Classification Algorithm of Sports Teaching Video Based on Wireless Sensor Network

Zhipeng Chen(✉)

Chongqing Vocational Institute of Engineering, Jiangjin 402260, China
chenzhipeng1993@126.com

Abstract. In view of the problems of low recall and accuracy caused by the huge amount of physical education teaching videos, a physical education teaching video classification algorithm based on wireless sensor network is proposed. The classification framework of physical education teaching videos based on wireless sensor networks is constructed, and the video node coordinates are located according to the structural relationship of physical education teaching videos. Initialize the histogram index, calculate the similarity of any two frames of the video, and set the clustering index of key frames of the physical education teaching video based on the distance between the two frames. Retrieve the video to be classified, find the sensor node with the largest weight, calculate the distance between the target and the detection sensor node, design the video classification steps of physical education teaching, and realize video classification. The experimental results show that the minimum recall rate of this algorithm is 87%, and the maximum classification accuracy rate is ninety-seven percent, which has the advantages of high classification recall and accuracy.

Keywords: Wireless Sensor Network · Physical Education · Video Classification · Accuracy Rate · Recall Rate

1 Introduction

Sports teaching video is an important video resource with a wide range of users. Due to the rapid development of information technology, the research on the classification of sports teaching videos has gradually become a hot issue concerned by relevant personnel. At present, China's sports cause is in a rapid development stage, and sports teaching video data also shows a massive growth. Efficient sports teaching video classification has important application value for high-quality sports teaching video browsing and retrieval. There is a large amount of sports teaching video information in the network. If there is no efficient sports teaching video retrieval system, the sports teaching video in the network will be disordered. Therefore, how to organize sports teaching video resources with high efficiency and precision, and realize the classification and sorting of sports teaching videos is conducive to helping users to obtain their own sports teaching video content efficiently.

L. Yun et al. (Eds.): ADHIP 2023, LNICST 547, pp. 291–305, 2024.
https://doi.org/10.1007/978-3-031-50543-0_20

In the past, sports teaching videos were managed and classified in the form of manual annotation, which not only wasted more human resources, but also had low classification accuracy due to a large number of subjective human factors. Therefore, the research and development of a method that can accurately and reasonably classify sports teaching videos has great scientific research value. Reference [1] proposed a key frame extraction method based on inter frame difference, which is to extract key frames based on the change value of image information (color, texture, brightness, etc.). When the change value is less than the preset threshold, it can be selected as a key frame. The key frame extraction method based on the difference between frames is simple and fast, but the calculation is large and the results are easily affected by the selection of starting frame and threshold setting. The key frame extraction method based on clustering clusters the frames with similar image characteristics, and takes the intermediate frame of each class as the key frame. Reference [2] proposed a method to obtain specific motion frames based on human posture estimation and clustering. Each frame of the motion video is classified according to similarity, and one frame is selected as the key frame for each category according to specific criteria; Reference [3] proposed a dynamic clustering method. First, the predicted number of classifications is calculated according to the similarity of adjacent video sequences and the characteristics of motion parameters to complete the video decomposition; Then, the ISODATA algorithm is used to adaptively calculate different class thresholds, and the video sequence clustering and key frame extraction are completed by combining and splitting operations. It can be seen that the above algorithm does not consider the motion expression ability of key frames, which easily leads to the distortion of motion sequence analysis.

From the existing academic achievements, due to the influence of action diversity, background complexity, etc., there is no universal method of action key frame extraction that can fully achieve the above goals. As far as the sports video review of physical education teaching is concerned, because it needs to judge whether the sports action is correct or not according to the frame sequence of interest in the aerobics video, the requirement for sports expression ability is higher. Therefore, a classification algorithm of sports teaching video based on wireless sensor network is proposed.

2 Video Key Frame Clustering Based on Wireless Sensor Networks

The classification framework of sports teaching video based on wireless sensor network is shown in Fig. 1.

According to the characteristics of video data, video content information is extracted from two aspects. First, data irrelevant to content information, including file path, playing date and time length; The second is data related to content information, including personal entities, semantic events and feature data [4]. The retrieval module is mainly responsible for extracting features according to the user's retrieval information, matching them with the information in the database, and returning the retrieval results. For the segmentation of the original video stream, video boundary detection and audio detection are combined to improve the efficiency of video segmentation. In fact, due to the characteristics of sports teaching videos, most of the other shot changes are abrupt, except for gradual change detection of the replay scene shot. The general algorithm based on color histogram threshold segmentation can achieve better results.

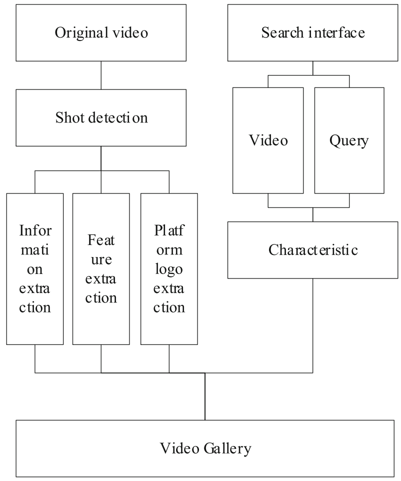


Fig. 1. Classification framework of sports teaching videos

2.1 Coordinate Positioning of Video Nodes in Wireless Sensor Networks

In addition to the general video data structure, the structure of sports teaching video content also has its own characteristics. First of all, the semantic information in the sports teaching video is relatively clear, reducing the fuzziness and subjectivity between the underlying semantic and high-level semantic. Secondly, different events have specific structures and rules. There are obvious differences between the number of people, groups, time, equipment and other factors in the event. These differences are conducive to the analysis and understanding of the video, it can help users quickly find the content they need from a large number of videos, and directly and accurately locate and browse [5]. Video is an unstructured data stream. For video structure, it can be divided into four levels from small to large: frame, shot, scene and video.

Frame represents a still image, which is the smallest unit of video. Shot is a video sequence composed of adjacent consecutive frames, and shot is the basic unit of video retrieval. In the retrieval process, one or more key frames are usually extracted to describe a shot, and then a video is retrieved through the matching similarity of key frames. A scene is composed of a set of semantically related shots that express a complete event, usually with the same background and character information.

In conclusion, the number of hops between nodes is used as the eigenvector to calculate the similarity. Based on the location information of beacon nodes, the positioning algorithm of wireless sensor networks classifies and determines the coordinate interval of unknown nodes for many times. Therefore, the process of determining the node location by the location algorithm is divided into three stages: hop count calculation stage, beacon broadcast stage, and node self location stage.

Hop count calculation stage: beacon node broadcasts its own location information packet to neighbor node, including hop count field and node number. The hop count field is initialized to 0. The receiving node records the minimum hops to each beacon node, ignoring packets with larger hops from the same beacon node. Then add 1 to the hop value and forward it to the neighbor node. With this method, all nodes in the network can record the minimum hops of each beacon node.

Beacon broadcasting stage: each beacon node forms an eigenvector according to the hops from other beacon nodes recorded in the first stage. Accordingly, the unknown node forms the eigenvector according to the hops away from the beacon node. Then, beacon nodes broadcast packets containing node numbers and eigenvectors to the network, and unknown nodes are only responsible for forwarding packets from beacon nodes to ensure that nodes in the network receive packets from each beacon node [6].

Based on the positioning algorithm of wireless sensor network, the X axis and Y axis of unknown nodes are classified respectively to determine the range of coordinates:

After the X axis is classified, its space is $X \in \left[x\frac{L}{2^n}, (x+1)\frac{L}{2^n}\right]$ after classification, the space of Y axis is $X \in \left[x\frac{L}{2^n}, (x+1)\frac{L}{2^n}\right]$, $Y \in \left[y\frac{L}{2^n}, (y+1)\frac{L}{2^n}\right]$, where L indicates the length of 2D space, n indicates the number of classifications. According to the above classification process, if the location algorithm of wireless sensor network is correct after 2^n classification, the space where the unknown node is located can be expressed as:

$$O_n \in \left[x\frac{L}{2^n}, (x+1)\frac{L}{2^n}\right] \times \left[y\frac{L}{2^n}, (y+1)\frac{L}{2^n}\right] \tag{1}$$

It can be seen from the formula that the unknown node is in the square area.

Node self localization stage: the unknown node starts the self localization stage after receiving the packets of all beacon nodes in the beacon broadcast stage. According to the feature vector, calculate the similarity with the beacon node, and the wireless sensor network localization algorithm classifies the X axis and Y axis of the unknown node respectively to determine the node coordinates.

2.2 Key Frame Clustering of Sports Teaching Videos

Features represent a certain target and some quantifiable attributes. For sports teaching videos, they mainly include general features and specific field features. Considering the efficiency of key frame extraction of sports teaching video, the image features of sports teaching video are set as color histogram and color distribution descriptor [7]. In general, the description of the color of the sports teaching video image belongs to the color space problem. The key frame extraction algorithm based on the similarity coefficient of the shot boundary is used. After the image is in the HSV color space and the color histogram is derived, the histogram index is initialized first. The similarity of the two frames of the sports teaching video histogram can be seen as:

$$sim_k(a, b) = \sum_{k=1}^{3} \min\{H_a(l_a, d_a, h_a), H_b(l_b, d_b, h_b)\} \tag{2}$$

In formula (2), l_a, d_a, h_a represent histograms respectively H_a length, width and height of; $H_b(l_b, d_b, h_b)$ represent histograms respectively H_b length, width and height of; k represents the number of calculations; a, b represents two frames of the histogram. In the formula, 0 describes that the color histogram difference between the two graphs is very large, and 1 describes that the color histogram difference between the two graphs is the same [8]. For the color segment descriptor, which is used to represent the spatial part of the color in the sports teaching video image, the feature extraction process is: sports teaching video image segmentation, dominant color selection, discrete cosine conversion of 64 pixels (brightness signal Y, blue difference b, red difference r) components, obtaining three groups of coefficients, and finally Zigzag scanning of the obtained discrete cosine coefficients, A new distribution descriptor [9] is established by selecting a few low frequency coefficients. At this time, the distance d_k between frames a, b is set as:

$$d_k = \sqrt{\omega \times \lambda(Y)^2} + \sqrt{\omega \times \lambda(b)^2} + \sqrt{\omega \times \lambda(r)^2} \tag{3}$$

In formula (3), ω represents weight; $\lambda(Y)$, $\lambda(b)$ and $\lambda(r)$ respectively represent the discrete cosine coefficients of Y, b and r describing the frame.

Because the same shot will appear repeatedly in the same sports teaching video, resulting in repeated key frame sequences obtained. In order to reduce the repeatability of the last key frame sequence obtained, cluster the key frame sequences: K average clustering, and finally set according to the clustering effectiveness method K value side. Clustering performance indicators are:

$$\eta = s_a \times V(a) + s_b \times V(b) \tag{4}$$

In formula (4), $V(a)$, $V(b)$ respectively a, b two frame key frame sequence class; s_a, s_b represent and describe the distance between two frames. Because the value ranges of these two items are very different, set a weight factor, and the maximum number of preset clusters obtained when this value is at the minimum value is the optimal number of clusters.

3 Video Classification Algorithm Based on Wireless Sensor Networks

3.1 Video Retrieval to Be Classified Based on Multi View Cooperative Operation

Correctly describing the relationship model between video data and establishing an efficient index can effectively help the system filter out a lot of irrelevant information [10]. In this paper, a hierarchical classification structure is given, which is illustrated by taking the project event type stratification as an example. The hierarchical structure is shown in Fig. 2.

Figure 2 shows the hierarchical relationship between project event entities. For example, sports events are first divided into ball games, field events and gymnastics, and then ball games are subdivided into football, basketball and volleyball, and gymnastics is subdivided into rings, vault and horizontal bar. Similarly, hierarchical relationships are also established for the extracted video information such as individuals and icons. When video retrieval is carried out, a multi view collaborative retrieval method is used, and the steps are as follows:

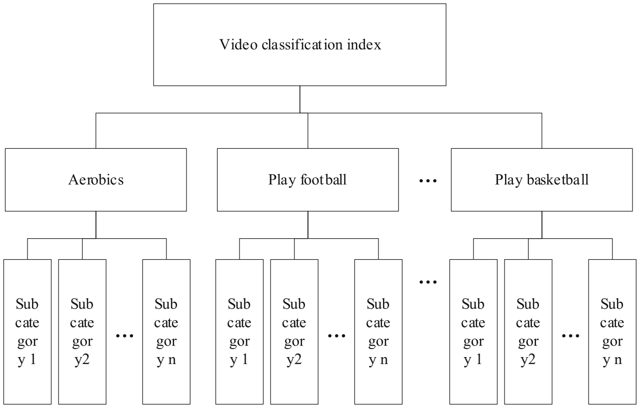


Fig. 2. Structure of Classified Events

Step 1: The information related to the retrieval of video content in the video library can be expressed by the following formula:

$$W = G \cap H \tag{5}$$

In formula (5), G is the video segment of the search video related to frame a; H is the video segment of the retrieved video related to the frame b. For multiple relationships, the intersection of multiple relationship sets is obtained according to the above method, and the content of the set is a candidate video clip.

Step 2: Let the total matching similarity between a video segment in W set and the retrieved video for the relationship between frame a and frame b be:

$$Sim' = \frac{Sim_a + Sim_b}{2} \tag{6}$$

In formula (6), Sim_a and Sim_b represent the similarity of two frames respectively. By analogy, the total similarity of multiple relationships can be expressed as:

$$Sim'' = \frac{\sum_{k=1}^{N} Sim_N}{N} \tag{7}$$

In formula (7), N represents the total number of mapping relationships.

Step 3: During video retrieval, query the video or auxiliary text information according to a segment given by the user, and get the video information with high similarity and return it to the user according to the extracted classification information after steps 1 and 2.

Because step 1 combines the relational semantics of multiple information, the information that is obviously irrelevant to the problem can be filtered out after the operation, which greatly reduces the amount of videos to be searched when retrieving videos. The comprehensive similarity is obtained through the multi view collaborative operation in step 2, and the retrieval efficiency and detection rate will be greatly improved.

3.2 Retrieval Video Tracking Based on Wireless Sensor Network

The master node sends request information to all other neighbor nodes that are one hop away from the master node, and takes them as slave nodes and forms a cluster directly. Each adjacent sensor node sends a join request message to the master node, which contains location information, and informs the master node to use it as a slave node. After the cluster is formed, the master node establishes a time division multiple access plan and sends it to its slave nodes accordingly. This mechanism avoids the conflict of sending messages. After all slave nodes know the TDMA scheduling, the slave node of each detection target will find its residual energy and use the TOA method to calculate its distance from the target. This data is sent to the master node through the slave node data message in the allocated TDMA timeslot. On the other hand, the master node obtains data from the node message and aggregates its own data to calculate the distance between the detection sensor node and the base station. The master protocol basically maintains three lists: the first list contains the distance from the target to the detection sensor node; The second type includes detecting the residual energy of sensor nodes; The third includes the distance between the detection sensor node and the receiving node. The master protocol uses these lists to find the sensor node with the highest weight according to the following formula:

$$w_{\max}(i) = \left(\frac{E(i)}{d_1 \times d_2}\right) \tag{8}$$

In formula (8), $E(i)$ represents the weight of detecting the sensor node to be selected as the master node i; d_1 and d_2 represent the distance from the detection sensor node to the base station and the distance from the detection sensor node to the target, respectively. If the current master node still has the highest weight value, it will still be the master node in the next round and reassign the same slave node. Otherwise, you must select a new master node. In addition, the first list maintained in the master node is used to estimate the target location in the location algorithm, which will be discussed later. After the target location is estimated, the main node records the estimated target location in the target archive, and then sends it to the base station through a cluster node message.

In order to calculate the distance between the moving target and the sensor node, TOA method is used. In the proposed algorithm, TOA represents the time from the sensor node to the moving target. Therefore, considering the signal speed, the distance between the target and the detection sensor node can be directly calculated according to the following formula:

$$d' = \Delta t \times v \tag{9}$$

In formula (9), Δt indicates signal delay; v indicates the signal speed. In some applications, obstacles may prevent the direct line of sight between the sensor node and the target. Therefore, TOA readings may be affected. However, the location algorithm used in IAH is the same as the cluster head adaptive clustering algorithm, which mainly depends on the circle exchange principle. The positioning algorithm uses the results of the following formula, combined with the position of the sensor nodes, to obtain a set of circular equations:

$$(x - i_k)^2 + (y - j_k)^2 = r_k^2 \tag{10}$$

Stay (i, j) each sensor node knows its position when the network is deployed; r_k It is the distance between the sensor node and the moving target. The position of the target tracking result is obtained by solving these equations.

3.3 Design of Video Classification Steps for Physical Education Teaching

On the basis of obtaining the key frame features of sports teaching video images, the sports teaching video image classification method based on the deep learning coding model is adopted to realize the sports teaching video classification. The detailed process is as follows:

The multi-level restricted Boltzmann machine is used to encode and learn the key frame feature library of sports teaching video images, and become a model visual dictionary. According to the spatial information of the key frame feature library of sports teaching video images, set the features of the adjacent key frame feature library of sports teaching video images as the input of RBM, train RBM with wireless sensor network positioning algorithm, and obtain the hidden layer features; Then the adjacent hidden layer features are regarded as the input of the lower layer RBM, and the output dictionary is obtained. When any weight belongs to the connection weight of RBM, RBM has an explicit layer and a hidden layer, while neurons based on the same layer in RBM do not have connection relationship. During network training, the hidden layer and the visible layer of RBM are connected according to the conditional probability distribution.

By setting the weight matrix ω_h and the hidden layer bias vector ε, the input layer features β can be encoded into the corresponding visual dictionary φ, and by setting the ω_h and the explicit layer bias matrix ϕ, the physical education teaching video features can be reconstructed through the visual dictionary. For a set of input layers and coding layers in RBM, its energy function is:

$$f(E) = -\sum_{k=1}^{i}\sum_{k=1}^{j}\beta_{ij}\omega_h\varphi_{ij} - \sum_{k=1}^{i}\sum_{k=1}^{j}\phi_{ij}\beta_{ij} \tag{11}$$

The calculation of the energy function can obtain the joint probability distribution characteristics of the energy function and the edge distribution of the joint distribution of the sports teaching video - the probability of the feature input node. RBM network training is mainly to maximize the probability of the input node. In general, the Monte Carlo Romankov chain method can be used to obtain the key frame feature vector of the sports teaching video.

The CD algorithm is used to implement rapid learning of RBM to improve the convergence efficiency of parameters. The updated amount of weight obtained is:

$$\Delta\omega_h = v_h\big(\langle\beta\varphi\rangle - \langle\beta\varphi\rangle'\big) \tag{12}$$

In formula (12), v_h describe learning speed; $\langle\beta\varphi\rangle$ and $\langle\beta\varphi\rangle'$ represent the actual weight and threshold of the visible layer and hidden layer respectively.

If all layers of sports teaching video features are trained at the same time, the time complexity will increase; If only one layer is trained at a time, the transmission of each layer will aggravate the degree of underfitting. Therefore, when using deep learning to encode the characteristics of sports teaching video, first obtain the visual dictionary through bottom-up unsupervised RBM hierarchical training.

According to the position $(x_t,\ y_t)$ of the visual dictionary and the previous position $(x_{t-1},\ y_{t-1})$, the current speed is obtained, and the calculation is as follows through message sending:

$$v_t = \frac{\sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}}{t - (t - 1)} \tag{13}$$

In formula (11), t represents time.

Predict the next position. Based on the current speed, assume that the speed and direction of the target movement remain the same as the original. Therefore, after a given time, the predicted position of the target represents:

$$\begin{cases} x'_{t+1} = x_t + v_t \Delta t \cos\theta \\ y'_{t+1} = y_t + v_t \Delta t \sin\theta \end{cases} \tag{14}$$

In formula (14), θ indicates the current direction; Δt represents a given time.

Step 4: When selecting the master node, use the predicted location as the classification parameter. Therefore, the new primary node selection criteria are defined as:

$$w'(i) = \left(\frac{E(i)}{d_1 \times d_2 \times d_3}\right) \tag{15}$$

In formula (15), d_3 indicates the distance between the predicted position and the detection node.

The number of active nodes in the formed cluster is determined according to the angle value obtained in the previous step. In other words, if the angle value is less than a certain error threshold, that is, the prediction error is relatively small, the number of active nodes in the formed cluster will be reduced to half, based on which the teaching video classification will be realized.

4 Experiment

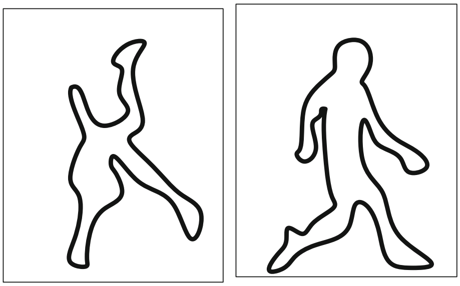
The experiment adopts the sports teaching video submitted by students on the hybrid teaching platform of a science and technology college, with the sampling frequency of 24 frames/s, the total length of the video of 2.03 min, and a total of 2953 frames. The simulation experiment is carried out in the Matlab2015 environment.

4.1 Experimental Objects

Through the extraction of action video clips based on beat, each video clip obtained represents a sports action clip. Because the key frame extraction only focuses on the moving human body in each sports action video clip. So first, with the video segment unit and the previous frame as the reference, the L-K method is used to calculate the optical flow, and it is detected that the moving target in each frame may be the background rather than the human body image at some times due to the background jitter and other noise effects in the complex scene, as shown in Fig. 3 (a). For this reason, after obtaining the moving target, HoG classifier is used to complete the human detection, and the bounding box containing the human body is obtained, as shown in Fig. 3 (b).

aerobics play football

(a) L-K optical flow diagram with the previous frame as reference



aerobics play football

(b) Human body bounding box obtained by HoG classifier

Fig. 3. Actual moving target detection

In order to describe the motion characteristics in the human body bounding box, after the human body bounding box is normalized, the previous frame is the reference frame. According to the cell unit (unit: pixel), the L-IC method is used to estimate the motion direction of each cell unit, and the motion characteristics of the human body bounding box in each frame are obtained.

4.2 Experimental Environment

In order to further verify the algorithm performance, CC2510 node is used to build a wireless sensor network experiment test platform in the open area, as shown in Fig. 4.

121 nodes in the laboratory (beacon nodes evenly distributed among them) cover 30m $\times$ 30 m area, the distance between nodes is 3 m, forming 10 $\times$ 10 Grid distribution. After testing, beacon nodes can “reach each other”. The experimental parameters are shown in Table 1.


Floor mounted
camera

Fig. 4. Experimental Test Platform

Table 1. Experimental Parameter Setting Table

Serial number	Parameter	Describe the content
1	Network type	Wireless sensor network
2	network topology	Network composed of multiple wireless sensor nodes
3	data acquisition	Wireless sensor nodes collect video data
4	data transmission	Wireless sensor networks transmit video data to algorithm processing nodes
5	Experimental evaluation index	Classification node energy consumption, classification recall and classification accuracy

4.3 Experimental Results and Analysis

The key frame extraction method based on inter frame difference, the specific motion frame acquisition method based on human posture estimation and clustering, the dynamic clustering method and the classification algorithm based on wireless sensor network are respectively used to compare and analyze the energy consumption of classification nodes. The comparison results are shown in Fig. 5.

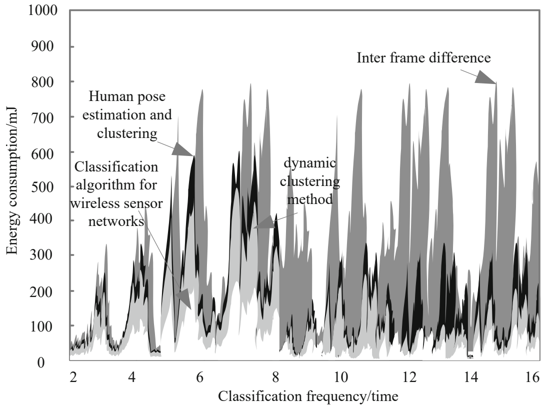


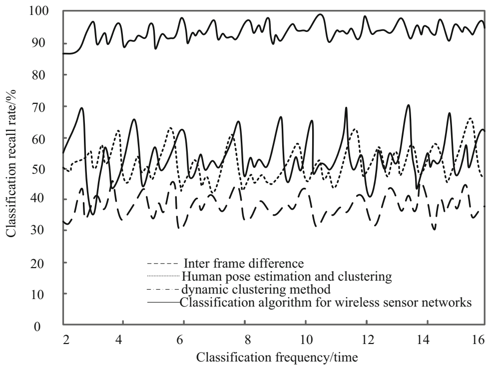
Fig. 5. Comparison and analysis of node energy consumption by different methods

As shown in Fig. 5, the key frame extraction method based on inter frame difference has the largest energy consumption of nodes, the reason is that this method needs to compare each frame and calculate the difference between frames, so as to extract key frames. And followed by the specific motion frame acquisition method based on human posture estimation and clustering, and the dynamic clustering method. Both methods need to estimate human posture or cluster analysis, and then extract specific motion frames. The wireless sensor network classification algorithm has the smallest energy consumption of nodes, with the maximum energy consumption value of only 240 mJ. The energy consumption of nodes should be considered when designing the video classification algorithm for physical education teaching. Choosing a low energy consumption method can reduce the energy consumption of nodes and prolong the network life.

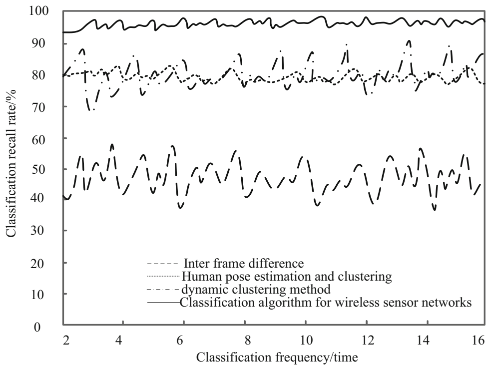
In order to further verify the effectiveness of the algorithm studied, the classification recall rates of the three methods are compared, and the comparison results are shown in Fig. 6.

It can be seen from Fig. 6 (a) that the recall rate of classification using the three traditional methods is less than 70%, because it is difficult for traditional methods to accurately classify complex sports actions, some key frames or specific sports frames are missed, which affects the recall rate. And the minimum recall rate using the algorithm studied can also reach 87%, compared with the traditional method, it has been significantly improved.

It can be seen from Fig. 6 (b) that the recall rate of specific motion frame acquisition method and dynamic clustering method based on human posture estimation and clustering exceeds 70%, it shows that these methods can better capture the key information of sports movements. The recall rate of key frame extraction method based on inter frame difference is less than 60%, because this method can't accurately capture the important changes of actions when extracting key frames. And the minimum recall rate of classification algorithm based on wireless sensor network can also reach 94%, it shows that the algorithm has a high recall rate in video classification of physical education teaching.



(a) Aerobics



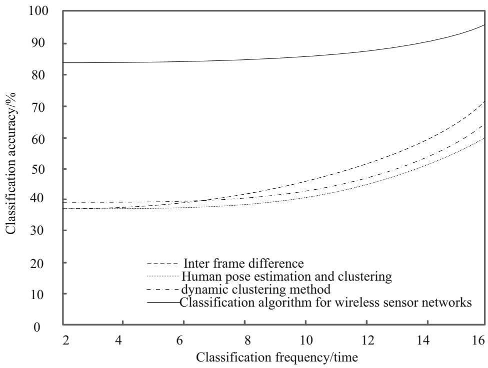
(b) Play football

Fig. 6. Comparative Analysis of Recall Rates by Different Methods

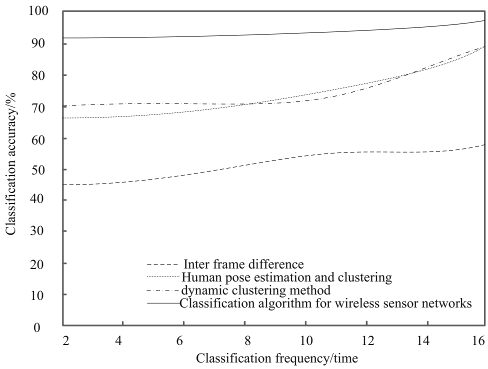
The classification accuracy of the three methods is compared as shown in Fig. 7.

It can be seen from Fig. 7 (a) that the maximum classification accuracy of the three traditional methods is 71%, there is a big classification error when dealing with complex sports movements, which leads to low accuracy of classification results. While the maximum classification accuracy of the algorithm studied is 96%, compared with the traditional method, it has been significantly improved.

It can be seen from Fig. 7 (b) that the maximum classification accuracy of specific motion frame acquisition method and dynamic clustering method based on human posture estimation and clustering is 87%, this shows that these methods can better capture the key information of sports movements and accurately distinguish different types of movements in the classification process. The maximum classification accuracy of key frame extraction method based on inter frame difference is 59%, because this method can not accurately extract representative key frames. And the maximum classification accuracy of the algorithm studied is 97%, it shows that the algorithm has high classification accuracy in video classification of physical education teaching.



(a) Aerobics



(b) Play football

Fig. 7. Comparison and analysis of classification accuracy of different methods

5 Conclusion

Aiming at the problem of poor classification effect of sports teaching videos using traditional methods, this paper proposes a classification algorithm of sports teaching videos based on wireless sensor networks. Experimental results show that, compared with existing algorithms, this algorithm can accurately track targets, and has lower energy consumption and higher classification accuracy, which shows the effectiveness and feasibility of using the proposed algorithm. The algorithm studied in this paper has a high recall rate in the classification of sports teaching videos, and can better capture the key information of sports actions. At the same time, the algorithm studied in this paper has high classification accuracy in video classification of physical education teaching, and can identify and classify different sports actions more accurately.

References

1. Cao, C.P., Yuan, K.G.: Key frame extraction algorithm of reinforcement learning based on multi-channel feature and attention mechanism. Appl. Res. Comput. **39**(4), 1274–1280 (2022)

2. Cai, M.M., Huang, J.F., Lin, X., et al.: Acquisition method of specific motion frame based on human attitude estimation and clustering. J. Graph. **43**(1), 44–52 (2022)
3. Huang, W., Wang, Y., Zhang, L., et al.: Key frame extraction algorithm for theodolite image sequence. J. Appl. Opt. **43**(3), 430–435 (2022)
4. Zhang, M.Q., Li, W.P.: An automatic classification method of sports teaching video using support vector machine. Sci. Program. **2021**(9), 1–8 (2021)
5. Ass, A., Sh, B., Jp, B., et al.: Classification of educational videos by using a semi-supervised learning method on transcripts and keywords. Neurocomputing **2021**(7), 637–647 (2021)
6. Lin, Y., Liu, H., Chen, Z., et al.: Machine learning-based classification of academic performance via imaging sensors. IEEE Sens. J. **21**(22), 24952–24958 (2021)
7. Zheng, Y., Shi, G.: Research on data retrieval algorithm of English microlearning teaching based on wireless network information classification. J. Sens. **2021**(7), 1–10 (2021)
8. Hong, S., Kim, J., Yang, E.: Automated text classification of maintenance data of higher education buildings using text mining and machine learning techniques. J. Archit. Eng. **28**(1), 1–10 (2022)
9. Liang, X., Yin, J.: Recommendation algorithm for equilibrium of teaching resources in physical education network based on trust relationship. J. Internet Technol. **23**(1), 133–141 (2022)
10. Li, J.: Application of mobile information system based on internet in college physical education classroom teaching. Mob. Inf. Syst. **2021**(9), 1–10 (2021)