

Retraction

Retracted: Semantic Analysis of Multimodal Sports Video Based on the Support Vector Machine and Mobile Edge Computing

Wireless Communications and Mobile Computing

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

Copyright © 2023 Wireless Communications and Mobile Computing. 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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] X. Chen, "Semantic Analysis of Multimodal Sports Video Based on the Support Vector Machine and Mobile Edge Computing," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 3511535, 9 pages, 2022.

Research Article

Semantic Analysis of Multimodal Sports Video Based on the Support Vector Machine and Mobile Edge Computing

Xu Chen 

Wuhan Technical College of Communications, Wuhan, Hubei 430065, China

Correspondence should be addressed to Xu Chen; chenxu@whtcc.edu.cn

Received 8 January 2022; Revised 27 January 2022; Accepted 29 January 2022; Published 22 February 2022

Academic Editor: Xin Ning

Copyright © 2022 Xu Chen. 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.

Even though sports video analysis and research have yielded some results, the competition video itself contains a large number of images, video, audio, text, and other elements. Video semantic analysis has always been a hot and difficult topic in video studies. Multimodal sports video data contain a wealth of data. Second, it has a complex structure with an ambiguous relationship between each video unit. A typical algorithm based on kernel function learning is a support vector machine. It has shown promising results in the prediction of landslide displacement time series. The semantic analysis of multimodal sports video using support vector machines and mobile edge computing is the subject of this study. The theoretical prediction results based on the monitoring data of multimodal sports video semantic analysis show that they are comparable to neural network prediction results using genetic algorithm. The semantic analysis of multimodal sports video based on support vector machine method proposed in this study has a better prediction ability, and its theoretical prediction results are close to the actual monitoring values.

1. Introduction

Video semantic analysis has always been a contentious and difficult topic in video studies. Data from multimodal sports video contain a wealth of information. To begin with, the content is diverse, ranging from abstract semantic descriptions of high-level events to low-level audio-visual perception content. Second, it is structurally complex, with an uncertain relationship between each video unit. Despite the fact that sports video analysis and research have yielded some results, the game video itself is a fusion of a large number of images, videos, audio, texts, and other features [1]. Because most of the video analysis methods are based on video's underlying features, semantic analysis and research of high-level video are still a lengthy process [2, 3]. Multimodal sports video has a huge market potential and a large audience. In addition to traditional TV users, sports programs are popular with new media users such as mobile and Internet. Sports videos are not only concerned by ordinary audiences but also favored by sports professionals. They are used in tactical and strategic analysis [4]. Multimodal semantic analysis of sports video is a major difficulty in this

study. Most of the current research methods only rely on the features of the video at the bottom level to assist in the analysis of exciting events. There is a lack of a model that integrates the features of video, audio, and text to study exciting events [5, 6].

Relatively speaking, the support vector machine method developed in recent years has more prominent advantages. It embeds or maps the data in the input space into a high-dimensional feature space where linear relations can be found, thus solving many problems that are difficult to be solved by linear methods in the original sample space [7]. Sports video classification results are also linked to classifier design, in addition to feature extraction. Support vector machine is used to create sports video classifiers at the moment. The choice of kernel function, on the other hand, determines the algorithm's learning performance. To determine whether a single landslide time series is representative, some studies have looked at the impact of kernel function selection on prediction results. The training speed of a support vector machine is slow when dealing with large amounts of data, whereas a least-square support vector machine is an improved support vector machine that

effectively overcomes the flaws of traditional machine learning algorithms such as overfitting and slow training speed. Support vector machine is a typical algorithm based on kernel function learning. It has achieved good results in the application of landslide displacement time series prediction [8–19].

Support vector machine is used to predict, then the calculated results of slope displacement in different empirical modes are accumulated and synthesized, and the predicted value of multimodal sports video semantic analysis is obtained [20]. The support vector machine method proposed in this study for multimodal sports video semantic analysis has stronger prediction ability, and its theoretical prediction results are in good agreement with the actual monitoring values. This theoretical method may contain some connection between observation results and deformation evolution mechanism, which provides a model worthy of reference for establishing a more practical and effective theoretical prediction model [21, 22]. Taking advantage of the incomplete and unclear information of D-S evidence theory and the good classification and generalization ability of support vector machine in small samples, the semantic analysis of sports video of support vector machine with a single feature is fused, the final recognition result is obtained according to the decision rules, and the semantic analysis effect of sports video of support vector machine is tested by simulation experiments [23]. The foundation of multimodal sports video semantic analysis must be based on the first principle of natural law. The prediction method with pattern recognition as the main means needs to study the underlying physical basis. Otherwise, there will be many deviations or even errors; that is, data analysis cannot be separated from the grasp and understanding of the physical basis. Even if this is just a statement from one family, this view at least objectively emphasizes the importance and necessity of multimodal sports video semantic analysis [24].

2. Related Work

According to reference [25], soccer video shots were classified based on object size and shot motion. Ma determines the video's motion mode and categorizes it into three categories: object motion, camera motion, and no motion. Using underlying features such as main color and golden section method, Ahmet Ekin and others divided football video shots into three categories. Other video features, such as audio and text features, have not been used in the literature [26] that is based on visual features using big data analysis methods. Because text and video are created simultaneously and independently of one another, multimodal sports video semantic analysis is currently primarily used for text event detection to aid video analysis. For video semantic shot annotation, multimodal sports video semantic analysis has not been used. It is finally classified into player lens, off-site audience lens, and competition field lens, according to literature research [27]. The features of the

court line and field color are extracted in this process, laying a solid foundation for the classification of the entire lens. Pitch ratio, texture, and object size are used to divide football video shots into four categories, according to reference [28]. Duan and his colleagues combined the underlying features of global motion, color, texture, shape, and lens length and divided sports video shots into eight semantic categories. Literature [29] uses a big data analysis method to analyze and detect stadium position changes, corner kick events, and shooting events using football video as the research object. Players, football field lines, football, and camera movement are among the underlying video data used. Video semantic shot tagging plays an important basic role in video semantic analysis, according to reference [30], and researchers both at home and abroad have conducted in-depth and extensive research. Semantic shot tagging can provide the content of a video's intermediate semantic description, which sits between low-level features and high-level semantics. It provides efficient semantic event detection, classification, and recognition services at a high level. The tagging results provide a solid foundation for semantic video data mining by providing a clear judgment basis for subsequent judgment of the cause and effect relationship between shot and some high-level events. Reference [31] proposed to detect the sports field by using the methods of edge detection and mathematical morphology, remove the sign line of the field by Hough transform, and finally segment the sports area by using the detected field information. Video frequently has two modes, namely, pure video stream and audio stream, which are the information sources for video semantic analysis when using the big data analysis approach [32]. We use these two modes for semantic analysis of sports video represented by football. Literature [33] shows that by taking football video as the research object, the progress of the game is analyzed and studied. The underlying video information mainly includes the color of the players' jerseys, the color of the football, and the color of the football field. According to these color features, we can track the football and players' movements. Literature [34] pointed out that video analysis based on multimodal information is an important research direction, but it is difficult to extract all kinds of information from the video. Video information is mainly about object detection, but in terms of audio information, it is difficult to extract audio information for auxiliary analysis because there is always noisy background sound in the video. The superimposed text of football game video can usually provide important information about the game, but it is difficult to extract text information from complex background.

This study presents the semantic analysis of multimodal sports video based on support vector machine and mobile edge computing. We must first analyze the video structure and then extract a specific video structure from unstructured video data in this process. Second, video features such as low-level visual features, auditory features, and text features are extracted. Finally, a manual interface is created to load video data and import it into the video database, making it easy for users to query and retrieve information based on

their specific needs. The content-based video semantic analysis method abstracts the underlying video information and then processes it using some specific methods.

3. Support Vector Machine and Mobile Edge Computing

Among all kinds of videos, sports videos are relatively structural. This is because the number of cameras around the stadium is limited and their positions are relatively fixed, and the broadcast director will organize the shots captured by different cameras in a specific way. The minimum unit of sports video is the frame in the video image. In order to effectively cover different types of sports videos, it is necessary to extract features that can describe the video types. The current features mainly include color, texture, edge, and so on. The semantics of sports video is expressed in the form of “event,” which is analyzed on the basis of grammatical lens and audio grammatical content. At present, we only analyze the key semantic events in football video, that is, goals and threatening shots, because this is the most needed shot for the audience and the one to be selected for the editing program of football game. This method not only has high solution accuracy but also can ensure that the found solution is the global optimal solution, so it has good generalization ability. Video often contains two modes, namely pure video stream and audio stream, which are the information sources of video semantic analysis. We use these two modes for semantic analysis of sports videos represented by football. According to the previous discussion, we propose the semantic analysis framework of sports video as shown in Figure 1.

First, a splitter separates the video stream from the audio stream. After that, the video and audio streams are examined separately. As mentioned in Part 2, the video stream is first divided into “physical shots” based on physical characteristics, and then “grammar shots” are determined. The low-level “physical content” is extracted first, and then the “grammatical content” is synthesized. Additional information is provided in the following section. Finally, by combining the grammatical shot and the grammatical content of audio, the specific semantic events are analyzed.

The basic principle of multimodal support vector machine model prediction method based on empirical mode decomposition (EMD) is shown in Figure 2. Given training samples $\{(x_1, y_1), (x_2, y_2), \dots, (x_j, y_j)\}$, EMD analysis is carried out on them, and several IMF components of the time series of displacement change are obtained. Then, the support vector machine regression method is used to predict each IMF obtained by decomposition, and the predicted values obtained by each IMF are accumulated, summed, synthesized, and reconstructed to obtain the predicted values of multimodal sports video semantic analysis.

Because sports video contains numerous interruptions, such as close-ups of off-site audiences and advertising, structured video segmentation is used for classification. Because the most effective feature shots for classification are long shots, the threshold method is used to eliminate nonfeature shots and keeps only the feature

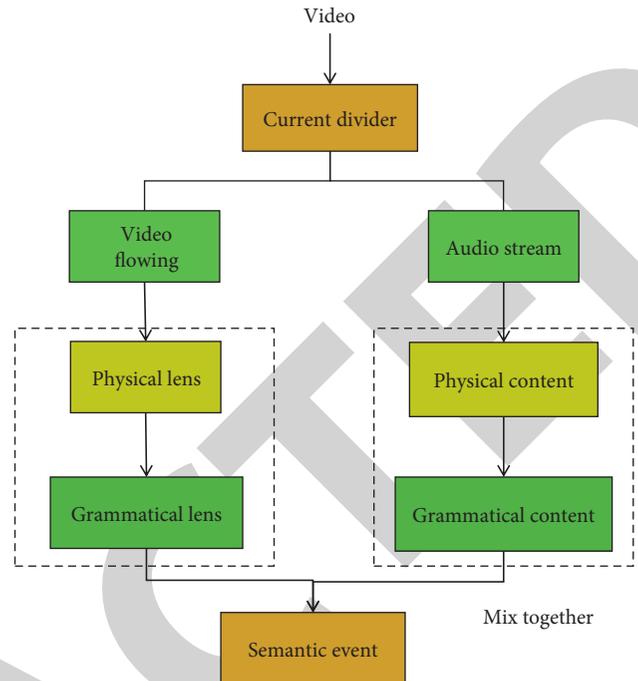


FIGURE 1: Multimodal semantic analysis framework.

shots. In this study, the double-comparison method combining edge segmentation and motion segmentation is used for athlete information segmentation and extraction. The object is relatively independent, and the contour is clear in football, table tennis, tennis, and other sports. Edge segmentation and motion segmentation methods can be used to obtain the results. After extracting the venue of a sports video using edge information from the video image and an attention model algorithm, the object features in the venue are extracted, and a classification system composed of support vector machine meta classifier is used to classify the sports video. The system structure is shown in Figure 3.

Specifically, for an original signal $x(t)$, all the extreme points are found, all the extreme points are interpolated by a cubic spline function curve, and the upper envelope $x_{\max}(t)$ of the original signal is fitted. Similarly, the lower envelope $X_{\min}(t)$ can be obtained. The upper and lower envelopes contain all the signal data. A mean value line $m_1(t)$ can be obtained by connecting the mean values of the upper envelope and the lower envelope in sequence, and its equation can be expressed as follows:

$$m_1(t) = \frac{[x_{\max}(t) + x_{\min}(t)]}{2}. \quad (1)$$

Then, subtract $m_1(t)$ from $x(t)$ to get $h_1(t)$, namely,

$$h_1(t) = x(t) - m_1(t). \quad (2)$$

For different signals, $h_1(t)$ may or may not be a natural modal function component. Generally speaking, it does not meet the conditions required by the inherent modal function. At this time, $h_1(t)$ is taken as the original signal, and the above steps are repeated K times until certain conditions are met.

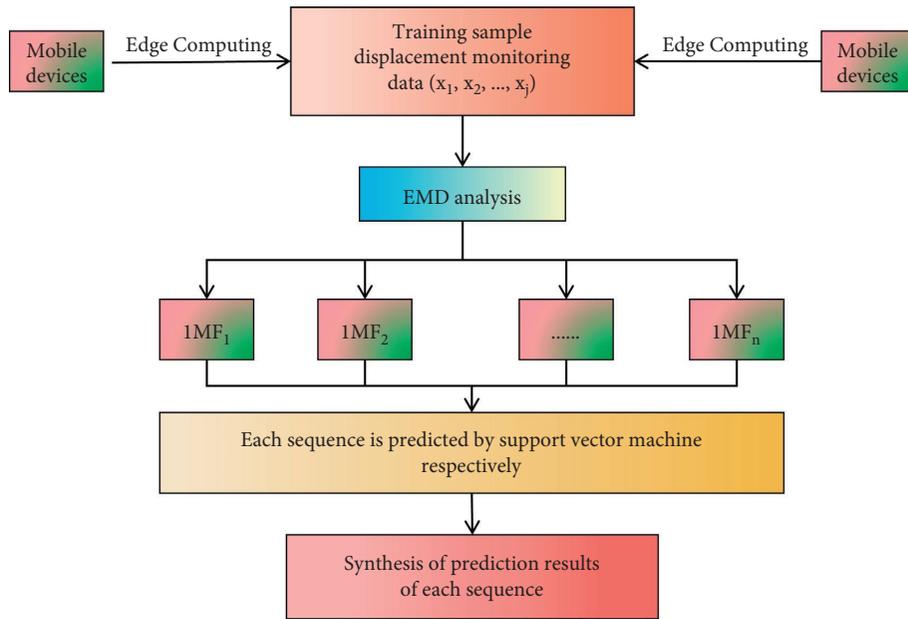


FIGURE 2: Basic principle of the multimodal support vector machine and MEC.

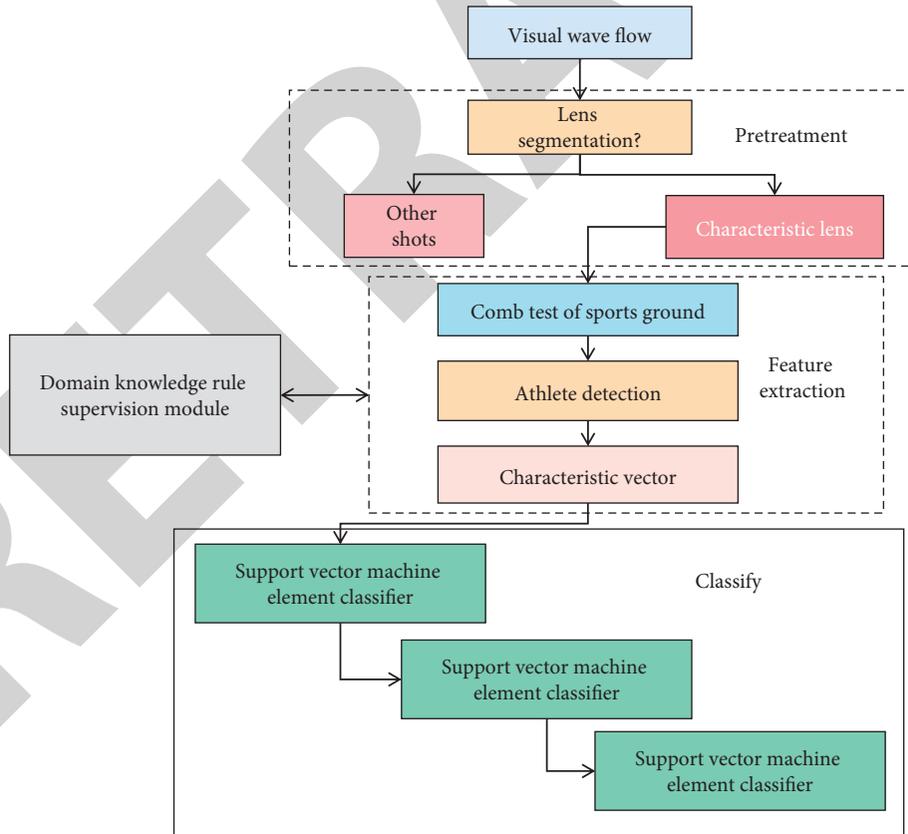


FIGURE 3: System structure diagram of the support vector machine.

$$c_1(t) = h_{1k}(t). \quad (3)$$

If $c_1(t)$ satisfies the judgment condition, that is

$$SD = \sum_{i=0}^T \left| \frac{[h_{1(k-1)}(t) - h_{1k}(t)]^2}{h_{1(k-1)}^2(t)} \right|. \quad (4)$$

$C_1(t)$ is regarded as a 1 MF; otherwise, the iterative calculation is continued. Experience shows that SD value is generally 0.2–0.3.

Subtract the residual signal of $c_1(t)$ from $x(t)$, that is, the residual error is as follows:

$$r_1(t) = x(t) - c_1(t). \quad (5)$$

Taking $r_1(t)$ as a group of new signals and repeating the above modal decomposition process, all the residuals $r_i(t)$ can be obtained after multiple operations:

$$r_{i-1}(t) - c_i(t) = r_i(t), \quad (i = 2, 3, \dots, n). \quad (6)$$

When $r_1(t)$ satisfies the condition that $Cn(t)$ or $Rn(t)$ is less than the predetermined error, or the residual $Rn(t)$ becomes a monotonic function, that is, it is impossible to extract the IMF component from it, the modal decomposition process is terminated. So far, the original signal $x(t)$ can be composed of n -order IMF component and residual $Rn(t)$, that is

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t) = \sum_{i=1}^n IMF_i + r_n(t), \quad (7)$$

where $rn(t)$ is called residual function.

The prediction accuracy of the model is evaluated by the average relative error MRE, and its specific expression is as follows:

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{(y_i - \hat{y}_i)}{y_i} \right| \times 100\%, \quad (8)$$

where y_i is the real value, and \hat{y}_i is the predicted value.

4. Research on Semantic Analysis of Multimodal Sports Video

4.1. Semantic Analysis of Multimodal Sports Video Based on the Support Vector Machine. Multimodal sports video retrieval technology based on support vector machine has gotten a lot of attention lately, thanking the release of large-scale semantic concept sets. Many researchers from all over the world have spent a lot of time and effort attempting to solve the problem of sports video type classification. The manual and automatic stages of sports video type classification research can be divided into two categories. Because the labor phase takes a long time, classifying a large number of sports videos is impossible, and the workload of sports video classification is relatively high. To some extent, the semantic concept can reflect the visual semantic information in news video. Different modal information reflects the semantic content in different aspects of video, for example, image contains the visual information of video, voice

contains the background music of the video or the voice characteristics of the speaker, and subtitles contain the description of the video content, because sports video is a collection of continuous frames and contains a variety of modal information such as image, voice, and text. The information sources for video semantic analysis are usually two modes: pure video stream and audio stream. For semantic analysis of sports video represented by football, we use these two modes.

In video, a shot is the basic unit to process the video stream of support vector machine. The shot is defined as an uninterrupted sequence of frames taken by a camera, which is the basic structure layer of further structural processing of video. In video processing, the first step is to find the cut point of the shot. Firstly, some symbols and terms of news video retrieval are introduced, and the news video retrieval problem is formally described to further standardize the subsequent discussion. Among all kinds of videos, sports videos are relatively structural. The minimum unit of sports video is the frame in the video image. It is necessary to extract features that can describe different types of sports videos in order to effectively cover them. Color, texture, edge, and other features are currently available. Color is one of the most widely used sports video features, and it is the one that is most likely to draw people's attention. Color space is the foundation for extracting color features. The broadcast director will organize the shots captured by different cameras in a specific way because the number of cameras around the stadium is limited and their positions are relatively fixed, and the broadcast director will organize the shots captured by different cameras in a specific way because the number of cameras around the stadium is limited and their positions are relatively fixed. For example, in the football video, the director will use different shots at the appropriate times to help the audience understand and appreciate the game. When attacking, for example, the director will switch to a midrange lens to show the game's key situations, then close-ups of key players, and often scenes of celebration such as spectators, before the game is replayed in slow motion and then returned to the live broadcast of the game.

The sound information in multimodal sports video based on support vector machine is mainly divided into voice, audience cheers, hitting sound, and so on. These sounds represent different meanings. The first problem to be solved in the processing of auditory information in sports video is to distinguish different kinds of sounds and then infer the content expressed in video according to the occurrence time of different kinds of sounds. The extraction of acoustic features is not only a process of information compression but also a process of signal deconvolution, in order to make the mode divider divide better. The multimodality of multimodal sports video of support vector machine integrates the multimodal information such as vision, audio, and text. It is required that the multimodal sports video retrieval of support vector machine should obtain meaningful content description and analysis from each mode as much as possible. In the first mock exam, the demand for video retrieval is increasing. On the other hand, the query topic

contains more and more high-level semantic information. The single-mode information or feature video query method is difficult to meet the user's query requirements. This is because the number of cameras around the stadium is limited, their positions are relatively fixed, and the broadcast director will organize the shots captured by different cameras in a specific way. In video processing, the first step is to find the cut point of the shot. Firstly, some symbols and terms of news video retrieval are introduced, and the news video retrieval problem is formally described to further standardize the subsequent discussion. The basic semantic expression unit of sports video is the lens. A video is composed of various "semantic events," and each "semantic event" is composed of several specific shot sequences. Compared with the "physical shots" directly distinguished according to the low-level features, these shots can be called "grammatical shots." We have determined the following lens types: long-range lens, medium-range lens, close-up lens, slow lens, on-site lens, and off-site lens. A lens can be given two grammatical meanings. For example, a lens can be given both "gas perspective lens" and "field lens." The sound information in multimodal sports video based on support vector machine is mainly divided into voice, audience cheers, hitting sound, and so on. In order to locate the action sequence from the video, two methods are adopted: the method based on audio recognition and the method based on player action trajectory recognition. In the football game, because the player's hitting power is large, and the hitting sound is clearly recognized in the video, the audio-based recognition method is adopted. For video information, multimodal information fusion can be defined as the ability to express predefined video semantic information by using a variety of information channels to represent specific content. The video mainly includes the following three kinds of modal information:

- (1) Visual modal information: all information that can be seen visually in the video, including both naturally formed and artificially generated visual information. Color, texture, edge, and other image features are the most commonly used visual features.
- (2) Audio modal information: all sound information that can be heard in the video, including voice, background music, and ambient sound. The analysis of sound information mainly includes speaker recognition, automatic speech recognition, audio classification, noise analysis, and so on.
- (3) Text modal information: including text resources describing video content, such as scripts and subtitle information, and text information obtained from video by optical character recognition, etc.

4.2. Experimental Results and Analysis. In the process of automatic classification of sports video types, it is necessary to extract the feature vectors of sports video types such as visual features, audio features, site area features, and sports features. Due to the complex changes of sports video, it is difficult to describe the type of sports video with a single

feature. Therefore, at present, we should mainly use a combination of multiple features to classify the type of sports video. The theory of multimodal sports video gives the actual displacement monitoring value of multimodal sports video and the theoretical prediction value obtained by genetic algorithm neural network method. Although the genetic algorithm neural network method has achieved a quite good prediction effect, it is not very ideal for the theoretical prediction before the occurrence of severe sliding. The prediction results of multimodal support vector machine show amazing consistency with the actual monitoring values. The multimodal support vector machine method can also predict the displacement in other periods. Three experiments were carried out as shown in Figures 4–6.

The experimental results show that, for the most part, the theoretical prediction results from the graph are consistent with the monitoring values, with some exceptions around the 16th, 30th, and 41st days. The deviation is not obvious from the picture due to the large scale of the selected drawing to show the entire picture, but the actual deviation still reaches the level of the learning sample's maximum residual error, and some points are slightly larger. Using a fusion method based on rules or specialized models, the extracted visual, aural, and stylistic elements of distinct movies are modeled, a mode of unified processing of different modal feature data is generated, and metadata that may properly describe the video content is created. The key frame in a video frame is one or more frames of video that can represent the characteristics of the video shot. The foundation and key steps of video structuring are extracting key frames and segmenting video shots, which have a direct impact on the efficiency of subsequent video detection. A video scene is made up of several video shots that are either semantically related or temporally adjacent.

Take out any shot logo shot from the logo shot set. If the shot appears an odd number of times, the shot is the logo shot before the start of the slow shot, and the next frame of the end frame of the shot is the start frame of the slow shot. In order to detect the effect of the slow shot detection algorithm implemented in this study, we collected four American tennis game videos and conducted four experiments, respectively. The format is avi, the resolution is 320×240 , and the frame rate is 20 frames per second. Recall rate and precision rate are used as evaluation performance indicators, and the test results are shown in Figures 7–10.

The experimental results show that the slow-motion detection algorithm based on the logo proposed in this study has high recall and precision, and this method can not only detect the slow motion made with pins but also the slow motion of videos played at normal speed by high-speed cameras. In this algorithm, only the lens frame images with video length between 10 and 20 frames are studied, which improves the detection speed and efficiency. With the increase in video resources, video classification has become the primary work in the field of video analysis, and video semantic shot labeling can obtain shots with specific semantics, establishing an effective link between video classification based on semantics and video classification based on underlying physical features, as well as providing

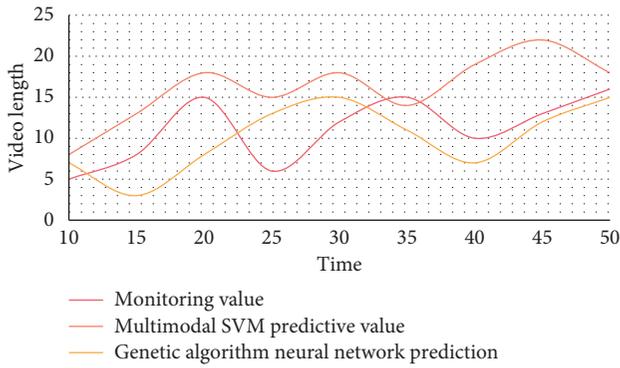


FIGURE 4: Comparison of multimodal sports videos and their theoretical prediction results.

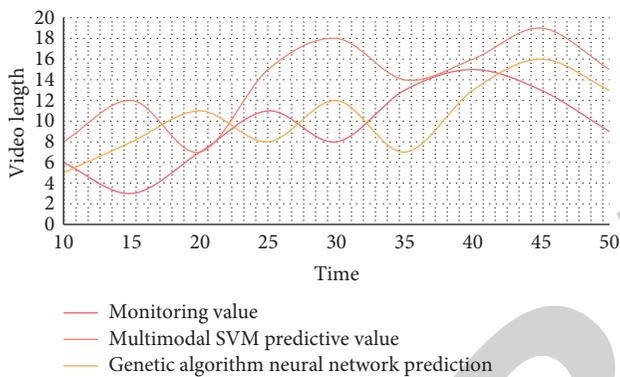


FIGURE 5: Comparison of multimodal sports videos and their theoretical prediction results.

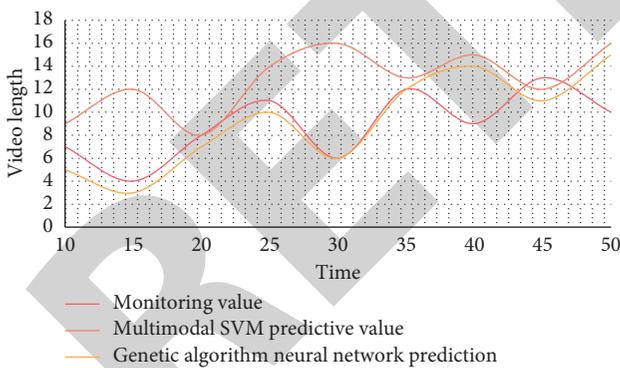


FIGURE 6: Comparison of multimodal sports videos and their theoretical prediction results.

an analytical foundation for achieving automatic video classification based on content. A frame in a video image is the smallest unit of sports video. It is necessary to extract features that can describe the video types in order to effectively cover different types of sports videos. Color,

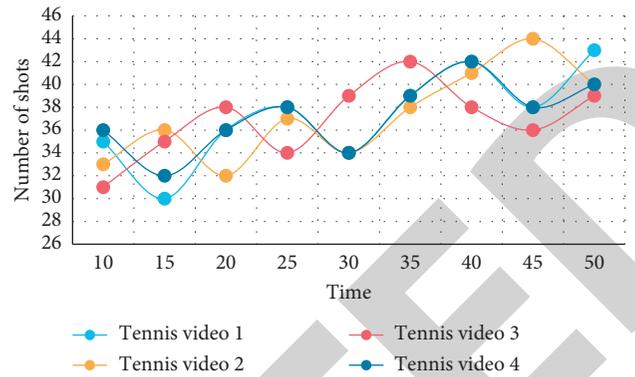


FIGURE 7: Experimental results of the first slow-motion detection algorithm.

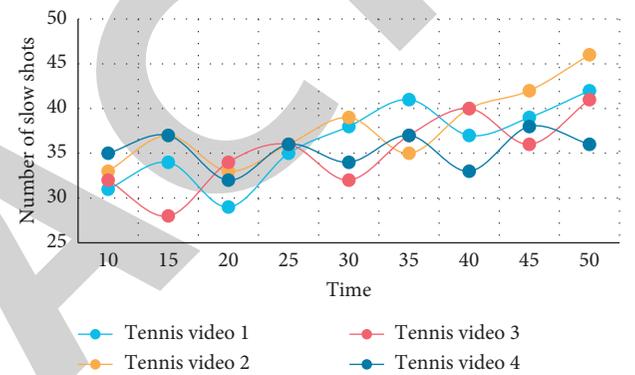


FIGURE 8: Experimental results of the second slow-motion detection algorithm.

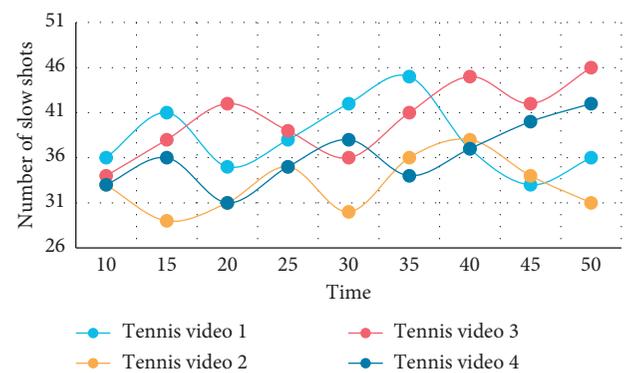


FIGURE 9: Experimental results of the third slow-motion detection algorithm.

texture, and edge are just a few of the current attributes. Color is one of the most widely used sports video features, which is most likely to attract people's attention. Color space is the basis of color feature extraction.

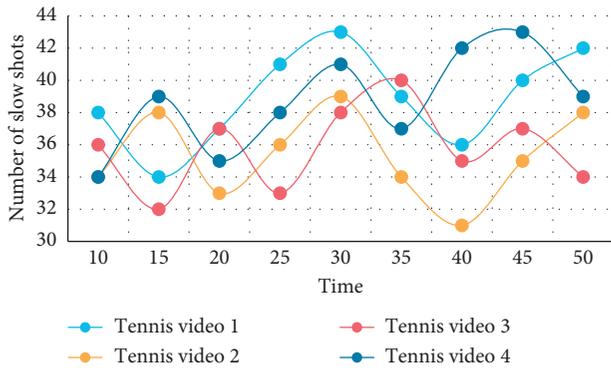


FIGURE 10: Experimental results of the fourth slow-motion detection algorithm.

5. Conclusions

Many researchers from around the world have spent a great deal of time and effort categorizing different types of sports videos. The manual and automatic stages of sports video research can be divided into two categories. The signal used for support vector machine learning modeling and prediction is a single empirical modal signal of multimodal sports video semantic analysis with a relatively simple change rule, which reduces the complexity of the nonlinear analysis signal, reducing the requirements for relevant calculation parameters and optimal kernel function selection. The first step in video postproduction is to locate the shot's cut point. This study examines and investigates commonly used shot classification algorithms before developing a tennis video shot classification algorithm based on the main color, the tennis court's ground wire, and the characteristics of target objects. The main color features of a tennis court, regular field line features, and audience shots, as well as the edge pixel features of player's close-up shot, are all included in this algorithm. Tennis video shots are divided into three categories: on-court global shot, audience shot, and player's close-up shot. Image frames are primarily analyzed by multimodal sports video semantics. Experiments have shown that this algorithm is effective at classifying data. Experiments on various sports game videos demonstrate that the proposed method for detecting and tracking players achieves satisfactory results by combining support vector machine classification with stadium segmentation.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] T. M. Le, V. Le, S. Venkatesh, and T. Tran, "Hierarchical conditional relation networks for multimodal video question answering," *International Journal of Computer Vision*, vol. 1, no. 8, pp. 1–24, 2021.
- [2] B. P. Joshi, A. Pant, X. Duan et al., "Multimodal video colonoscope for targeted wide-field detection of nonpolypoid colorectal neoplasia," *Gastroenterology*, vol. 150, no. 5, pp. 1084–1086, 2016.
- [3] Y. Wei, X. Wang, W. Guan, L. Nie, Z. Lin, and B. Chen, "Neural multimodal cooperative learning toward micro-video understanding," *IEEE Transactions on Image Processing*, vol. 29, no. 1, pp. 1–14, 2020.
- [4] H. Rafael, K. Marion, S. Bernd et al., "Effect of compression garments on the development of delayed-onset muscle soreness: a multimodal approach using contrast-enhanced ultrasound and acoustic radiation force impulse elastography," *Journal of Orthopaedic & Sports Physical Therapy*, vol. 48, no. 11, pp. 887–894, 2018.
- [5] A. Zadeh, R. Zellers, E. Pincus, and L.-P. Morency, "Multimodal sentiment intensity analysis in videos: facial gestures and verbal messages," *IEEE Intelligent Systems*, vol. 31, no. 6, pp. 82–88, 2016.
- [6] A. Collet, C. Ming, P. Sweeney et al., "High-Quality streamable free-viewpoint video," *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 69.1–69.13, 2016.
- [7] P. Kumar, S. Mukherjee, R. Saini, P. Kaushik, P. P. Roy, and D. P. Dogra, "Multimodal gait recognition with inertial sensor data and video using evolutionary algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 5, pp. 956–965, 2019.
- [8] S. M. Short, P. A. Anloague, and D. S. Strack, "Rehabilitation and return to sport following surgical repair of the rectus abdominis and adductor longus in a professional basketball player: a case report," *Journal of Orthopaedic & Sports Physical Therapy*, vol. 46, no. 8, pp. 697–706, 2016.
- [9] D. Li and G.-g. Ma, "Analgesic efficacy and quadriceps strength of adductor canal block versus femoral nerve block following total knee arthroplasty," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 24, no. 8, pp. 2614–2619, 2016.
- [10] W. Marc, L. Wen, and Y. Fumio, "Learning change from synthetic aperture radar images: performance evaluation of a support vector machine to detect earthquake and tsunami-induced changes," *Remote Sensing*, vol. 8, no. 10, pp. 1–21, 2016.
- [11] D. Ryu, O. Choi, and J. BaIk, "Value-cognitive boosting with a support vector machine for cross-project defect prediction," *Empirical Software Engineering*, vol. 21, no. 1, pp. 43–71, 2016.
- [12] M. Balachandran and L. Jooyoung, "SVMQA: support-vector-machine-based protein single-model quality assessment," *Bioinformatics*, vol. 33, no. 16, pp. 2496–2503, 2017.
- [13] J. Sun, H. Fujita, P. Chen, and H. Li, "Dynamic financial distress prediction with concept drift based on time weighting combined with adaboost support vector machine ensemble," *Knowledge-Based Systems*, vol. 120, 2017.
- [14] S. Kim, J.-H. Jhong, J. Lee, and J.-Y. Koo, "Erratum to: meta-analytic support vector machine for integrating multiple omics data," *BioData Mining*, vol. 10, no. 1, p. 8, 2017.
- [15] P. T. Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery," *Sensors*, vol. 18, no. 1, p. 18, 2018.
- [16] G. Milan, S. Shahaboddin, R. Zaidi, D. Petković, S. Ch, and S. Trajkovic, "Long-term precipitation analysis and estimation of precipitation concentration index using three support vector machine methods," *Advances in Meteorology*, vol. 2016, Article ID 7912357, 11 pages, 2016.
- [17] W. M. Lin, C. S. Tu, R. F. Yang, and M. T. Tsai, "Particle swarm optimisation aided least-square support vector

- machine for load forecast with spikes,” *IET Generation, Transmission & Distribution*, vol. 10, no. 5, pp. 1145–1153, 2016.
- [18] K. Y. Bae, S. J. Han, and K. S. Dan, “Hourly solar irradiance prediction based on support vector machine and its error analysis,” *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 935–945, 2017.
- [19] X.-B. Jin, W.-Z. Zheng, J.-L. Kong et al., “Deep-learning temporal predictor via bidirectional self-attentive encoder-decoder framework for IOT-based environmental sensing in intelligent greenhouse,” *Agriculture*, vol. 11, no. 8, p. 802, 2021.
- [20] I. Mladenovic, S. Sokolov-Mladenovic, M. Milovancevic, D. Marković, and N. Simeunović, “Management and estimation of thermal comfort, carbon dioxide emission and economic growth by support vector machine,” *Renewable and Sustainable Energy Reviews*, vol. 64, no. 2, pp. 466–476, 2016.
- [21] B. R. Froz, A. Filho, A. C. Silva et al., “Lung nodule classification using artificial crawlers, directional texture and support vector machine,” *Expert Systems with Applications*, vol. 69, pp. 176–188, 2016.
- [22] M. A. Ghorbani, H. A. Zadeh, M. Isazadeh, and O. Terzi, “A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction,” *Environmental Earth Sciences*, vol. 75, no. 6, pp. 1–14, 2016.
- [23] H. Samma, C. P. Lim, J. M. Saleh, and S. A. Suandi, “A memetic-based fuzzy support vector machine model and its application to license plate recognition,” *Memetic Computing*, vol. 8, no. 3, pp. 1–17, 2016.
- [24] C. Li, S. Li, and Y. Liu, “A least squares support vector machine model optimized by moth-flame optimization algorithm for annual power load forecasting,” *Applied Intelligence*, vol. 45, no. 4, pp. 1–13, 2016.
- [25] H. Wang, Z. Zhou, and Y. Xu, “An improved ν -twin bounded support vector machine,” *Applied Intelligence*, vol. 48, no. 3, pp. 1–13, 2017.
- [26] A. Asfaram, M. Ghaedi, M. H. A. Azqhandi, A. Goudarzi, and M. Dastkhooon, “Statistical experimental design, least squares-support vector machine (LS-SVM) and artificial neural network (ANN) methods for modeling the facilitated adsorption of methylene blue dye,” *RSC Advances*, vol. 6, no. 46, pp. 40502–40516, 2016.
- [27] K. Kesorn and S. Poslad, “Enhanced sports image annotation and retrieval based upon semantic analysis of multimodal cues,” in *Pacific-rim Symposium on Image and Video Technology*, pp. 817–828, Springer, Berlin, Germany, 2009, Advances in Image and Video Technology.
- [28] K. He and X. Li, “A quantitative estimation technique for welding quality using local mean decomposition and support vector machine,” *Journal of Intelligent Manufacturing*, vol. 27, no. 3, pp. 525–533, 2016.
- [29] M. Buyukyildiz and S. Y. Kumcu, “An estimation of the suspended sediment load using adaptive network based fuzzy inference system, support vector machine and artificial neural network models,” *Water Resources Management*, vol. 31, no. 4, pp. 1343–1359, 2017.
- [30] S. Cogill and L. Wang, “Support vector machine model of developmental brain gene expression data for prioritization of Autism risk gene candidates,” *Bioinformatics*, vol. 2, no. 23, pp. 3611–3618, 2016.
- [31] M. A. Ghorbani, R. Khatibi, A. Goel, and M. H. Fazelifard, “Modeling river discharge time series using support vector machine and artificial neural networks,” *Environmental Earth Sciences*, vol. 75, no. 8, pp. 1–13, 2016.
- [32] S. Zhu, J. Zhou, L. Ye, and C. Meng, “Streamflow estimation by support vector machine coupled with different methods of time series decomposition in the upper reaches of Yangtze River, China,” *Environmental Earth Sciences*, vol. 75, no. 6, pp. 531.1–531.12, 2016.
- [33] W. C. Hsu, L. F. Lin, C. W. Chou, Y.-T. Hsiao, and Y.-H. Liu, “EEG classification of imaginary lower limb stepping movements based on fuzzy support vector machine with kernel-induced membership function,” *International Journal of Fuzzy Systems*, vol. 19, no. 2, pp. 1–14, 2016.
- [34] Y. Huang, L. Zhang, G. Lian et al., “A novel mathematical model to predict prognosis of burnt patients based on logistic regression and support vector machine,” *Burns*, vol. 42, no. 2, pp. 291–299, 2016.