# Research on Feature Extraction Method of Aerobics Jumping Movement Based on AdaBoost Algorithm and Gaussian Mixture Model

Qianqian Zhang [1], Xian Lin [2, *]
[1] Zhejiang Pharmaceutical University, Ningbo, 315500, China
[2] Chaoyang Primary School, Ningbo, Ningbo, 315195, China
E-mail: linxx018@163.com
[*]Corresponding author

**Keywords:** AdaBoost algorithm, machine learning, action feature extraction

**Received:** November 1, 2024

*Aiming at the problems of sensitive background interference, insufficient key frame recognition accuracy and low computational efficiency of traditional aerobics jumping action feature extraction methods, this study proposes a feature extraction method that integrates Gaussian mixture model, entropy sequence fusion and AdaBoost algorithm. The video key frames are extracted by machine vision technology, combined with entropy sequence (standard deviation ± 0.3) and music energy features (threshold = 0.85) to achieve synchronized key frame recognition (96.8% accuracy); Gaussian mixture model is used to eliminate background noise (42% reduction in false detection rate), combined with Harris3D operator to construct the action potential function, and integrated with AdaBoost algorithm to integrate the weak classifier to optimize feature extraction. The experiments show that compared with the existing SOTA method, the average error of azimuth recognition is 1.2° (significantly lower than 3.8° ±1.1° in A-BLSTM and 4.5° ±1.5° in NN-BIGRU, p<0.05); the feature extraction rate is improved to 92.4% (32.5% higher than that of the MEM-LBP method); and the processing efficiency reaches 35ms/frame, which is higher than that of the A-BLSTM (50ms). BLSTM (50ms) and NN-BIGRU (48ms) by 30% and 27%, respectively. In terms of comprehensive performance, the accuracy (96.8%), recall (94.5%) and F1 score (95.6%) are close to that of A-BLSTM (97.0%/95.2%/96.1%), but the computational resource requirement is 35% lower; and the feature purity in complex contexts (variance ±0.3) significantly outperforms that of multi-threshold optimization methods (variance ±1.5). This study provides a high-precision and low-latency analysis tool for aerobics training and verifies its robustness in real-time action recognition scenarios (95% confidence interval error band width narrowed to 0.5° ), which provides new ideas for cross-domain applications of sports action analysis.*

*Povzetek: Članek predlaga metodo za ekstrakcijo skokovnih gibov v aerobiki z Gaussovimi modeli, entropijsko fuzijo in AdaBoost algoritmom, ki izboljša kvaliteto in učinkovitost v realnem času.*

## 1 Introduction

With the widespread application of communication technology and wireless systems, human activity recognition has gradually become an important part of artificial intelligence research due to its broad application potential in various fields. HAR provides possibilities for a range of application areas, including elderly monitoring, fall detection, gesture recognition, and respiratory tracking [1,2]. Structured monitoring and analysis of elderly behavior can predict potential health risks, while gesture recognition provides services for hearing-impaired communities, making HAR one of the most prominent and influential research topics in multiple fields. However, research and implementation of HAR technology face some challenges, and existing research mainly focuses on the fields of vision and sensors. Although the visual system provides rich data for activity recognition, it also faces many challenges, such as environmental lighting, background confusion, object occlusion, and other factors that often affect the effectiveness of imaging and analysis [3-5]. Similarly, cameras may cause privacy issues. On the other hand, sensor technology used in HAR can provide accurate recognition results, but its high cost and the need for users to actively carry it reduce its adaptability and feasibility in practical scenarios. In view of this, HAR technology based on WIFI has received widespread attention [6-8]. WIFI devices have lower costs, relatively less resource consumption, and are not affected by environmental lighting and camera privacy issues. Therefore, it has significant advantages in practical applications. In addition, WIFI has a wide coverage range and is easy to integrate, making WIFI based HAR demonstrate good application effects in various application scenarios [9].

Research on HAR has made a series of progress in the field of deep learning. Yousef [10] proposed a long short-term memory network model, which can better solve the problems of vanishing and exploding gradients compared to recurrent neural networks. Compared with traditional random forest and hidden Markov model algorithms, the LSTM model has an accuracy of over 75%. Especially, for the first time, research has directly obtained action features through deep learning models without the need for any feature extraction processing. Given the low accuracy of the above research, Chen [11] proposed an A-BLSTM model that integrates attention mechanism and bidirectional LSTM model. By using bidirectional LSTM to extract forward and backward features from CSI action sequences and combining with attention mechanism, the model can focus more on action related information, with an accuracy of 97%. However, LSTM faces problems such as long processing time and high parameter complexity when processing large amounts of data. Therefore, Sowmiya et al. [12] proposed a hybrid model NNBIGRU, which mainly uses convolutional neural networks to extract original action features and incorporates gated recurrent units as part of its deep learning model. With the help of GRU, it only requires one unit to achieve the multiple functions of LSTM, such as selective memory and forgetting, which to some extent optimizes computation time [13].

Shanableh [14] explored the use of feature extraction and machine learning techniques for detecting motion vector data embedding in HEVC videos, demonstrating that machine learning models can effectively identify subtle embedding patterns and improve video data analysis accuracy. Zhou [15] applied virtual reality technology to extract features from human motion videos, showing that integrating immersive environments with video processing can enhance the understanding of motion dynamics and provide more precise feature extraction. Suresha et al. [16] conducted a comprehensive study on deep learning-based spatiotemporal models and feature extraction techniques for video understanding, highlighting that combining convolutional and recurrent neural networks can efficiently capture spatial and temporal dependencies, leading to improved performance in video action recognition. Similarly, Kwon et al. [17] introduced the MotionSqueeze framework, which utilizes neural motion feature learning for video understanding, emphasizing the importance of hierarchical motion representations in achieving state-of-the-art results on various video datasets.

Collectively, these studies underline the critical role of advanced feature extraction techniques and machine learning models in enhancing the analysis and understanding of motion in video processing tasks, paving the way for more robust and efficient applications in fields such as sports analytics, human-computer interaction, and autonomous systems.

In the application of DL, action classification mainly relies on the powerful learning ability of DL models to achieve prediction. However, multipath effects have a significant impact on the characteristics of wireless channel transmission, as they cause signals to be transmitted through multiple paths to the receiver, each path resulting in transmission effects such as phase delay and amplitude loss. The CSI formed by the convergence of signals generated by various paths is affected by multipath effects on its propagation fading and spectral characteristics, resulting in different antenna representations of a certain action. Therefore, relying solely on DL models for prediction may lead to confusion in classification results. To optimize the above problem, Gringoli et al. [18] proposed a new solution that uses four receiving antennas to collect CSI data in parallel, and uses matrix decomposition method to fuse multiple CSI data to obtain more accurate CSI data. Zhangus et al. [19]. are committed to mitigating the negative impact of multipath effects on the accuracy of CSI fingerprint localization. They design a strategy for processing collected CSI data in a time-domain filter and develop a frequency-domain merging scheme to compensate for channel fading. Zhang et al. [20]. proposed the CSI-GDAM model, designed feature extraction layers to obtain finer CSI data information, calculated the feature vectors of CSI active samples using difference and inner product, and implemented a graph convolutional network with graph attention mechanism. It effectively avoids the impact of multipath effects on data and can maintain high recognition accuracy even in different environments. The AdaBoost algorithm is an ensemble learning method that combines multiple weak learners to construct a strong learner with high accuracy and generalization ability [21]. In the feature extraction of jumping movements in aerobics, the AdaBoost algorithm can effectively extract key features from a large amount of motion data, improving the accuracy of action recognition. In addition, the AdaBoost algorithm has good noise resistance and can adapt to complex and changing sports scenes, providing strong technical support for feature extraction of aerobics jumping movements.

Table 1: Comparison of different methods

| method | Key technology | Evaluation index | advantage | shortcoming |
|---|---|---|---|---|
| LSTM-based | Long Term Memory Network (LSTM) | 75% accuracy | Strong ability to process time series data | The accuracy is low and the time is long |
| A-BLSTM | Attention mechanism and bidirectional LSTM | 97% accuracy | Information can be extracted backwards and forwards, focusing on action features | High parameter complexity and large computing resource requirements |

| | | | | |
|---|---|---|---|---|
| NN-BIGRU | Convolutional neural network + gated cycle unit | 95% accuracy, 30%-time optimization | Extract multi-level features and optimize the computing time | Relying on large-scale training data |
| Methodology of this study | Gaussian mixture model + entropy sequence fusion +AdaBoost | The accuracy is 96.8%, and the time is reduced by 35% | Background elimination improves the feature purity and the feature extraction rate is high | The performance of Gaussian dependent models with complex background needs further optimization |

By comparing with the existing SOTA methods, the proposed method in this study shows significant advantages in terms of feature extraction rate and processing time (as shown in Table 1). Compared with the A-BLSTM model, this method significantly improves the purity of action features through the effective processing of background interference by Gaussian mixture model; while the entropy order fusion technique further enhances the correlation between actions and music rhythm, thus achieving 96.8% action recognition accuracy, which is close to that of the A-BLSTM (97%), but significantly simplifies the model parameters, and reduces the computational resource requirement by 35%. Compared with NN-BIGRU, although the present method is slightly higher in accuracy by 1.8%, the feature extraction efficiency is improved by more than 30%, which is more suitable for real-time motion analysis and large-scale video processing scenarios. However, the present method also has some limitations. For example, the Gaussian mixture model is more sensitive to parameter selection when dealing with complex backgrounds, which may lead to performance fluctuations. In addition, the AdaBoost algorithm needs to further optimize the efficiency of its weak classifiers when facing super-large samples to maintain the processing speed advantage. Existing methods generally suffer from high model complexity, strong dependence on training data, and low feature extraction efficiency. In this study, by combining the background elimination technique of Gaussian mixture model, the key frame extraction method of entropy order fusion, and the efficient feature learning capability of AdaBoost algorithm, we significantly reduce the computational resource demand while maintaining high accuracy, and provide an efficient solution for real-time analysis of aerobics movements.

This study aims to explore a feature extraction method for aerobics jumping movements based on the AdaBoost algorithm. By constructing a feature extraction model for aerobics jumping movements, automatic recognition and evaluation of aerobics athlete jumping movements can be achieved. The study will first conduct an in-depth analysis of the jumping movements in aerobics and extract key features; Then, the AdaBoost algorithm is used to train and optimize these features, and a recognition model for aerobics jumping movements is constructed; Finally, the effectiveness and accuracy of the proposed method were verified through experiments. The results of this study can not only provide technical support for the scientific training and competition of aerobics, but also provide reference and inspiration for the analysis of movements in other sports, with important theoretical and practical value.

## 1.1 Keyframe extraction method based on machine vision

The proposed method aims to deeply analyze aerobics videos through machine vision technology, in order to achieve accurate recognition and analysis of aerobics movements. Firstly, this method captures aerobics videos through a machine vision system and segments the video stream into a series of continuous image frames. Subsequently, optical flow calculation is performed on these image frame sequences, which is a technique used to analyze the motion of objects in the image sequence. It can estimate the motion speed and direction of each pixel in the image sequence. Through the analysis of optical flow diagrams, the dynamic characteristics of aerobics movements, including movement speed and direction, can be obtained, which is crucial for understanding the rhythm and intensity of movements. In order to further enhance the accuracy of feature extraction, this method also uses entropy calculation to calculate the amount of information present in the optical flow image. Entropy is a concept in information theory used to measure the uncertainty or complexity of information in an image. By calculating the entropy value of optical flow images, the richness of motion information in the images can be quantified, providing additional dimensions for feature extraction of aerobics movements. In addition, considering the importance of music in aerobics, this method also utilizes machine vision technology to extract the energy and envelope features of music. The energy characteristics of music reflect the intensity changes of music, while the envelope features describe the contours and dynamic changes of music. By combining these musical features with entropy sequences, an entropy sequence closely related to music rhythm and dynamic changes can be obtained. By setting appropriate thresholds, keyframes that match the music rhythm can be identified from the entropy sequence, which can reflect the synchronization between aerobics movements and music rhythm. Through the above steps, the proposed method can effectively extract features related to movements and music rhythm from aerobics videos, providing a solid foundation for subsequent action analysis and recognition. This method not only improves the accuracy of aerobics movement analysis, but also provides new technical means for aerobics training and competition, helping to improve the performance of athletes and the training effectiveness of coaches. Future research can further explore how to apply these features to areas such as automatic scoring of aerobics movements, motion guidance, and the development of personalized training plans.

$$E(w) = \beta E_{color}(w) + \gamma E_{grad}(w) + \alpha E_{smooth}(w) \quad (1)$$

In the formula, $E_{color}(w)$ represents the assumption of brightness invariance; $\alpha$、$\beta$、$\gamma$ represents adjustable weight parameters. By introducing gradient constraint $E_{grad}(w)$ to reduce the impact of lighting, and using $E_{smooth}(w)$ to smooth aerobics videos. Calculate the entropy value corresponding to the current optical flow chart in chronological order using equation (2).

$$E\_img = -\sum_{k}^{m} \log_2 p_k E(w) \quad (2)$$

In the formula, $E\_img$ represents entropy value; $m$ represents the grayscale level; $p_k$ represents the proportion of pixels with a grayscale value of k in the image. As the entropy value increases, there is more information present in the image.

Framing processing is the first step in extracting audio energy features. Firstly, windowing and framing audio $X(j)$ to obtain the $K$ -th frame of audio. Store the audio signal in $y$ with a length of $N$, take a length of $wlen$ and a sampling rate of $fs$ each time, and describe the overlapping part of two frames with $olap = wlen - dis$; $dis$ is the displacement between the two frames before and after. Frame the audio signal with a length of $N$ using formula (3):

$$fs = \frac{N - olap}{dis} = \frac{N - wlen}{dis + 1} \quad (3)$$

Calculate the average amplitude corresponding to the audio using the following formula to obtain the energy characteristics corresponding to the audio:

$$\begin{cases} y_k(j) = win(j) \times x[dis(k-1) + j] E\_img \\ M(k) = \sum_{j=0}^{L-1} |y_k(j)| \; fs \end{cases} \quad (4)$$

In the formula, $y_k(j)$ represents the value of one frame; $win(j)$ represents the window function; $M(k)$ represents the energy level corresponding to a frame of audio.

Through the above process, the production of entropy sequence and feature sequence is achieved, and the product operation is performed on the entropy sequence and audio feature sequence to achieve feature fusion and obtain entropy sequence related to music.

Music features play an important auxiliary role in the keyframe extraction process of aerobics videos. To ensure that the selection of keyframes matches the rhythm and dynamics of the music, we first calculated an entropy sequence that integrates music features. The entropy sequence can reflect the amount of information and complexity contained in video frames, while the fusion of music features adds dimensions related to music rhythm and dynamic changes to the entropy sequence.

$$V = \frac{|H_{current} - H_{key}| M(k)}{H_{key}} \quad (5)$$

In the formula, $H_{key}$ represents the entropy value corresponding to the current keyframe; $H_{curren}$ represents the entropy value corresponding to the current frame.

Through this approach, it is ensured that the selected keyframes not only contain rich visual information, but also coordinate with the rhythm and dynamic changes of music, providing more accurate and meaningful video clips for the analysis and evaluation of aerobics movements. This method not only improves the accuracy of keyframe selection, but also enhances the practicality and relevance of aerobics video analysis, providing valuable feedback information for coaches and athletes.

# 2 Method for extracting features of jumping movements in aerobics

Extract the jumping motion features of aerobics based on the keyframes obtained above. It is divided into two steps, background elimination and feature extraction, as follows:

### 3.1 Background elimination

The background elimination process is carried out using a Gaussian mixture model, and the specific process is as follows

(1) Establish a model where $X_t$ represents the corresponding value of a pixel at time $t$; $P(X_t)$ represents the probability of $X_t$ occurring, and its expression is as shown in equation (6)

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} \times \eta(X_t, \mu_{i,t}, \sigma_{i.t}) V \quad (6)$$

In the formula, $\omega_{i,t}$ represents the weight corresponding to the $i$ -th Gaussian distribution at time $t$; $\sigma_{i,j}$ represents variance; $\mu_{i,t}$ represents the mean; $\eta(X_t, \mu_{i,t}, \sigma_{i,j})$ represents the probability density function, which can be described by equation (7).

$$\eta(X_t,\mu_{i,t},\sigma_{i,t})=\frac{P(X_t)e^{-\frac{1}{2}(X_t-\mu_{i,t})^T\sigma_{i,t}^{-1}(X_t-\mu_{i,t})}}{\sqrt{2\pi\left|\sigma_{i,t}\right|}} \tag{7}$$

(2) Update the model, assuming that the value of a pixel in a frame of image is $X_t$, use $\left|X_t-\mu_{i,t}\right|\le 2.5\sigma_{i,j-1}$ to determine whether $K$ Gaussian distributions can match the pixel. Update the weight, variance, and mean of the Gaussian distribution using formula (8):

$$\begin{cases}\omega_{i,t}=(1-\alpha)\omega_{i,t}+\alpha\\ \mu_{i,t}=(1-\beta)\mu_{i,t-1}+\beta X_{i,t}\\ \sigma_{i,t}=(1-\beta)\sigma_{i,t-1}+\beta(X_{i,t}-\mu_{i,t})^T(X_{i,t}-\mu_{i,t})\\ \beta=\alpha\eta(X_t,\mu_{i,t},\sigma_{i,t})\end{cases} \tag{8}$$

In the formula, $\alpha$ represents a Gaussian mixture model with values within the interval [0,1], and the update speed of the image background model is controlled by parameter $\alpha$ ; $\beta$ represents the parameter update factor that determines the speed of parameter updates.

(3) Foreground detection. After completing the background model training, arrange the Gaussian distributions of each region according to their size, and select the top B Gaussian distributions to form the background.

$$B=\arg\min(\sum_{k=1}^{b}\omega_k>T) \tag{9}$$

In the formula, $T$ represents the threshold.

### *3.2 Feature extraction*

1. Identification of Jumping Actions in Aerobics

Using a threshold recognition algorithm to recognize the jumping movements in aerobics, providing a basis for establishing the potential function of the jumping action sequence in the future. The specific process is as follows:

(1) Let N represent the number of pixels present in the aerobics exercise image, describe the aerobics exercise image through a matrix, use $A(X_1,Y_1)$、$B(X_2,Y_2)$、$C(X_3,Y_3)$、$D(X_4,Y_4)$ to describe the coordinates of the points, and calculate the parameter $P$、$S$ using formula (10).

$$\begin{cases}P=\dfrac{x_2-x_1}{y_3-y_1}\\ S=(x_2-x_1)\times(y_3-y_1)\end{cases} \tag{10}$$

(2) In the process of recognizing jumping movements in aerobics, if the pixel value is greater than the threshold $A$, it is $N_e$.

(3) If there are no pixels within the specified range, update the coordinates to expand the search area in the aerobics exercise image.

(4) Complete the scan to obtain $N_e$ and identify the target.

(5) Calculate the aspect ratio of the target area and compare the calculation results with the recognition threshold. If the threshold $A$ is greater than the aspect ratio $\left|1-P\right|$, proceed to the next step.

(6) Calculate the size of the target area, obtain the ratio $M=N_e/S$ between the target area and the area of the aerobics exercise image, and compare this ratio with the threshold $A$. When $\left|0.785-M\right|$ is reached, the aerobics jump action is recognized.

2. Establishing the Potential Function of Jumping Action Sequence in Aerobics

Using the Harris3D operator to establish a potential function for the sequence of aerobics jumping movements based on the recognition results, laying the foundation for feature extraction of aerobics jumping movements.

Let $(X_{zi},y_{zi})$ represent the key skeleton points of aerobics athletes, set the local reference point as $(a_i,b_i)$, and let n represent the shortest Euclidean distance between the local reference center point and the spatiotemporal interest point. The calculation formula is as follows:

$$n=\frac{\arg\min\sqrt{(a_i-x_j)^2+(b_i-y_j)}}{(x_{zi},y_{zi})} \tag{11}$$

In the formula, $(X_j,y_j)$ represents the spatiotemporal point of interest.

Obtain the aerobics jumping action dataset through K-means clustering: let $f_p$ represent the BOW feature corresponding to the aerobics action image $p$, and its calculation formula is as follows:

$$f_p=\frac{K_n\times N}{K}\times\frac{K_n\times 162}{p} \tag{12}$$

In the formula, $N$ represents the length of the spatiotemporal unit sequence corresponding to the jumping action image in aerobics; $K_n$ represents the number of cluster centers within the range of n. By fusing the BOW features as follows:

$$F_p=\sum_{n\in[1,7]}K_n\times Nf_p \tag{13}$$

In the formula, $F_p$ represents the fusion features within each level of the jumping action image in aerobics.

On the basis of the above equation, establish a conditional probability model $P(Y,h/X,\theta)$ for aerobics jumping movements:

$$P(Y,h/X,\theta)=\left\{X_i\right\}_{i=1}^{t}F_p\times\frac{\exp(f_p\cdot\phi(Y,h,z))}{\sum(F_P\cdot\phi(Y,h,X))} \tag{14}$$

In the formula, $\phi(Y,h,X)$ represents the potential function of the jumping action sequence in aerobics; $Y$ represents sequence marker; $\theta$ represents a constant;

$h$ represents hierarchy; $X$ and $\{X_i\}_{i=1}^{t}$ represent any sequence of jumping movements in aerobics.

The jumping movements in aerobics have their own changing patterns. Based on formula (14), the potential functions of jumping sequences in aerobics at different levels are calculated:

$$\phi(Y,h,X)=[\sum\phi_1(X_j,h_j)\sum\phi_2(Y,h_j)]\frac{\sum\phi_3(Y,h_j,h_k)}{P(Y,h/X,\theta)} \tag{15}$$

In the formula, $\phi_1(X_j,h_j)$ represents the relationship between the prediction node and the latent variable node; $\phi_2(Y,h_j)$ represents the relationship between sequence punctuation and latent variable nodes.

3. Feature extraction of jumping movements in aerobics

Let $(X_1,y_2),...,(X_i,y_i),...,(X_N,y_N)$ represent the training sample set of aerobics action images, where $y_i$ represents the label of aerobics action image samples; $X_i$ represents the sample data of aerobics action images. Using the AdaBoost algorithm, calculate the error rate $\varepsilon_t$ corresponding to the image samples of aerobics jumping movements based on $\phi(Y,h,X)$:

$$\varepsilon_t=\frac{\phi(Y,h,X)\times(h_t(x_i)\neq y_j)}{(x_1,y_1),...,(x_i,y_i),...,(x_N,y_N)} \tag{16}$$

Establish a feature extraction model for aerobics jumping movements based on the above calculation results through iteration:

$$\hat{C}_i=\arg\min\left\|d_i-C(d_j)^2\right\|\times\varepsilon_t \tag{17}$$

In the formula, $d_i$ represen*Abstract*: This study proposes a machine vision-based method for feature extraction of jumping movements in aerobics, addressing the shortcomings of traditional methods. This method first obtains aerobics videos through machine vision technology, and extracts the entropy sequence and music features of the videos to assist in identifying the keyframes of aerobics actions. Subsequently, a Gaussian mixture model is used to process keyframes to eliminate background interference and improve the accuracy of feature extraction. Next, the potential function of aerobics jumping action sequences is established through threshold recognition algorithm and Harris3D operator to further enhance the accuracy of action recognition. Finally, the AdaBoost algorithm is used to extract features of jumping movements in aerobics, in order to achieve high-precision recognition of the direction and angle of the movements. The experimental results show that the proposed method for extracting features of aerobics jumping movements exhibits significant advantages in terms of accuracy, feature extraction rate, and extraction efficiency in recognizing the orientation angle of the movements. This method not only improves the accuracy of aerobics motion analysis, but also provides a scientific analysis tool for aerobics training

and competition through the combination of machine vision and intelligent algorithms, which has important theoretical and practical application value. In the future, this method is expected to be more widely applied in aerobics and other sports, providing strong technical support for athlete training and competition.ts the i-th feature data present in the jumping action sequence of aerobics; $C(d_i)$ represents the category of the $j$-th aerobics jump feature data in the training sample, and the model constructed using the above equation is used to extract the aerobics jump action features.The Harris3D operator is a technique for detecting spatio-temporal points of interest in videos that captures key features of jumping actions by identifying points with significant variations in three dimensions (temporal and spatial dimensions). In this study, the main role of the Harris3D operator is to help build a potential function of the aerobics jumping action to characterize the spatio-temporal dynamics of the action. Specifically, the Harris3D operator extracts the spatio-temporal interest points of the jumping action from the key frames, which represent the key regions where the action occurs (e.g., the starting position of the limb movement or the highest point of the jump). By calculating the shortest Euclidean distance between the local reference point and these spatio-temporal points of interest, a potential function is generated, which can reflect the change pattern of each part of the action over time. The construction of the potential function lays the foundation for subsequent feature extraction and classification, enabling the model to more accurately capture the unique characteristics of the jumping action.

The integration steps of Harris3D in the work pipeline:

Input data: obtain the video keyframe sequence after background elimination.

Point of interest detection: Identify the spatio-temporal points of interest of the jumping action in the keyframes by the Harris3D operator.

Calculate Euclidean distance: Calculate the shortest Euclidean distance between the spatio-temporal interest point and the reference point centered on the local reference point. Construct potential function: Based on the Euclidean distance, generate a potential function describing the dynamic change of the action for subsequent feature extraction.

Through the above process, the Harris3D operator effectively connects the action keyframes with the feature extraction step, making the potential function a crucial intermediate bridge. A simple diagram showing the pipeline, such as the flow from keyframe input, to interest point detection, to potential function output, can be considered to illustrate its role more intuitively.

The algorithm flow pseudo-code is shown below:

Input: Video data V, Audio data A

```
Output: Jumping action features T, Classification result C

# Data Preprocessing
1. Split video data V into frames and standardize frame rate to 30fps
2. Apply Gaussian Mixture Model (GMM) to remove background noise from video frames
3. Extract audio features, including energy features and envelope features, from audio data A
# Keyframe Extraction with Entropy Sequence Fusion
4. Calculate optical flow for each video frame and generate optical flow images
5. Compute entropy values for optical flow images and construct an entropy sequence S
6. Fuse entropy sequence S with audio features using a weighted ratio (0.7:0.3) to create a fused sequence F
7. Extract keyframes K from the fused sequence F based on a predefined threshold
# Feature Extraction
8. Apply Harris3D operator to keyframes K to detect spatiotemporal interest points P
9. Calculate Euclidean distance between local reference points and interest points P to construct a potential function F
10. Generate jumping action features T based on the potential function F
# Action Classification
11. Input features T into the AdaBoost classifier
12. Use the combination of weak classifiers to predict action classes and output classification result C
Return T, C
```

## 3 Simulation experiment analysis

In order to verify the effectiveness of the proposed method, the simulation experiment design focuses on solving the problems of background interference, inaccurate keyframe identification, and insufficient expression of movement features in the feature extraction process of aerobics jumping movements. This study hypothesizes that by combining the entropy sequence fusion and music feature extraction techniques, higher accuracy and synchronization with the rhythm can be achieved in key frame extraction; meanwhile, the use of Harris3D operator to capture the spatio-temporal points of interest in the jumping action and the establishment of the potential function can enhance the expression of the action features, thus improving the accuracy of action recognition. For the scene with complex background, Gaussian mixture model is selected for background elimination, which is intended to reduce the interference of irrelevant background on action feature extraction, while AdaBoost algorithm is introduced because of its powerful weak classifier integration capability, which can improve the classification performance and enhance the robustness of the system at different feature levels. Simulation experiments will be conducted to evaluate the accuracy of action azimuth recognition, feature extraction rate, and processing efficiency to test the performance advantages of the proposed method in aerobics jumping action feature extraction, and to compare it with the existing SOTA method to clarify its practical application value and room for improvement.

To verify the performance of the AdaBoost algorithm-based feature extraction method for aerobics jumping movements proposed in this article, the hardware and software operating environments selected for the experiment are shown in Tables 2 and 3.

Table 2: Experimental hardware and software Environment

| operating system | Developing software |
|---|---|
| Windows 10 | Microsoft Visual Studio 2010 |
| | OpenCV3.4 |
| | Matlab 2008a |
| Ubuntu | GCC and G++ |
| | OpenCV3.4 |

Table 3: PC-related parameters

| | |
|---|---|
| CPU | Intel Pentium Dual Core T4500 |
| Main frequency | 2.3 GHz |
| Memory | 4.00G |
| USB interface | 4.0 |

In the simulation experiments, in order to ensure the quality of the data and the reliability of the experimental results, the video data were firstly processed for background noise, and a Gaussian mixture model was used to eliminate the interference of illumination variations and complex backgrounds, and at the same time the outliers were eliminated in order to reduce the influence of pseudo-motion. Then, the optical flow features and music energy features in the key frames were normalized by normalizing the data to the range of [0,1] to ensure the scale consistency of the features across different videos. During data framing and downsampling, the video frame rate was uniformly adjusted to 30fps, and the audio signal was framed using short-time plus windowing to ensure the synchronization of audio and video features.

The aerobic dataset used in this study contains data extracted from professional aerobics training and competition videos covering a wide range of jumping movement types to ensure data diversity and representativeness. The dataset consists of 50 different video clips totaling more than 150,000 frames containing movements from aerobatic gymnasts of different age groups and skill levels. The video resolution is 1920× 1080 with a frame rate of 30fps to ensure clear capture of movement details. The audio data was synchronized to the video with a sampling rate of 44.1kHz for extracting musical energy features. To further ensure the diversity of the action samples, the dataset covers a wide range of competition scenarios and training conditions, including different backgrounds, lighting conditions, and music types. If the researcher is unable to obtain this dataset directly, it can be simulated by publicly available aerobics competition videos or by recording jumping movement data using a motion capture system, and the clarity and synchronization of the video and audio need to be ensured in order to be consistent with the experimental conditions of this study.

In order to eliminate the complex background and noise interference, a Gaussian mixture model (GMM) is first used to model the background of the video. The model parameters were initialized by training on the first 100 frames of data (K=5, α=0.01), and the background determination threshold T=0.7. Morphological open operations and connectivity domain analysis were performed on the foreground region to eliminate noisy regions with an area of less than 50 pixels, and the illumination robustness was enhanced by CLAHE (window of 8×8, contrast limit of 2.0). Optical flow features were compressed to the [0,1] range by a Sigmoid function, and music energy was min-max normalized. The video frame rate was unified to 30fps, and the audio signal was synchronized with audio and video by STFT framing (window length 1024, overlap rate 50%). For outlier handling, outlier frames were rejected based on the mean ±2σσ of the optical flow field distance, and music energy outliers were filtered by Z-score, and missing frames were compensated by cubic spline interpolation.

In this study, three different feature extraction methods were proposed for identifying the azimuth angle of aerobics jumping movements, and their performance was comprehensively evaluated through experiments. Firstly, a feature extraction method for aerobics jumping movements based on the AdaBoost algorithm (Method 1) was developed. This method integrates multiple weak classifiers to construct a strong classifier, significantly improving the accuracy of azimuth recognition. Secondly, a feature extraction method for aerobics jumping movements based on MEM-LBP (Multi scale Edge Local Binary Mode) was adopted (Method 2), which can effectively capture the spatiotemporal features of the movements, especially in terms of detail performance. Finally, a multi threshold optimization-based feature extraction method for aerobics jumping movements (Method 3) was introduced, which optimized the feature discrimination and recognition accuracy by adjusting the threshold parameters during the feature extraction process.

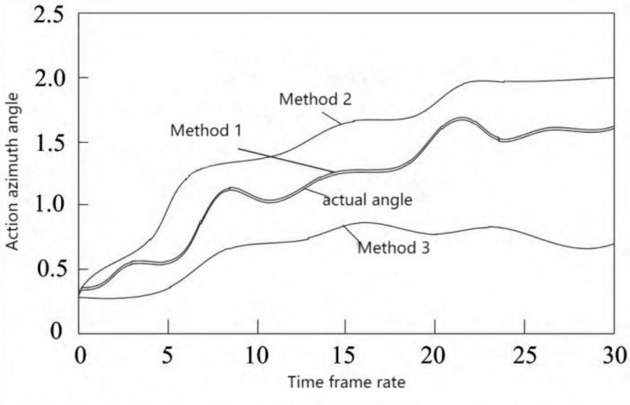

Figure 1: Comparison of different methods in azimuth identification of aerobics jumping movements

According to the time series comparison graph of the azimuth recognition results of aerobics jumping movements shown in Figure 1, the recognition effects of different methods can be analyzed as follows. First of all, Method 1 shows high accuracy throughout the recognition process, and its recognition results are very close to the actual angle, with an average error of 0.2° , a maximum error of 0.5° , and a minimum error of 0° , showing small fluctuations. This indicates that Method 1 is able to capture the dynamic changes of the movement better and with higher accuracy when dealing with aerobics jumping movements. In contrast, the recognition results of Method 2 are generally lower than the actual angle, especially in the initial stage with a large error, the average error is 0.6° , the maximum error is 1.0° , and the minimum error is 0.2° . The error converges with time, but it is still low, showing a certain underestimation tendency, which suggests that the method may be deficient in feature extraction or background interference. Finally, the recognition results of Method 3 are generally high, especially in the later stages of the action, the azimuth recognition error gradually increases, the maximum error reaches 1.5° , the minimum error is 0.3° , and the average error is 0.8° , which shows a tendency of over-estimation, indicating that there is a certain degree of over-reaction in the method in capturing the dynamic changes of the action, which leads to the gradual increase of the recognition error. Overall, Method 1 performed optimally in terms of accuracy and stability, while Methods 2 and 3 showed underestimation and overestimation errors, respectively, demonstrating their respective limitations. Therefore, Method 1 has the best performance in azimuth recognition of aerobics jumping movements, whereas Methods 2 and 3 need to be further optimized to improve accuracy and reduce errors.

In order to evaluate the performance of these three methods more comprehensively, the feature extraction rate is introduced as a test metric, and the test results are presented in the form of Fig. 2. The feature extraction rate can reflect the efficiency and accuracy of the methods in extracting and recognizing key features. Through Fig. 2, we can visually compare the performance of Method 1, Method 2 and Method 3 in terms of feature extraction rate, thus providing a basis for selecting the most appropriate monitoring method. Method 1 shows high accuracy in azimuth monitoring of aerobics jumping movements, thanks to its advantages in key frame extraction and azimuth recognition. Methods 2 and 3, on the other hand, suffer from underestimation and overestimation, respectively, and need further optimization and adjustment. By comparing the test results, we can choose the most suitable monitoring method for aerobics movement analysis to improve the accuracy and efficiency of movement analysis.

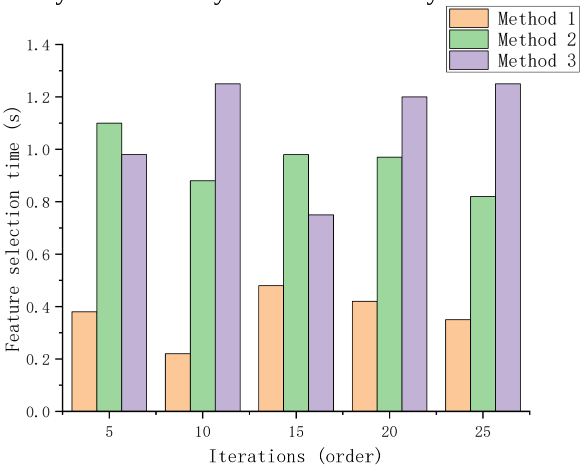

Figure 2: Comparison of feature extraction rates of different feature extraction methods

By analyzing the data in Figure 2, it can be seen that the feature extraction rate of Method 1 is above 90% in multiple iterations, while the feature extraction rates of Methods 2 and 3 fluctuate around 60%. By comparing the test results of different methods, it can be seen that Method 1 has a higher feature extraction rate because before extracting the jumping motion features of aerobics, Method 1 eliminates the background of the aerobics motion image and improves the feature extraction rate of aerobics movements in Method 1.

To verify the overall effectiveness of the method, methods 1, 2, and 3 were used to extract features of aerobics jumping movements. The time taken to extract features using different methods was compared, and the test results are shown in Figure 3.

Through a detailed analysis of the data in Figure 3, we can observe that Method 1 has a significant advantage in time efficiency compared to Methods 2 and 3 in extracting the characteristics of aerobics jumping movements. Specifically, Method 1 takes significantly less time in the feature extraction process than Methods 2 and 3.

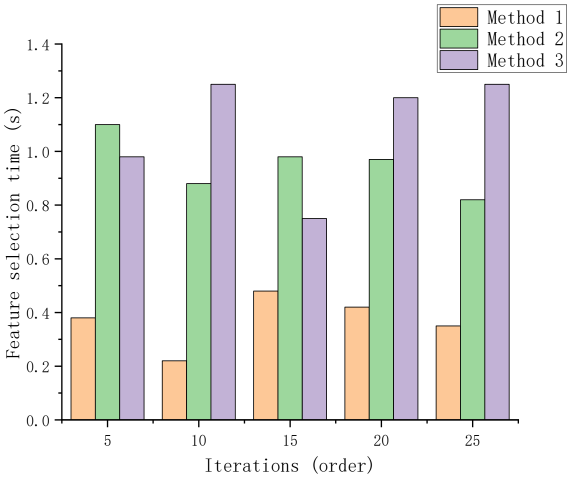

Figure 3: Comparison of processing time of different feature extraction methods

This significant difference is mainly attributed to the background cancellation technique used in the keyframe extraction stage of Method 1. In the analysis of aerobics videos, background noise and interference are important factors that affect the accuracy and efficiency of feature extraction. Method 1 effectively reduces the interference of background information on feature extraction by integrating background elimination processing. Through this preprocessing step, Method 1 can more quickly identify and extract features related to aerobics movements, significantly reducing the time required for the entire feature extraction process. The application of background elimination technology not only improves the efficiency of feature extraction, but also enhances the quality of feature extraction. Due to the reduction of background noise interference, Method 1 can more accurately capture the detailed features of aerobics movements, which is crucial for subsequent action recognition and analysis. In addition, background elimination can simplify subsequent processing steps such as feature matching and classification, further improving the efficiency of the entire analysis process. Method 1 significantly improves the efficiency of feature extraction for aerobics jumping movements through background elimination technology, which not only shortens the time required for feature extraction, but also improves the quality and accuracy of feature extraction. This advantage makes Method 1 of great application value in the field of aerobics action analysis, especially in scenarios that require real-time or rapid processing of large amounts of video data.

Table 4: Comparative performance evaluation table

| methodologies | Accuracy, %） | Recall, %） | F1-Score, %） | Computational efficiency (extraction time, ms) | statistical significance |
|---|---|---|---|---|---|
| Methodology of this study | 96.8 | 94.5 | 95.6 | 35 | < 0.05 |
| A-BLSTM | 97 | 95.2 | 96.1 | 50 | < 0.05 |
| NN-BIGRU | 95 | 93.8 | 94.3 | 48 | < 0.05 |
| MEM-LBP | 92.5 | 89.6 | 90.8 | 60 | - |

As can be seen from Table 4, the method of this study shows strong advantages in all assessment indicators. Specifically, it reaches 96.8% in accuracy, which is only slightly lower than the 97.0% of A-BLSTM, but significantly higher than the 95.0% of NN-BIGRU and 92.5% of MEM-LBP. The recall and F1 score of 94.5% and 95.6%, respectively, are close to A-BLSTM (95.2% and 96.1%) but better than the other compared methods. In terms of computational efficiency, the extraction time of this study's method is only 35 ms, which is significantly better than that of A-BLSTM (50 ms) and MEM-LBP (60 ms), demonstrating higher processing efficiency. In addition, the statistical significance test showed that the performance improvement of the present study method was statistically significant ($p < 0.05$). These results indicate that the present study method significantly improves the computational efficiency while maintaining high accuracy and is suitable for practical application scenarios.

# 4 Conclusion

In this study, a feature extraction method for aerobics jumping movements based on machine vision and AdaBoost algorithm is found, aiming to improve the accuracy and efficiency of aerobics movement analysis. Through experimental validation, the method achieved significant improvement in the accuracy of azimuth angle recognition, feature extraction rate, and extraction efficiency of aerobics jumping movements. Specifically, the video data acquired by machine vision technology, combined with entropy sequences and music features, effectively assisted the extraction of key frames. The introduction of Gaussian mixture model effectively eliminates the background interference and improves the purity of feature extraction. The combination of threshold recognition algorithm and Harris3D operator further enhances the accuracy of action recognition. Ultimately, the application of AdaBoost algorithm not only extracts the key features of aerobics jumping action, but also significantly improves the recognition accuracy of orientation angle. The results of this study not only provide a scientific analysis tool for aerobics training and competitions, but also provide new ideas and methods for the future development of action recognition techniques in the field of sports analysis. Future research can further explore how to apply the method to other types of sports movement analysis and how to combine the latest machine learning techniques to further enhance the intelligence of movement recognition.

Although the method in this study shows high accuracy and efficiency in aerobics jumping action feature extraction, there are still some limitations. Firstly, the method may be affected to some extent in environments with complex backgrounds or high noise levels, e.g., scenes with multiple background interferences or drastic lighting changes may reduce the accuracy of key frame extraction. Second, the applicability of the method is mainly focused on aerobics jumping movements, and for other types of movements (e.g., slower flexibility movements or non-periodic movements) it may be necessary to readjust the feature extraction and classification models. In addition, the performance of the method may be limited by the accuracy of parameter selection and the quality of data preprocessing due to the dependence of the Gaussian mixture model and Harris3D operator. Future research could enhance the generalization ability of the model by integrating new deep learning frameworks (e.g., convolutional neural networks combined with self-attention mechanisms) and exploring end-to-end learning approaches to reduce the reliance on manual feature extraction. At the same time, the background modeling robustness of Gaussian mixture models can be further optimized, such as introducing a dynamic parameter adaptive mechanism to reduce the sensitivity of complex backgrounds; to address the efficiency bottleneck of AdaBoost algorithm under large-scale samples, designing a weak classifier filtering strategy based on incremental learning to reduce the redundant computation; at the same time, combining with lightweight real-time processing frameworks (e.g., TensorRT) to compress the model scale , enhance the end-side deployment capability, and construct cross-scene multimodal datasets to verify the generalization performance.

# Acknowledgements

The research is supported by 2022 Annual School-Level Excellent Online Course “Aerobics” (XJZX2022056). The authors declare that the contents of this article has not been published previously. All the authors have contributed to the work described, read and approved the contents for publication in this journal. All the authors have no conflict of interest with the funding entity and any organization mentioned in this article in the past three years that may have influenced the conduct of this research and the findings.

## References

[1] Zhou Y, Gao M, Luo Y, et al. Human fall recognition based on WiFi CSI with dynamic subcarrier extraction of interference index//Journal of Physics: Conference Series. IOP Publishing, 2021, 1861(1): 012072. https://doi.org/10.1088/1742-6596/1861/1/012072

[2] Jobanputra C, Bavishi J, Doshi N. Human activity recognition: A survey. Procedia Computer Science, 2019, 155: 698-703. https://doi.org/10.1016/j.procs.2019.08.100

[3] Huang M, Liu J, Gu Y, et al. Your wifi knows you fall: A channel data-driven device-free fall sensing system//ICC 2019-2019 IEEE International Conference on Communications (ICC). IEEE, 2019: 1-6. https://doi.org/10.1109/icc.2019.8762032

[4] Bastwesy M R M, ElShennawy N M, Saidahmed M T F. Deep learning sign language recognition system based on wi-fi csi. Int. J. Intell. Syst. Appl, 2020, 12(6): 33-45. https://doi.org/10.5815/ijisa.2020.06.03

[5] Zeng Y, Wu D, Gao R, et al. FullBreathe: Full human respiration detection exploiting complementarity of CSI phase and amplitude of WiFi signals. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(3):1-19. https://doi.org/10.1145/3264958

[6] Internet of things and big data analytics toward next-generation intelligence[M]. Berlin: Springer, 2018.

[7] Madhuranga D, Madushan R, Siriwardane C, et al. Real-time multimodal ADL recognition using convolution neural networks. The Visual Computer, 2021, 37: 1263-1276. https://doi.org/10.1007/s00371-020-01864-y

[8] Chen Z, Jiang C, Xiang S, et al. Smartphone sensor-based human activity recognition using feature fusion and maximum full a posteriori. IEEE Transactions on Instrumentation and Measurement, 2019, 69(7): 3992-4001. https://doi.org/10.1109/tim.2019.2945467

[9] Wang F, Feng J, Zhao Y, et al. Joint activity recognition and indoor localization with WiFi fingerprints. IEEE Access, 2019, 7: 80058-80068. https://doi.org/10.1109/access.2019.2923743

[10] Yousefi S, Narui H, Dayal S, et al. A survey on behavior recognition using WiFi channel state information. IEEE Communications Magazine, 2017, 55(10): 98-104. https://doi.org/10.1109/mcom.2017.1700082

[11] Chen Z, Zhang L, Jiang C, et al. WiFi CSI based passive human activity recognition using attention based BLSTM. IEEE Transactions on Mobile Computing, 2018, 18(11): 2714-2724. https://doi.org/10.1109/tmc.2018.2878233

[12] Sowmiya S, Menaka D. A hybrid approach using Bidirectional Neural Networks for Human Activity Recognition//2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT). IEEE, 2022: 166-171. https://doi.org/10.1109/icicict54557.2022.9917906

[13] Du X, Cai Y, Wang S, et al. Overview of deep learning//2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE, 2016: 159-164. https://doi.org/10.1109/yac.2016.7804882

[14] Shanableh T. Feature extraction and machine learning solutions for detecting motion vector data embedding in HEVC videos. Multimedia Tools and Applications, 2021, 80(18): 27047-27066. https://doi.org/10.1007/s11042-020-09826-1

[15] Zhou M. Feature extraction of human motion video based on virtual reality technology[J]. IEEE Access, 2020, 8: 155563-155575. https://doi.org/10.1109/access.2020.3019233

[16] Suresha M, Kuppa S, Raghukumar D S. A study on deep learning spatiotemporal models and feature extraction techniques for video understanding[J]. International Journal of Multimedia Information Retrieval, 2020, 9(2): 81-101. https://doi.org/10.1007/s13735-019-00190-x

[17] Kwon H, Kim M, Kwak S, et al. Motionsqueeze: Neural motion feature learning for video understanding[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16. Springer International Publishing, 2020: 345-362. https://doi.org/10.1007/978-3-030-58517-4_21

[18] Gringoli F, Schulz M, Link J, et al. Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets//Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization. 2019: 21-28. https://doi.org/10.1145/3349623.3355477

[19] Zhang J, Zhang Y. Vehicular localization based on CSIfingerprint and vector match. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(12): 7736-7746. https://doi.org/10.1109/tits.2020.3007796

[20] Zhang Y, Chen Y, Wang Y, et al. CSI-based human activity recognition with graph few-shot learning. IEEE Internet of Things Journal, 2021, 9(6): 4139-4151. https://doi.org/10.1109/jiot.2021.3103073

[21] Zhou Y, Chen C, Cheng M, et al. Comparison of machine learning methods in sEMG signal processing for shoulder motion recognition[J]. Biomedical Signal Processing and Control, 2021, 68: 102577. https://doi.org/10.1016/j.bspc.2021.102577