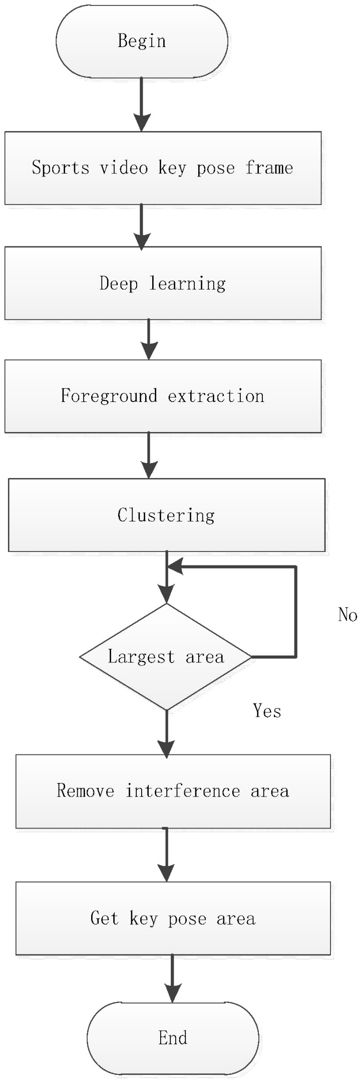
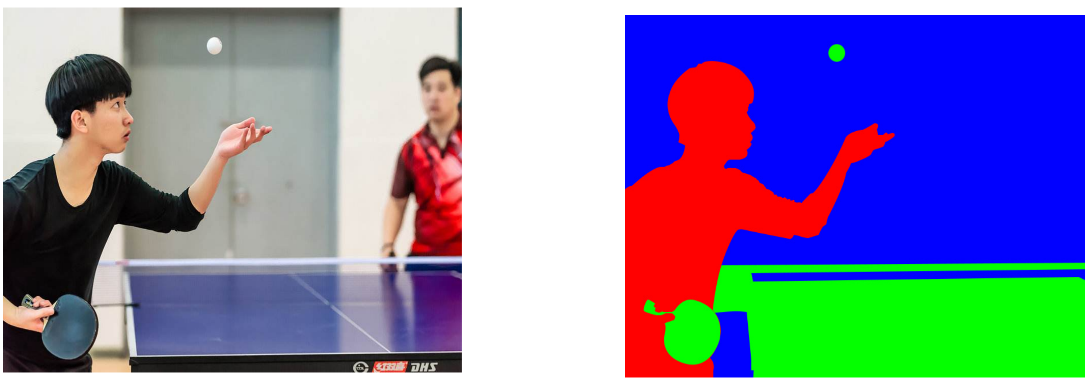


Fig. 1. Key pose area acquisition process

example, Fig. 2(a) shows the original image of table tennis image, and Fig. 2(b) shows the foreground information of image key posture frame extracted based on depth learning technology.



(a) Original (b) Image foreground information

Fig. 2. Table tennis simulation

Using 0 and 1 analog signal matrix to make the training tag of deep learning, as follows:

$$P = \begin{Bmatrix} 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \end{Bmatrix} \tag{1}$$

where, P means the training label for deep learning.

The training tags are combined with the Caffe open source framework to perform deep learning training on the sports video key pose frame pascal dataset, and obtain the trained caffemodel. The caffemodel initialization parameters are used to adjust the key pose frame images of sports videos to obtain the foreground extraction parameters of key pose frames for sports videos [10]. When adjusting the image, set the input parameters of the first layer of AlexNet to 2, which correspond to the athlete, sports equipment area and background area in the key pose frame of the sports video. Countless iterations of the caffemodel until a small range of loss is obtained [11], so that the caffemodel plays the best effect in the region segmentation of key pose frames of sports videos. Use caffemodel to segment the key pose frames of sports videos, and then use the clustering algorithm to optimize the segmentation results. First, perform coarse segmentation on the binarized image after region segmentation: set the label of all the points in the figure as 0, and the image is traversed by pixels, starting from the upper left point of the image. The specific pixel traversal process satisfies is formula (2).

$$\mathrm{label}_{\mathrm{xy}} = \begin{cases} \mathrm{label}_{\max} + 1 \\ \min(\mathrm{label} \mid \mathrm{label} \in LABEL_{\mathrm{xy}}) \end{cases} \tag{2}$$

The pixel traversal process satisfies:

$$pixel_{x,y} = 255 \tag{3}$$

$$pixel_{x+1,y} = pixel_{x-1,y} = 0 \tag{4}$$

$$pixel_{x,y+1} = pixel_{x,y-1} = 0 \tag{5}$$

where, label_{xy} means the label of the current point whose original value is 0; $LABEL_{xy}$ means the integration of the labels whose neighboring region is 0 at the point (x, y); $\mathrm{label}_{\max}$ means the maximum value of label number of each pixel in the image; $pixel_{x,y}$ means the color value of the current point [12-13], which is usually 255 or 0. Then perform fine-tuning on the coarsely segmented image to remove the interference area in the key pose frame of the sports video, so that the label value of the image no longer changes, and to obtain the key pose area of the sports video: $label(x)'$.

$$label(x)' = \min(label \mid label \in LABEL_{xy}) \tag{6}$$

2.2 Incremental Clustering of Key Pose Data for Sports Videos

Conduct incremental clustering of data in key pose areas of sports videos through the SOFM network [14]. First train the SOFM network, and use the trained SOFM network to incrementally cluster the data: Initialize the weights of the SOFM network, i.e., assign a relatively small random number to each W_{ij} as the original value, and input the training samples [15]. Set the original value of the time t as 0, and calculate the distance between the ownership weight and the model X according to the following formula:

$$D^2{}_i = \left\| W_{ij}(t) - X \right\|^2 \tag{7}$$

where, D_i^2 means the distance between the video ownership vector and the model X. Mark the unit coordinate with the minimum distance as r_c, and use D_i^2 to adjust W_{ij}:

$$W_{ij}(t+1) = \frac{W_{ij}(t)}{D^2_{\ i}} r_c^{\ 2} \tag{8}$$

When the value of $W_{ij}(t+1)$ is greater than 1, complete the training of the SOFM network. Then use it to incrementally cluster data from key pose regions of sports videos to obtain the membership matrix U and each cluster center V :

$$U = \{u_{ij}\} \tag{9}$$

$$V = \{v_1, v_2, \ldots, v_n\} \tag{10}$$

Use the time gap T as the basic unit of the incremental clustering, and use the clustering dataset X_i as the center for clustering. Set the cluster center to the core point and the threshold to Ω. When the distance between two clusters is less than Ω, combine these two clusters. After several times of combination, when the data of the key pose area of the sports video meets the clustering accuracy requirements, incremental clustering of the key pose data of the sports video is realized [16-18].

2.3 Incremental Mining of Key Pose Data for Sports Videos

The incremental clustering results of sports video key pose data are used to build an incremental mining model of sports video key pose data to perform data conjunction operations. Incremental mining parameters for key pose data of sports videos are obtained using the conjunction paradigm, and the mining parameters are input into the mining model. The incremental mining of data is achieved using the bwmorph method.

The main structure of the model is incrementally mined by video key pose data. It is shown in Fig. 3.

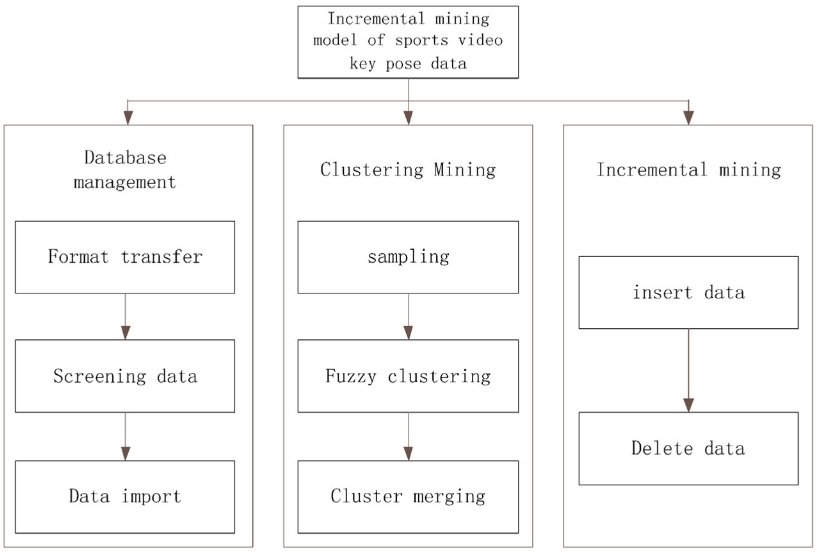


Fig. 3. Incremental mining model of sports video key pose data

Fig. 3 shows that the incremental mining model for sports video key pose data is mainly divided into three parts, namely database management, cluster mining and incremental mining. Among them, the function of database management is to convert video formats, filter video data, and import filtered video data for incremental mining. Clustering mining is realized through sampling, fuzzy clustering, and small class merging, which provides a certain data foundation for incremental mining. Because sports video is constantly changing, the image information in it changes very quickly. After completing a data mining, the result often cannot represent the current motion state. Therefore, after cluster mining, incremental mining is needed in some modes, the newly added data is used to mine again, and the invalid data is deleted after the data conjunction operation.

Use the incremental mining model of sports video key pose data to perform data conjunction operations on the incremental clustering results of the data to obtain the conjunction paradigm of sports video key pose data:

$$L = \wedge W_{ij} L_{ij} \tag{11}$$

where, L means the conjunctive paradigm of key pose data for sports videos; L_{ij} means extraction logic expression of key posture data of sports video [19-20]. Incremental mining parameters of key pose data for sports videos using conjunctional paradigm.

$$\lambda = \frac{L^2}{2} \sum_{i=0} L_{ij} \tag{12}$$

The mining parameters are input into the incremental mining model of key pose data for sports videos, and the incremental mining of key pose data for sports videos can be realized using the bwmorph method. The operation flow of the bwmorph method. It is shown in Table1.

Table 1. Operation flow of bwmorph method

Operation name	clean	diag	hbreak	supr
Specific operation	Remove orphaned pixels	Filling 8 connected regions with diagonal	Remove connected pixels	Remove orphaned pixels
Operation category	Expansion before corrosion	Closed operation	Corrosion before expansion	Closed operation

Table 1 shows that for the morphological operation of sports video binary images, the images are refined until the images no longer change.

3 Experiment Design and Discussion

3.1 Experimental Environment and Data Set

In order to detect the incremental mining algorithm for key pose data of sports videos based on deep learning, a comparative experiment is designed. Sports videos are from the Kinetics (https://deepmind.co,/research/open-source/open-source-datasets/kinetics) video data set. 800 randomly selected frames of the sports video were selected for this experiment, including 132 basketball goal frames, 105 football goal frames, and 130 table tennis serve frames Count, 133 ice hockey pass frames. 173 dash line frames and 127 dive pictures. The parameters of this experiment are shown in Table 2.

Table 2. Experimental parameters

Project	Parameters / execution range	Remarks
Frames of key posture in sports video	800	
Basketball goal frames	132	
Basketball goal frames	105	
Table tennis service frames	130	Experimental platform: MATLAB
Ice hockey pass frames	133	
dash line frames	173	
dive pictures	127	
Data processing tools		Programming language: Java

3.2 Experimental Indicators

The analysis of experimental indicators is as follows:

(1) Key frame pose classification time: Key frame pose classification is helpful for better and more convenient management of motion capture data. It is very time-consuming to find the target motion in the

motion capture data with a large amount of data. Effective pose classification can reduce the data storage capacity.

(2) Statistical similarity: Similarity is the comparison of the similarity of two things. Generally, the distance between the features of a thing is calculated. If the distance is small, the similarity is large; if the distance is large, the similarity is small.

(3) Number of key-frame frames: It is the complete frame of the image in the video, which represents the changed frame. Set the key-frame in the encoder to control the frequency of creating key-frames (i-frames) in the video. The higher the key-frame interval, the more compression is usually applied to the content.

3.3 Experimental Results and Analysis

During the pose mining phase, due to the gradual nature of the spatial pose, similar characteristics of a certain pose may be mined in successive multi-frame images.

Literature [5] proposed an action video key frame extraction algorithm combining pose estimation and specific part tracking; Literature [6] proposed a technique for extracting key frames of motion video combined with flexible pose estimation and spatial-temporal feature embedding, and Literature [7] proposed a method for mining massive high-resolution video image data from different perspectives as a comparison algorithm. Incremental mining algorithm for sports video key pose data proposed in this paper performs data mining on selected sports video key pose frames.

Comparison of Key Pose Classification Performance. Compare results of key pose classification performance of each algorithm in mining.

As shown in Fig. 4, the key pose classification time of the incremental mining algorithm for sports video key pose data proposed in this paper is much lower than the pose estimation, flexible pose estimation, and massive high-resolution methods. The main reason is that the algorithm in this paper uses deep learning analog signal matrix to make training labels, and segment the key pose frames of sports videos, reducing the influence of background on extracting key pose areas of sports videos. This proves the superiority of the proposed incremental data mining algorithm in key pose classification performance.

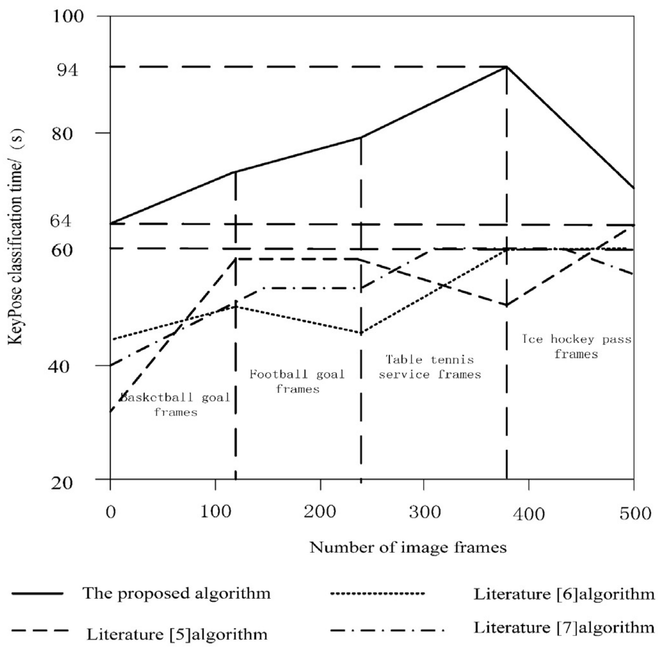


Fig. 4. Comparison of key pose classification performance

Analysis of Statistical Similarity Fitting Comparison Results. In order to compare the mining performance of the traditional method and the method of this paper, for the same video sequence, a certain key pose corresponds to different methods for pose mining. Fig. 5 shows the actual value of statistical similarity. The closer the experimental result is to the actual value, the higher the accuracy of the method.

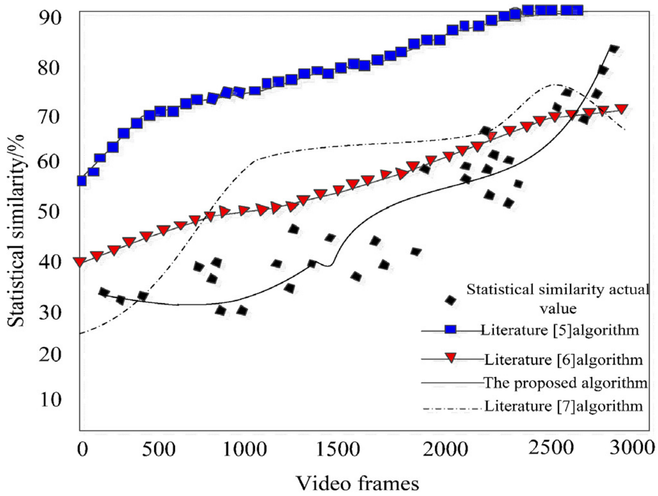


Fig. 5. Comparison of statistical similarity fitting results

Comparative Analysis of Mining Accuracy. Because the number of motion features is too large, it will cause a large number of cumulative errors and reduce the performance of the key frame algorithm. If the number of motion features is too small, the human pose feature discrimination vector cannot accurately represent the original motion features, and the performance of key frame algorithms will also be reduced. Therefore, a certain weightlifting sports video is selected as the training set, 30 samples are extracted from each pose, and the samples are horizontally mirrored. Calculate the accuracy of key frame mining:

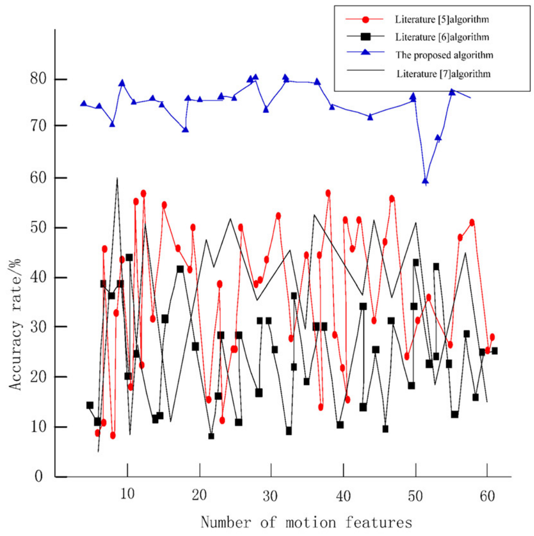
$$Accuracy = \frac{\sum_{i=1}^{m} \delta(f_i, r_i)}{n} \tag{13}$$

where, n means the number of key frames; $\delta(\cdot)$ means the similarity comparison function between the key frame f_i extracted from the key framework function and the manually extracted framework r_i. The experimental video image is shown in Fig. 6.

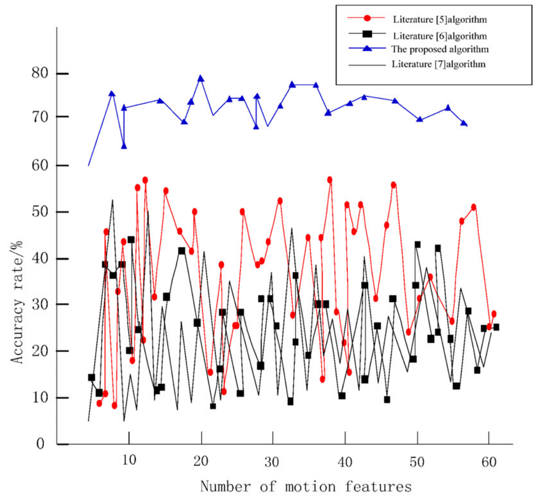
(a) Sports Video 1 (b) Sports Video 2

Fig. 6. Experiment video

Fig. 7 shows the comparison results of keyframe mining accuracy of motion video.



(a) Sports Video 1



(b) Sports Video 2

Fig. 7. Analysis of comparison results of key frame mining accuracy of motion video

Fig.7 shows that the accuracy of the proposed method is much higher than that of pose estimation, flexible pose estimation, and massive high-resolution methods in motion video 1, which is about 75%; In motion video 2, as the number of motion features changes, the mining accuracy of the pose estimation method fluctuates, maintaining an average of about 40%. The mining accuracy of the flexible pose estimation method is close to that of the massive high-resolution method, about 35%. It can be seen intuitively that the proposed method has much higher mining accuracy than pose estimation, flexible pose estimation, and massive high-resolution methods. The main reason is that the clustering algorithm is used to remove the interference areas in the key pose frames, which makes the key pose classification time of the algorithm much lower than the traditional sports video key pose data mining algorithm, which improves the mining accuracy.

4 Conclusions

Incremental mining algorithm for key pose data of sports videos based on deep learning. Incremental clustering of key pose data for sports videos and key pose data for sports videos. An incremental mining model of key pose data for sports videos is established to realize incremental mining of key pose data for sports videos. However, the effect of improving the key frame capture rate of this algorithm needs to be improved. In future research, the key frame capture rate of the algorithm needs to be improved. The proposed incremental mining algorithm for sports video key pose data has profound significance for the incremental mining of sports video key pose data, and its application needs to be expanded.

The key frame algorithm in this paper has the following shortcomings: When the number of human parts of the action parameters changes greatly or the action changes drastically, the algorithm cannot guarantee that the optimal key frame mining result is always obtained. When there is a lot of motion concealment, it will lead to a large error of key frame mining in motion video and affect the overall performance. Therefore, the next research direction considers combining dense logic technology to further improve the optimal mining result of key frames.

References

[1] Z.M. Li, Design of Proofreading Visual Surveillance Platform in National Sports Teaching. Modern Electronic Technology 41(1)(2018) 178-181.

[2] Y.X. Wang, Research on Data Mining Technology in Monitoring Sports Events in the Internet. Television Technology 42(49)(2018) 1121-1124.

[3] L.M. Bai, M.X. Jia, Keypoint Localization Algorithm and Recognition Algorithm Based on Multi-pose Face. Computer and Digital Engineering 35(37)(2018) 173-178.

[4] G.G. Yan, X.K. Yang, J. Weng, et al. A General Method for Solving Attitude Non-Exchangeable Error Compensation Coefficients. Journal of Astronautics 38(47)(2017) 723-727.

[5] N.F. Shi, X.J. Hou, P. Zhang, et al. Keyframe extraction of motion video based on pose estimation and tracking. TV Technology 41(4-5)(2017) 37-46.

[6] N.F. Shi, X.J. Hou, P. Zhang, Extraction of Key Frames in Motion Video with Local Preservation of Spatiotemporal Features. Journal of Computer Applications 37(9)(2017) 2605-2609.

[7] H.N. Zhang, Data mining methods for massive high-resolution video images from different perspectives. Science Technology and Engineering 17 (26)(2017) 257-261.

[8] W.J. Zou, Y.J. Xiao, Contrast Research on Correct Attitude Contour Extraction Methods of Motion Picture. Modern Electronic Technology 518 (15)(2018) 73-76.

[9] Z.X. Xu, R.Z. Zhang, X.J. Shi, et al. Real-time multi-person pose estimation and tracking in deep learning. Journal of China Academy of Electronic Sciences (4)(2018) 491-496.

[10] Z. Wang, J. He, Z.H. Hu, Multi-ROI background modeling method suitable for high-definition surveillance video. Microcomputer System 39 (6)(2018) 1190-1194.

[11] Y. Wei, Y. Zhao, C. Lu, et al. Cross-Modal Retrieval with CNN Visual Features: A New Baseline. IEEE Transactions on Cybernetics 47(2)(2017) 449-460.

[12] P. Xu, Q. Yin, Y. Huang, et al. Cross-modal Subspace Learning for Fine-grained Sketch-based Image Retrieval. Neurocomputing 278(22)(2018) 75-86.

[13] R. Yan, Researches on hybrid algorithm for moving target detection and tracking in sports video. Cluster Computing (2)(2018) 1-10.

[14] L. Zheng, H. Hong, Y. Guo, et al. Supervised Adaptive Incremental Clustering for data stream of chunks. Neurocomputing 219(C)(2017) 502-517.

[15] H. Ordoñez, J.C.Corrales, C. Cobos, et al. Grouping of business processes models based on an incremental clustering algorithm using fuzzy similarity and multimodal search. Expert Systems with Applications 67(2017) 163-177.

[16] A.M. Abdo, X. Zhao, Z. Rui, et al. MU-MIMO Downlink Capacity Analysis and Optimum Code Weight Vector Design for 5G Big Data Massive Antenna Millimeter Wave Communication. Wireless Communications and Mobile Computing 2018(2018) 1-12.

[17] H. Elwahsh, M. Gamal, A.A. Salama, et al. Modeling Neutrosophic Data by Self-Organizing Feature Map: MANETs Data Case Study. Procedia Computer Science 121(2017) 152-159.

[18] S. Gaglio, G.L. Re, M. Morana, Human Activity Recognition Process Using 3-D Posture Data. IEEE Transactions on Human-Machine Systems 45(5)(2017) 586-597.

[19] C. Zhang, L. Hao, F. Li, Optimization and improvement of data mining algorithm based on efficient incremental kernel fuzzy clustering for large data. Cluster Computing (7)(2018) 1-10.

[20] P. Kirst, F. Rigterink, O. Stein, Global optimization of disjunctive program. Journal of Global Optimization 69(2)(2017) 283-307.