

Retraction

Retracted: Research on Sports Video Image Based on Clustering Extraction

Mathematical Problems in Engineering

Received 25 July 2023; Accepted 25 July 2023; Published 26 July 2023

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

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] J. Ma, "Research on Sports Video Image Based on Clustering Extraction," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9996782, 9 pages, 2021.

Research Article

Research on Sports Video Image Based on Clustering Extraction

Jun Ma 

Sports Department, Guizhou University of Finance and Economics, Guiyang 550025, Guizhou, China

Correspondence should be addressed to Jun Ma; majun@mail.gufe.edu.cn

Received 2 April 2021; Revised 7 May 2021; Accepted 17 May 2021; Published 14 June 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Jun Ma. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Today, with the continuous sports events, the major sports events are also loved by the majority of the audience, so the analysis of the video data of the games has higher research value and application value. This paper takes the video of volleyball, tennis, baseball, and water polo as the research background and analyses the video images of these four sports events. Firstly, image graying, image denoising, and image binarization are used to preprocess the images of the four sports events. Secondly, feature points are used to detect the four sports events. According to the characteristics of these four sports events, SIFT algorithm is adopted to detect the good performance of SIFT feature points in feature matching. According to the simulation experiment, it can be seen that the SIFT algorithm can effectively detect football and have good anti-interference. For sports recognition, this document adopts the frame cross-sectional cumulative algorithm. Through simulation experiments, it can be seen that the grouping algorithm can achieve a recognition rate of more than 80% for sporting events, so it can be seen that the recognition algorithm is suitable for recognizing sports events videos.

1. Introduction

In recent years, with the improvement of people's living standards, more and more attention has been paid to sports and video processing has been gradually deepened in the field of sports. A large number of sports and national fitness information, as well as video and image forms, are stored in various fitness guidance systems. In order to better promote public fitness and facilitate learning and viewing, image segmentation motion video player has become a hot topic of digital image processing. To overcome the slow split, irregular movement, and susceptible to the influence of light of the shortcomings of the moving object, the researcher based clustering presents a motion video sequence moving foreground object extraction algorithm. Experiments show that the algorithm is effective, simple, workable, and has less calculation, and the effect is satisfactory.

Motion video object exercise posture has relatively random and blurred image change tendency, and the divided region is not easily determined. Moving video images may present segmentation exercise posture, unnecessary region of the object, pixel area divides into foreground and background, and extracts the movement of the object.

Launched in 2000, with an increase in MPEG-4 video content semantic search feature, you can split the background and foreground images into different semantic objects. Coding efficiency is improved, but the noise in the encoding process is not quickly eliminated. Temporal segmentation and frequency domain are the first segmentation method for segmentation of moving video images presented specifically. After a lot of experiments, both methods can not accurately describe the attitude of the object, and the image segmentation is not clear. Therefore, research on sports video images based on clustering extraction has greatly changed sports videos and will help make better use of sports video analysis and images.

As research deepens, sports video image segmentation technology has made great strides. In 2009, David and Zhang Shensheng used a binary grouping method, which showed a good recognition rate in various images and achieved good results in the experiment. However, there are also some weaknesses in the method. The result is not obvious when the brightness of the object surface is affected by many lighting factors, such as dimming reflection, high light reflection, and fuzzy texture. Fan Cuihong proposed a segmentation method of video moving objects based on

regional differences. The RGB space of video image is transformed into HSV space. The closed contour of moving object is extracted by chroma, saturation, and brightness. According to the moving area, background area, and occlusion area of video, the edge of moving object is detected by edge detection operator, and finally the moving object in video is segmented. Experiments show that the improved algorithm can improve the segmentation accuracy and meet the real-time requirements. Ouyang Yi proposed a method based on Markov chain Monte Carlo (MCMC) to track human posture in monocular video images. Firstly, the projection maps of human appearance in basic human motion database acquired by motion capture equipment were clustered under different perspectives. Using HOG to detect human body in monocular video images can segment the position of human limbs more accurately. Finally, the appearance model of the three-dimensional human posture reasoning algorithm is used to analyze each frame, and then the time-constrained analysis model is used to track the target. Constraint graph-driven MCMC and basic action library are combined to construct a model for video data modeling, and the model is applied to data driven online behavior recognition to improve human pose modeling ability. Zhang Jiawen et al. proposed a convenient and practical method for human motion tracking and motion reconstruction in video and initially achieved the goal of obtaining human motion data from video resources such as video surveillance and video recordings. This paper is a useful attempt to track and acquire human motion in monocular video, and some satisfactory preliminary research results have been achieved. In the last part of this paper, the author puts forward his own opinions on the problems of human motion tracking and motion reconstruction methods and their further improvement. Wu Tianai, Yang Ling, and others have proposed a moving human body detection algorithm based on space-time combination in color environment. The algorithm combines temporal segmentation with spatial segmentation to obtain the moving human body with precise edges and eliminates the shadow of the moving human body. The experimental results show that the above algorithm can detect the moving human body from the color image sequence in real time and effectively, eliminate the shadow of the moving human body, and finally detect the moving human body is the color. With more and more people researching sports video images, many achievements have been made in recent years.

Because of the complexity of the algorithm and the great difference between the segmentation results and the reality, the application of the segmentation results is limited. The main reason is that the change of the blur factor of the image, the uncertainty of the segmentation, and the time-consuming information loss. This loss is often due to the boundary information generated in the classification process. Clustering extraction as a clustering method has been successfully applied in many research classifications. Liu Guodong et al. proposed a threshold adaptive online clustering color background reconstruction algorithm and objectively evaluated the reconstructed color background. Finally, the background subtraction method was used to

extract moving objects. Jiang Yuan et al. put forward that clustering is an important tool for data mining. According to the similarity of data, the database is divided into several categories, in which the data should be as similar as possible. Based on possibilistic C-mode clustering, the main tone and subtone were selected to describe the features of video images so that the key frames of video could be extracted directly without shot segmentation. Experiments show that this method can effectively extract the most representative key frames according to the complexity of video content and has high timeliness. Leskovec et al. proposed the K_SC clustering algorithm for topic time series in 2010, which has high accuracy and can better describe the inherent trend of topic development. However, the K_SC algorithm is highly sensitive to the center of the initial class matrix and has high time complexity, which makes it difficult to apply in the actual high-dimensional large data sets.

Due to the advantages of the cluster extraction algorithm [1], cluster extraction has been used as a classification tool in many areas of clustering research, and good results have been achieved. The extraction algorithm proposed sequential cross-sections of cross-based video frames that form moving foreground objects. The simulation experiment is performed under the simulation environment of MATLAB 2014 B. The experiment proves that the algorithm is efficient, simple, and feasible, with fewer calculations and can have more satisfactory results. Due to the inherent noise in the original frame, it is necessary to reduce the noise in the original video sequence to improve the accuracy and efficiency of moving objects in video segmentation, to reduce noise, and to enhance the effect of calculating the frame difference. This algorithm uses the most common square median filter module for pre-processing, which is fast and feasible, can reduce the loss of image details, optimize image quality, and be suitable for the next segment. It can also better protect the edges of the image and effectively remove noise. Research on sports video images based on clustering extraction and these fast 3D camera modeling broadcast court network sports videos, deep learning models and transfer learning sports videos, mixed reality systems, and hyperparameter optimization based on convolutional neural networks. In comparison with the sports video summary scene classification algorithm based on classification and transfer learning, the clustering extraction algorithm can be more convenient, fast, safe, and reliable.

2. Proposed Method

2.1. Binary Processing Based on Sports Video Image. Among many image segmentation methods, binarization is a simple and effective method. The purpose of image binarization [2] is to separate the moving object and background from the image and to provide a basis for subsequent classification, detection, and recognition. Typically, the method of threshold segmentation [3] is used for binarization. Its principle is to select an appropriate threshold to determine whether each pixel in the image belongs to the target or the background by making use of the

difference between the object and the background in the image so as to get the target to be detected.

Set the input image as $f(x, y)$ and the output image as $f_b(x, y)$, then

$$f_b(x, y) = \begin{cases} 1, & f(x, y) \geq T, \\ 0, & f(x, y) < T, \end{cases} \quad (1)$$

where T represents the threshold of the binarization process. In actual processing, 255 is used to represent the background and 0 is used to represent the moving target. The process of binarization is relatively simple. The key problem is the selection method of binarization. In this paper, the OTSU algorithm [4] is used to improve the selection of binarization image threshold.

2.2. Morphological Processing of Sports Video Image. The binary image has some misjudgement points; that is, there are a small number of “burrs” or “holes,” which requires further refinement. In order to get more accurate segmentation results, this paper chooses appropriate structural elements and uses mathematical morphology to filter the corrosion and expansion. In fact, it uses image open and close operation.

In corrosion operation, the result of etching a binary image is to narrow the edge of the image and shrink inward, which seems to be eroded by the surface. Its principle is to define a subimage whose size is negligible and relative to the image to be processed as a structural element. Typically, a $2 * 2$ or $3 * 3$ pixel size template is selected, which specifies a pixel in the template as the origin and assigns a value (1 or 0) to each position in the template. The structure elements are used to scan the image to be processed point by point and perform matching operations. Whenever a subimage identical to the structural element is found (no matching of the zero pixels in the structural element), the location of the corresponding pixel in the subimage corresponding to the original position of the structural element is marked. The set of all such pixels is the result of the corroded operation of the binary image, which is defined as formula (2)

If I is the image to be corroded and B is the selected structure “probe,” then the definition of target image I corroded by probe B [5] is as follows:

$$I \ominus B = \{x | (B)_x \supseteq I\}. \quad (2)$$

Among them, X represents the displacement of set translation, \ominus represents the operator of corrosion operation, and $(B)_x$ represents the translation of structural elements.

Morphological corrosion means that every time the translation structural element detects the image to be processed; if a subimage identical to the structural element is found, the location of the pixels in the subimage and the origin of the structural element is marked. The set of all the marked pixels is the result of corrosion. In fact, in the image to be processed, the original pixels of the subimage with exactly the same shape of the structural elements are marked and retained. Expansion operation corresponds to

corrosion operation. Expansion processing of a binary image often enlarges the edge of the image. Therefore, if there are black spots in the white foreground area of the image or two white areas are blocked by very thin black lines, then the black pore in the corroded image will be filled into a white image block which is similar to the surrounding image block, and the two image blocks which are not connected by themselves will also become a complete connected block. The principle of image expansion is to define a subimage whose size is negligible relative to the image to be processed as a structural element. Typically, a template with 2×2 or 3×3 pixel size is selected to specify a pixel in the template as the origin. And a value (1 or 0) is assigned to each position in the template, the image is scanned to be processed point by point, and matching operation is performed by using structural elements. Whenever a pixel point intersecting with the structural element is found (only one position in the structural element and the image to be processed are foreground points), the point of the image to be processed relative to the original of the structural element is marked as foreground points. The set of all these marker points is the result of image expansion. The definition of image expansion is as formula (3).

If I is the source image to be processed and B is the selected structural element, then the mathematical definition of the expansion of the target image I by the structural element B is as follows:

$$I \oplus B = \{x | x = i + b, i \in I, b \in B\} = \cup_{b \in B} I_b. \quad (3)$$

Among them, $i + B$ denotes that I is translated by vector b , x denotes the two-dimensional value after operation, and \oplus is the symbol of expansion operation.

The meaning of morphological dilation is that as long as the intersection point with the structural element is not empty in the image, the pixels corresponding to the original position of the structural element in the image to be processed are marked. The set of all the symbolic conditions is the result of the dilation operation.

Corrosion and expansion operations have different effects. Expansion can connect two separate regions and make two isolated “islands” connected. Corrosion can eliminate the pore in the image, make the original isolated “island” disappear in the image, and play the role of filtering noise.

2.3. Feature Point Detection. In feature point detection [6], what kind of feature is used to describe objects is an important part of feature matching. Whether the feature selection is appropriate or not is the decisive factor affecting the success or failure of the next matching work. In this paper, we use feature points, but there is no definite concept at present. Generally speaking, we represent points with feature properties in images, such as extreme points, points with zero second derivative, intersection points of lines, and lines in images. The feature points represent the important local feature information in the image, which effectively reduces the image information and plays a great role in the analysis and understanding of the image. Using feature

points for image matching can enhance the validity and reliability of image matching.

2.3.1. Harris Feature Point Detection. The core idea of the Harris feature point detection algorithm is to use small windows to judge the gray level change by moving on the image. If the gray level changes obviously in the process of moving, there will be feature points in the window.

If the gray level does not change or does not change in one direction, there is no feature point in the window. By constructing a mathematical model, the problem can be expressed as follows:

$$E(u, v) = \sum_{x,y} w(x, y) [I(x+u, y+v) - I(x, y)]^2. \quad (4)$$

In formula (4), (u, v) represents window translation transformation, $E(u, v)$ is gray level change, $w(x, y)$ is window function, and $I(x, y)$ is image gray level. The window function is Gauss function as follows:

$$w(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \quad (5)$$

This function can be understood as calculating the weight of gray level in the window, which changes with the direction from the center point to the edge smaller and smaller, so as to eliminate the influence of noise on it.

The Taylor expansion for formula (10) can be expressed as follows:

$$E(u, v) = \sum_{x,y} w(x, y) [I_x u + I_y v + o(u^2, v^2)]^2. \quad (6)$$

M is the partial derivative sparse matrix and it is expressed as

$$E(u, v) \cong [u, v] M \begin{bmatrix} u \\ v \end{bmatrix}. \quad (7)$$

In this formula, I_x is the difference in direction x , I_y is the difference in direction y , and M is the partial derivative sparse matrix and is expressed as follows:

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}. \quad (8)$$

For Harris feature point monitoring, the size of two eigenvectors of M matrix can be used. The two eigenvalues represent two directions of the fastest and slowest change, respectively. Both of them are large and can be judged as characteristics. One is the edge, and the other is the edge. Both of them are small and flat areas without change. So, we can express the problem as follows:

$$R = \det M - k(trM)^2, \quad (9)$$

where $\det M$ is the determinant of a matrix M which is the product of two eigenvalues, trM is the trace of a matrix M which is the sum of two eigenvalues, and k is a constant. From the analysis, it can be seen that the judgment of Harris

feature points is related to the value R . When the value R is larger and positive, it can be judged as the feature point. When the value R is larger and negative, it can be judged as the edge, and when the value R is smaller, it can be judged as the flat area without change. In practical applications, the horizontal and vertical difference operators are used to evaluate the matrix M of the pixels, and then the Gauss filter is used to determine the feature points directly by R . The algorithm is simple, efficient, insensitive to illumination changes, and has rotation invariance. However, it does not have scale invariance; when the threshold is too large, the sensitivity of focus detection is not enough and the number of feature points is too small, and false feature points are easy to appear when the threshold is too large.

2.3.2. ORB Feature Point Detection. ORB feature point detection [7, 8] is a SIFT algorithm for extracting and describing image features. FAST (Features from Accelerated Segment Test) is used to extract feature points. In the algorithm based on image feature detection and matching, through the GLAMPoints paper, we can learn the exact matching points greedily learned. And through the feature detection and description of image matching, we can learn from manual design to deep learning. It is best to talk about key point detectors and descriptors based on random samples through RSKDD-Net. The core idea of the FAST algorithm is as follows: for a pixel, a circular region is constructed with the center of the pixel, and the gray value of the pixel in the center of the circle is compared with that of all the pixels in the circle. When the pixel value of enough points is larger or smaller than that of the center of a circle, the pixel is considered to be a feature point. It can be expressed as follows:

$$D = \sum_{\text{circle}(p)} |I(x) - I(p)| > \varepsilon. \quad (10)$$

In formula (10), $I(x)$ is the gray value of any point in the circle, $I(p)$ is the gray value of the central pixel, and ε is the set thresholds. If the D value is greater than the set threshold, the corner can be judged.

In view of the fact that FAST features do not satisfy scale changes, the ORB algorithm establishes scale pyramids, similar to the SIFT algorithm in building scale pyramids. For each layer of image, FAST feature points are extracted. Finally, the extracted features are regarded as a set of features extracted from all layers, so as to meet the scale invariance. Aiming at the problem that FAST feature points have no direction, the ORB algorithm gives the gray centroid position in the neighborhood where R is the center of feature points and the direction of gray centroid is the direction of feature points, and then the formula for calculating centroid C is as follows:

$$m_{pq} = \sum_{x,y \in r} x^p y^q I(x, y), \quad (11)$$

$$C = \begin{pmatrix} m_{10} & m_{00} \\ m_{01} & m_{00} \end{pmatrix}.$$

In formula (11), $I(x, y)$ is the gray value of the neighboring pixel, and the direction of the vector \overline{OC} of the feature point O and the center of mass C is as follows:

$$\theta = \arctan \frac{m_{01}}{m_{10}}. \quad (12)$$

Since FAST feature is only an algorithm for feature detection and does not involve the formation of feature descriptors, the ORB algorithm uses Rbrief algorithm for feature description. The Rbrief algorithm is that the descriptor generated by the Brief algorithm [9] adds rotation angle information. The Brief algorithm takes n pairs of pixels randomly around the position of the extracted feature points to form an image block and then compares the two pixels in the image block, the result of which is expressed by 0 or 1. A series of binary digit strings with length n are generated, and the generated binary digit strings are the Brief descriptors of the feature points. In order to enhance the descriptor's robustness to noise and illumination, the image block is smoothed by Gauss filtering. The gray level of $5 * 5$ neighborhood of a point in its $31 * 31$ neighborhood is compared with that of point pair. The comparison criterion of the ORB algorithm after Gaussian smoothing of image blocks can be expressed as follows:

$$\tau(p; x, y) := \begin{cases} 1, & p(x) < p(y), \\ 0, & p(x) \geq p(y). \end{cases} \quad (13)$$

In formula (13), the average gray levels of the two pairs of pixels X and Y in the image block p are, respectively, $p(x)$ and $p(y)$. The generated binary digit string can be represented as follows:

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; x, y). \quad (14)$$

In formula (14), n can generally be taken as 128, 256, and 512. It can be seen that the Brief descriptor does not have rotation invariance. In order to describe the direction of the generated feature points using the descriptor, ORB improves the Brief descriptor. Let the original Brief algorithm represent a $2 * n$ matrix by selecting n pairs of points around the feature points:

$$S = \begin{pmatrix} x_1, \dots, x_n \\ y_1, \dots, y_n \end{pmatrix}. \quad (15)$$

According to the angle θ between FAST feature points and centroid, the rotation matrix R_θ is constructed, and then the corresponding direction of rotation of the point set matrix can be expressed as follows:

$$S_\theta = R_\theta S. \quad (16)$$

At this time, the original Brief descriptor can be expressed as follows:

$$g_n(p, \theta) := f_n(p) | (x_i, y_i) \in S_\theta. \quad (17)$$

In order to ensure the separability of feature point descriptors, the ORB algorithm improves the original Brief algorithm by using statistical principles, namely, Rbrief

algorithm. Each point is arranged in columns according to the binary digits taken above, and the matrix Q is generated. The average value of Q of each column is calculated. The first column in the matrix R is rearranged according to the distance from the draw value to 0.5. The correlation between Q of all the columns and all the columns in the matrix R is calculated. If the result is less than the set threshold, put the column in the matrix R , until 256 columns in R stop. The feature points extracted by the ORB algorithm and the descriptors generated by the ORB algorithm are fast and have good rotation invariance. However, when adapting to scale change, the effect is general. Because its descriptor is binary digit string, the matching process is not stable enough, which will cause some difficulties in matching.

2.3.3. SIFT Feature Point Detection. The SIFT algorithm [10] is a computer vision feature extraction algorithm which satisfies our needs very well. The feature extracted by this algorithm is a scale rotation invariant, so it has excellent robustness and is convenient for feature matching. The main idea of SIFT is to collect images in continuous scale, find extremum points in continuous scale space, remove unstable extremum points, and extract and acquire local features of rotation and scale invariance around stable extremum points. Finally, 128-dimensional descriptors are generated.

SIFT features have the following advantages: strong robustness and good adaptability to geometric deformation, image noise, and brightness change. An image can generate a large number of SIFT feature points, which are rich in data. The local invariant features corresponding to the two images have good repeatability. Firstly, we consider the scale invariance of SIFT features. In order to adapt to scale transformation, feature points need to be detected in all image scales. Therefore, it is necessary to establish the scale space, and the Gaussian kernel function [11] is the only smoothing function in the scale space, so the Gaussian kernel function can be used to build the scale space. The scale transformation of an image can be expressed as follows:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (18)$$

where $I(x, y)$ is expressed as an image, (x, y, σ) is a Gaussian kernel function, and $L(x, y, \sigma)$ is an image in different scales, and it is a scale space. The Gauss kernel function can be expressed as follows:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-[(x^2+y^2)/2\sigma^2]}, \quad \sigma \geq 0. \quad (19)$$

In order to improve the reliability and stability of feature points, the preliminary identified feature points are screened. The screening steps are divided into two steps. The first step is to remove the low contrast points, i.e., some noise-sensitive points, and Taylor expansion is carried out for equation (21):

$$D(x) = D + \frac{\partial D}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x. \quad (20)$$

Seeking extreme points,

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x}. \quad (21)$$

In formula (21), \hat{x} is the position of the extreme point and x is the offset of the extreme point. $D(\hat{x})$ can be used as a basis for judgment. Generally, feature points of $D(\hat{x})$ less than 0.3 are removed. The second step is to remove the edge response. The points at the edge are very sensitive to noise, and the Gauss difference function has a large principal curvature for a peak across the edge, so the SIFT algorithm excludes the edge response by calculating the principal curvature. The principal curvature can be calculated by a 2×2 Hessian matrix H as follows:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}. \quad (22)$$

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2},$$

$$\theta(x, y) = \alpha \tan 2 \left\{ \frac{[L(x, y+1) - L(x, y-1)]}{[L(x+1, y) - L(x-1, y)]} \right\}. \quad (24)$$

In formula (24), L is expressed as the scale of each point. The gradient value of each point represents the main direction of the neighborhood gradient and takes the main direction as the direction of the feature point. After determining the main direction, the position, scale, and direction of each feature point have been reflected. At this point, a descriptor is generated for each feature point. To solve the gradient value on the image smoothed by the corresponding Gauss feature points, each feature point can generate a 128-dimensional descriptor, that is, a 128-dimensional feature vector. In order to remove the influence of illumination changes, the feature vector can be normalized and SIFT features can be extracted.

2.4. Cluster Analysis Method. Cluster analysis is a method of quantitative classification with mathematical tools. Cluster analysis algorithm [12] is a typical unsupervised learning algorithm and one of the important algorithms of data mining. It is mainly used to study classification problems; that is, similar data are automatically divided into the same category. When using the clustering algorithm, we usually analyze the similarity relationship between data to group data and divide data into different classes. The greater the similarity between similar data, the smaller the similarity between different classes of data; that is, the greater the difference, the better the classification effect.

2.4.1. Clustering Algorithm Based on Cumulative Frame Difference Intersection. The basic idea of the clustering algorithm based on the intersection of the cumulative frame difference is based on the calculation of the cumulative frame difference and cross and cluster the two cumulative

Since the principal curvature is proportional to the eigenvalue of matrix H , if the two eigenvalues of matrix H are, respectively, α or β , then there are

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} < \frac{(r+1)^2}{r}. \quad (23)$$

In formula (23), $\text{Tr}(H)$ is the trace of matrix H , $\text{Det}(H)$ is the determinant of matrix H , and r is the constant which is generally taken as 10 to determine whether the principal curvature is in a certain range.

After the stable feature points are selected, the appropriate descriptor is generated for the feature. In order to make the descriptor rotate invariant, the gradient modulus and direction can be expressed as follows:

frame differences so that the changing area can accurately converge to the foreground edge, and then the area binarizes the mask to obtain a differential image frame that is to ensure the real-time performance, and greatly improve the segmentation effect. It is also suitable for sports video sequences with fast-moving objects. The main steps of the algorithm are as follows:

- (1) After the median filtering, the image processing sequence difference between adjacent frames and frame interval difference is as follows:

If the current image frame n can be expressed as $f(k)$, the frame image k_{+1} and k_{+2} can be expressed as $f(k+1)$ and $f(k+2)$, and the interframe difference of the adjacent difference categories can be expressed as follows:

$$\begin{aligned} &|f_{k+1}(x, y) - f_k(x, y)|, \\ &|f_{r+1}(x, y) - f_r(x, y)|. \end{aligned} \quad (25)$$

The cumulative results of the n frames are as follows:

$$D_1 = \sum_{r=1}^k |f_{r+1}(x, y) - f_r(x, y)| = E \cup N_1,$$

$$D_2 = \sum_{k=1}^k |f_{k+2}(x, y) - f_k(x, y)| = F \cup N_2, \quad (26)$$

where E and F are the change areas in the cumulative results and N_1 and N_2 are noise.

- (2) By intersecting the two cumulative frame differences, the pixels belonging to the change area in the two cumulative results can be effectively concentrated



FIGURE 1: Four kinematic video frames.



FIGURE 2: Grayscale of video frame image.

around the foreground contour, thereby obtaining an ideal moving foreground contour.

$$D_1 \cap D_2 = (A \cup N_1) \cap (B \cup N_2) = EF \cup (EN_2 \cup FN_1 \cup N_1N_2). \quad (27)$$

Since EN_2 and FN_1 are clustering of moving regions and noises, formula (5) can be rewritten as follows:

$$D_1 \cap D_2 = AB + N. \quad (28)$$

(3) For further clustering method to remove background pixels, clustering step procedure is described as follows:

- Step 1. Randomly determine two points and set $5 * 5$ and cycle number 1P as the centers.
- Step 2. For each pixel in a rectangular window,

$$D_1 = \frac{1}{N_K^2} \sum_{N_1 \cap N_2 \cup EF \cap EN_1} \sqrt{(x_i - x) - (y_i - y)}. \quad (29)$$

Step 3. Compare the distance D_1 with the threshold D_j . If $D_i \leq D_1$, binning of pixels moved to the edge of the moving foreground object. Otherwise, the pixel will be merged into the background noise edge pixel.

Step 4. Move the rectangular window from the horizontal and vertical directions, increase the number of pixels and follow Step 2 until all the binning classes they belong to, that is, the number of cycles reaches the specified number, and the cluster ends.

3. Experiments

In this document, the experimental platform configuration is based on a 64 bit flagship version of the Win7 operating system, 8 G physical memory, a 2.2 GHz quad-core Intel Core I5-5200U CPU, and MATLAB 2014b-based simulation software.

In order to reflect the universality of the experiment, the material used in this document comes from the network and not from a dedicated video library. The sports video used in the simulation experiment comes from sports gates of large gates. The video frame rate is 5 frames per second, and the output image resolution is 480×360 . Tennis, volleyball, water polo, and baseball were selected as the four sports videos. In this article, the SIFT algorithm is used to derive the characteristics of the sport, and then the cumulative cross-sectional algorithm is used to identify sports, as shown in Figure 1.

4. Discussion

Before the simulation experiment, it is necessary to preprocess the actual video object to improve the video characteristics. Firstly, all four types of motion video frames are gray, as shown in Figure 2.

In the image processing process, the presence of noise is inevitable. In this article, denaturing is also required in the preprocessing process. This document enhances traditional filter average processing, saves processing time efficiently, and improves processing results, which are useful for detecting and tracking images in the later stages of football. In this paper, we simulate and analyze the noise added to the gray image and use the traditional median filter and the

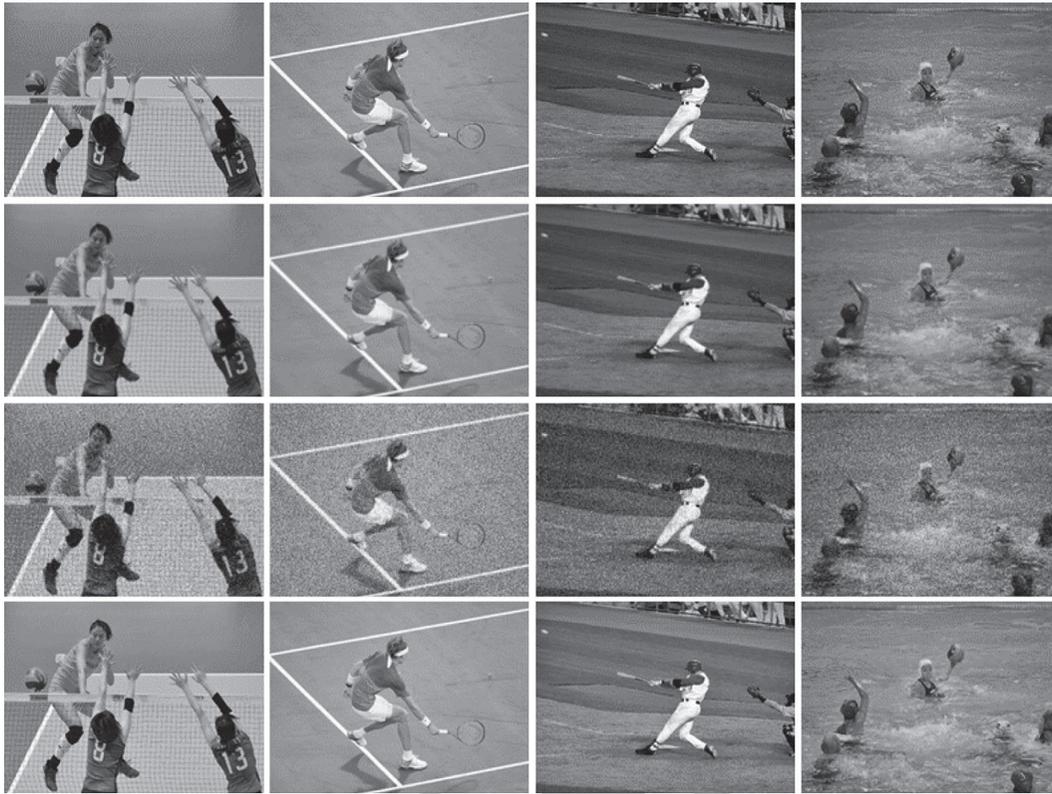


FIGURE 3: Video frame image noise and filtering.

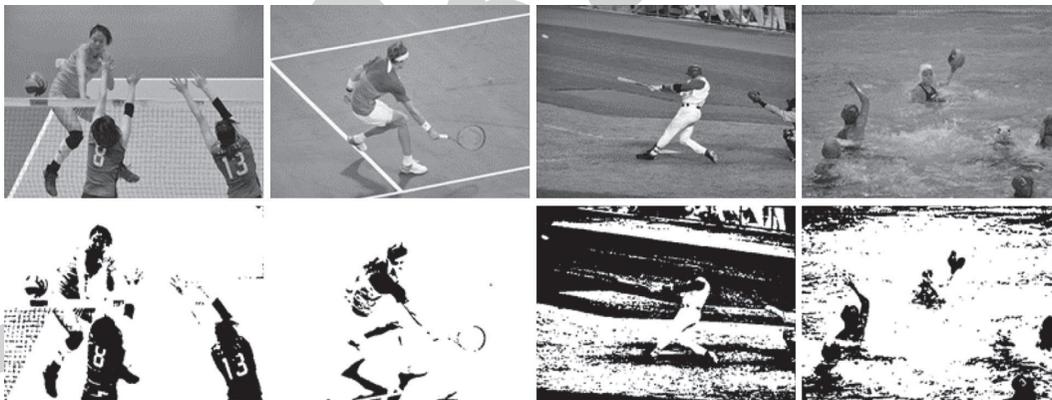


FIGURE 4: Two values.

TABLE 1: Motion recognition rates.

	Volleyball (%)	Tennis (%)	Baseball (%)	Water polo (%)
Accuracy	86.5	91.2	87.4	83.2
Recognition rate	82.3	93.4	88.4	84.3

improved median filter, respectively, to avoid the noise influence in the process of processing, as shown in Figure 3.

After filtering, we edge the image with the result of filtering. After gray level processing, there are 256 gray levels. By choosing appropriate threshold, the gray level of the gray image can be divided into two parts, and then the binarization of the image can be obtained. It keeps the region of

interest in the image to the greatest extent and shields all the irrelevant information as shown in Figure 4.

After the preprocessing results, the first step is to carry out simulation experiments on sports detection. The processed video frame images are recognized, and the accuracy and recognition rate of the four kinds of motion are compared, respectively, as shown in Table 1.

From this table, we can see that tennis has the highest recognition rate, while the rest of the sports are relatively low, which may be related to the video background.

5. Conclusions

In recent years, with the improvement of people's living standards, more and more attention has been paid to sports. The research of video data processing has high theoretical significance and commercial value. In this paper, four kinds of sports videos are analyzed, and four kinds of sports images are extracted. After image graying, image denoising, and image binarization, a detection method based on SIFT feature points is designed. In the simulation experiment, image preprocessing, sports item detection, and sports item recognition are analyzed, respectively. By comparing the accuracy and recognition rate of these four sports, we can conclude that the recognition rate is more than 80%. It can be seen that the recognition rate is still very high, and the tennis recognition rate is the highest. This shows that the SIFT algorithm and cumulative cross-section grouping algorithm proposed in this paper are suitable for sports video recognition.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] J. Lu, G. Zhou, C. Yang et al., "Steganalysis of content-adaptive steganography based on massive datasets pre-classification and feature selection," *IEEE Access*, vol. 7, pp. 21702–21711, 2019.
- [2] O. Choi, M. Kim, A. E. Cho et al., "Fates of water and salts in non-aqueous solvents for directional solvent extraction desalination: effects of chemical structures of the solvents," *Membrane Water Treatment*, vol. 10, no. 3, pp. 207–212, 2019.
- [3] W.-Y. Chang, "Research on sports video image based on fuzzy algorithms," *Journal of Visual Communication and Image Representation*, vol. 61, pp. 105–111, 2019.
- [4] Hume, Patria, Knudson et al., "Isbs 2018 Auckland Conference Sprinz sport and exercise biomechanics teaching archive programme," *ISBS Proceedings Archive*, vol. 36, no. 1, p. 278, 2018.
- [5] Y. Li, "Research on sports video image analysis based on the fuzzy clustering algorithm," *Wireless Communications and Mobile Computing*, vol. 2021, no. 3, pp. 1–8, 2021.
- [6] L. Zhang, "Evaluation and simulation of sports balance training and testing equipment based on medical video image analysis," *IEEE Sensors Journal*, vol. 20, no. 20, pp. 12005–12012.
- [7] M. Takahashi, "Video analysis technologies in sports broadcast programs," *Journal of the Japan Society for Precision Engineering*, vol. 84, no. 12, pp. 949–952, 2018.
- [8] Q. Fu, L. Ma, C. Li, Z. Li, Z. Zhu, and Z. Lin, "Detection method of sports scene conversion for MPEG compressed video based on fuzzy logic," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 3107–3115, 2021.
- [9] N. Sasikala, P. V. V. Kishore, D. A. Kumar, and C. R. Prasad, "Localized region based active contours with a weakly supervised shape image for inhomogeneous video segmentation of train bogie parts in building an automated train rolling examination," *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 14917–14946, 2019.
- [10] B. Joanna and K. Zuzanna, "The impact of social media on managing the image of the Polish national football team," *Physical Culture and Sport, Studies and Research*, vol. 87, no. 1, pp. 46–55, 2020.
- [11] M. D. Basar, A. D. Duru, and A. Akan, "Emotional state detection based on common spatial patterns of EEG," *Signal, Image and Video Processing*, vol. 14, no. 3, pp. 473–481, 2020.
- [12] K. I. Sainan, M. Mohamad, Z. Mohamed et al., "Athletes tracking using homography method: a preliminary study," *International Journal of Engineering & Technology*, vol. 7, no. 4, pp. 6–10, 2018.