Research Article

An Automatic Classification Method of Sports Teaching Video Using Support Vector Machine

Zhang Min-qing and Li Wen-ping

Department of PE, Zhejiang Sci-Tech University, Hangzhou 310018, China

Correspondence should be addressed to Li Wen-ping; zhangminqing@zstu.edu.cn

Received 1 September 2021; Accepted 7 October 2021; Published 19 October 2021

Academic Editor: Le Sun

Copyright © 2021 Zhang Min-qing and Li Wen-ping. 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.

There are many different types of sports training films, and categorizing them can be difficult. As a result, this research introduces an autonomous video content classification system that makes managing large amounts of video data easier. This research provides a video feature extraction approach using a support vector machine (SVM) video classification algorithm and a mix of video and audio dual-mode characteristics. It automates the classification of cartoons, ads, music, news, and sports videos, as well as the detection of terrorist and violent moments in films. To begin, a new feature expression scheme, the MPEG-7 visual descriptor subcombination, is proposed based on an analysis of the existing video classification algorithms, with the goal of addressing the problems in these algorithms. This is accomplished by analyzing the visual differences of the five video classification algorithms. The model was able to extract 9 descriptors from the four characteristics of color, texture, shape, and motion, resulting in a new overall visual feature with good results. The results suggest that the algorithm optimizes video segmentation by highlighting disparities in feature selection between different categories of films. Second, the support vector machine's multivideo classification performance is improved by the enhanced secondary prediction method. Finally, a comparison experiment with current related similar algorithms was conducted. The suggested method outperformed the competition in the accuracy of video classification in five different types of videos, as well as in the recognition of terrorist and violent incidents.

1. Introduction

The rapid expansion of Internet and 4G network bandwidth has resulted in the explosive proliferation of various multimedia video streams in recent years. A good example is the rise of sports and entertainment. People in current culture can access all types of sports games on the Internet at any time and from any location. The large demand has resulted in a massive number of video resources; however, the traditional human annotation approach for text retrieval has significant limits and cannot cope with the increasing demand in its existing form. As a result, the researchers propose that the video frequency information is retrieved and processed using an automatic video categorization method. In entertainment, education, security, military, and other industries, automatic video classification has a large commercial potential. Reference [1] proposes a network video classification method based on two-way

communication of heterogeneous information. Wen Xian proposes a video image distortion detection and classification method based on convolutional neural network [2]. Reference [3] proposed an improved local feature vector classifier for fast video classification. However, all the above methods train and classify the features of the whole video without considering the training of video frame classifier.

Traditional text-based information query technology cannot meet the expectations of users due to the enormous number of video information data and the absence of structure in the form. Despite the fact that current digital video can be labeled in a variety of ways at the production stage, automatic video classification technology is still needed. The following are the reasons: to begin with, many previously existing videos have not been labeled, making classification extremely difficult. While human tagging can fix this problem, it is a time-consuming and costly operation, and keeping up with the continually rising amount of films

will be challenging in the near future. Another point, which is also the most important, is that while current video watermarking technology is gradually improving, such that video can add watermark to produce phase or label, its resistance to attack ability is limited after adding watermarks or video label, and all of this information can be lost if some man-made error or accident occurs. While the content-based method can prevent these issues, the categorization detection results of the video will not change because it is based on the actual material. As a result, automatic categorization technology is still required to manage and alter video. As an example, television is one of the current household's developing directions toward intelligent personalized development. It is hoped that with automatic content identification, TV video filtering, and brightness contrast based on broadcast content of adaptive control and intelligent function in the implementation process, these functions cannot be independent of the automatic classification of video. Therefore, the development of video classification technology is of great significance in various fields.

This article, in conjunction with this paradigm, focuses on two types of important technologies for video information transmission control: automatic categorization based on video content and covert tag based on video concealed channel. Media websites can automatically classify video content and refine it to all kinds of programs (such as comedy movies, perfume advertisements, MTV, badminton programs, and weather news) using automatic video classification technology and perform automatic preliminary screening of bad video information. Implementation technology based on video concealed identity can distinguish between different video sources, the media must release video information unified hidden identification technology processing on its website according to management requirements, and by detecting the hidden on the video identification it can distinguish between publishing media sites and provide real-time monitoring and tracking video. Therefore, this paper proposes an automatic sports video frequency classification method based on support vector machine. The experimental results show that the proposed method can realize classification quickly and obtain high classification accuracy.

2. Video Classification Methods

The phenomenal growth of various multimedia video streams has coincided with the rapid expansion of Internet and 4G network bandwidth. A good example is the rise of sports and entertainment. People in current culture can access all types of sports games on the Internet at any time and from any location.

Literature [4] suggested a sports video classification method based on a type marking shot and a bag of words model, which resulted in high classification accuracy. Furthermore, literature [5] proposes the notion of principal component analysis based on an updated SVM algorithm to further improve classification efficiency, resulting in a rapid classification speed. As a result, this research provides a

support vector machine (SVM)-based automatic sports video categorization approach that combines the visual bag of words model with principal component analysis technology from the previous literature. Figure 1 depicts the visual automatic categorization approach for sports videos used in this paper, as well as its principle. Frame recognition and video classification are the major components of this approach. The upper half is divided into frame recognition, which essentially completes the development of an image training database and the collection and extraction of key characteristics, as shown in Figure 1. The lower section is divided into video classification, which mostly completes key frame extraction, outlier elimination, and type judgment in the video stream. Visual vocabulary, word frequency vector generation, and video classifier are three of the most popular components.

Video content analysis is a more advanced step of video processing that investigates the use of machines to analyze and identify the content or semantics of video in order to reach a definitive conclusion. However, in order for the computer to recognize the video, it is required to first obtain different visual features. Video feature extraction is the process of extracting these features, which are a form of representation and description of the video content. The original attributes or properties of video, encompassing both visual and auditory modes, are referred to as video features. Some of these aspects are natural features that humans can sense directly, such as a region's color, texture, or intensity; others, such as transform spectra, histograms, and moments, are artificial features that require variation or measurement. The traits that identify one sort of video from another are extracted as a result of feature extraction. The results of these feature extraction are expressed in certain ways for the computer to understand. In this paper, a series of representative audio and video features are extracted from visual and audio modes, which serve as the basis for constructing an automatic video classification algorithm.

2.1. Analysis and Extraction of Bimodal Features. Bimodal feature refers to both visual and audio modes. Feature analysis is the basis of video content analysis. It exists in each process of video content analysis technology. It directly restricts the description ability of media object content and affects the quality of subsequent content analysis and the effectiveness of application system. This section analyzes video features from visual and audio modes.

Human vision is a significant source of information. Visual perception accounts for 80% of the information people acquire from their surroundings. As a result, visual elements play a critical role in video classification research. Domain-specific visual features are related to specific applications, such as facial features, fingerprint features, and handwriting, whereas universal visual features are used to characterize the common features of all movies, such as color and texture. The object studied in this paper is the classification of video types, rather than the classification of specific objects, so the general visual features are mainly used. The following is a detailed introduction [6].

FIGURE 1: Schematic diagram of the automatic classification method of sports video.

2.1.1. Color Features. In content-based video feature extraction, there are a variety of color expression approaches. Other methods include the cumulative histogram, color correlation vector, color correlation graph, color clustering, and mass-tone-based algorithms [7].

In computer image processing, the image must be represented by data, color space is a means of using data to represent color, and there can be a range of data representation modes, or color spaces, for a color representation. The two color spaces utilized in this article are introduced briefly as follows. There are two types of color spaces: RGB and HSV [8].

The RGB (Red-Green-Blue) spatial model: this is the most basic image processing model, which is based on Cartesian coordinates and is frequently used in image display. However, the color space has nothing to do with the intuitive color concept. Each color image, according to this concept, is made up of three primary color planes, with each pixel point consisting of red, green, and blue primary colors.

HSV spatial model (Hue-saturation-Value): Alvy Ray Smith established this concept in 1978. It is a three-color model that has been transformed in a nonlinear way. This color model is a perception-based color scheme. H stands for the chrominance of dominating spectral hues, S for color saturation, and V for brightness in this spatial model. HSV color attribute mode is a way of defining color based on hue, saturation, and lightness, the three primary color characteristics [9].

2.1.2. Texture Features. Texture, like color, is a significant aspect of an image. However, there has not been a consistent and clear meaning for the term texture. When we talk about

the texture of an image, we usually refer to its degree of roughness, smoothness, and regularity. It refers to changes in the grayness or hue of picture pixels that are spatially related to statistics in image processing. Image texture can be analyzed in a variety of ways, the most common of which are statistical and structure approaches. The statistical method is the first texture analysis method proposed and has been playing an important role; this method is the use of texture statistical characteristics and laws to describe the texture. It is applicable to the widespread existence of natural texture and also to artificial texture, and it has been in a relatively mature stage [10].

- (1) Methods based on spatial domain, such as gray histogram statistics and gray cooccurrence matrix: the principle of them is simple, easy to implement, and suitable for natural texture.
- (2) Methods based on transform domain features, such as Fourier transform and Gabor transform: according to the human vision mechanism, this kind of methods uses the filter which obtains the good localization feature simultaneously, so as to obtain the more ideal texture characteristic [11].
- (3) Methods based on models, such as the simultaneous model and Markov random field model: its downside is that it requires a lot of computation and that expressing natural texture with a single model is challenging. Structure analysis, unlike statistical methods, shows that complex texture can be made up of simple texture primeless arranged and merged according to particular criteria. The issue comes in extracting pimples and defining arranging rules that can be used with standard texture photos. Since we

analyze various kinds of videos, including artificial video and natural video, and the texture is not necessarily regular, this paper adopts widely applicable statistical methods to analyze the texture.

2.2. Classification Modeling Module. This module is divided into two parts: the classifier training module and the classification determination module. The training module's major duty in this system is to train the SVM classifier using the training data as input. The selection of SVM classifier parameters, such as the kernel function and parameter determination in the kernel function, is the fundamental problem addressed in this module. This system's training module is expected to determine the best optimal parameters for reducing structural risk. The classification decision module's major purpose is to evaluate video classification results and predict unknown video types. The fundamental difficulty for the SVM classification model is how to address numerous classification problems with a binary classifier, and classification outcomes will differ depending on the decision-making approach used. The categorization decision's substance and enhanced methodology will be discussed in depth below. Based on the fusion of MPEG-7 visual descriptors and support vector machine (SVM) classifier, this paper proposes the process of automatic video content classification algorithm [12]. The algorithm process can be divided into the following seven steps:

- (1) Input video samples and preprocess the original video data after sample screening, including mirror head segmentation, video frame extraction, and other contents
- (2) Based on the analysis of video content and style, the method proposed in Section 2 is used to extract 9 Kinds of MPEG-7 visual descriptors and fuse them as the overall features of the video
- (3) The video features are combined into feature vector space and divided into five types of training data to prepare for the learning and training of support vector machine
- (4) The improved classification algorithm proposed in this paper is used to construct the multiclassification model of support vector machine
- (5) The training data in step 3 is used for learning and training, and the best parameters of the classifier are obtained by cross validation method
- (6) The optimal parameters C and γ were used to train the whole training set to obtain the support vector machine model
- (7) The obtained classification model is used to predict the samples of unknown categories, so as to judge the video category

3. Support Vector Machine (SVM) Theory

Vanpik [13, 14] and his colleagues from Bell LABS proposed the support vector machine (SVM) as a new machine learning technique based on statistical learning theory. It employs a

FIGURE 2: Comparison of classification hyperplane.

novel method of learning. SVM's learning and training criteria, unlike earlier learning methods, are based on the objective function rather than traditional minimization. Training SVM is comparable to finding the objective function with the biggest boundary, i.e., solving a large-scale quadratic programming problem, according to the idea of least structural risk (QP).

It is a method for determining the best classification surface in the original space or the high-dimensional space produced after projection in order to distinguish between two types of samples. Maximizing the gap between the two categories ensures the lowest experience risk (0) in statistical learning theory; maximizing the classification gap actually lowers the confidence range in the generalization bounds and so minimizes the real risk. We hope to identify a hyperplane in the R^d space for a set of data, and that this hyperplane can partition this set of data into two groups (such as class A and class B). Figure 2 depicts the distinction between classes A and B.

By comparing the left and right Figure 2, we can find that the hyperplane (dotted line) found in the left figure has a close distance between two parallel hyperplanes (solid line) tangent to two groups of points [15], while the right figure has a large margin. Since we hope to find the parameter that separates the two groups of data points more widely, we consider the figure on the right to be a better hyperplane [16]. Given the training data set,

$$\{x_i, y_i\}, i = 1, 2, \dots, n, x_i \in R^d, x_i \in \{1, -1\}. \quad (1)$$

We hope to use the training data to find an optimal hyperplane H in order to classify the unknown x_i . Principle of SVM is shown in Figure 3.

In Figure 3, the solid line is the hyperplane we found, $H1$ and $H2$ are called support hyperplane, and we hope to find the best classification hyperplane to maximize the gap between the two support hyperplanes.

$$\begin{aligned} H1: w^T x + b &= 1, \\ H2: w^T x + b &= -1. \end{aligned} \quad (2)$$

3.1. Support Vector Machine Binary Classification Algorithm. The reason why we choose SVM as the classifier is that SVM is not only guaranteed by statistical learning theory as its generalization, but also, on the premise of correct use, its accuracy is not far from that of K nearest neighbor, neural networks,

FIGURE 3: Principle of SVM.

decision tree, and other methods. However, the advantage of SVM is that it is easier to use [17]. The following is a brief introduction to SVM.

Let the linearly separable set of samples $\{x_i, y_i\}, i = 1, 2, \dots, n, x_i \in R^d, x_i \in \{1, -1\}$ be category markers. Assuming that the training set can be linearly divided by a hyperplane, the hyperplane is denoted as $w^T x + b = 0$, which satisfies

$$\begin{aligned} \forall y = 1, wx + b &\geq +1, \\ \forall y = -1, wx + b &\leq -1. \end{aligned} \quad (3)$$

If the sum of the closest distances between the two training points and the hyperplane is satisfied at the same time, the interval reaches the maximum. Because the distance between the support vector and the hyperplane is $1/\|w\|$, the distance between the support vectors is $2/\|w\|$. Therefore, the problem of constructing the optimal hyperplane is transformed into the following constraint minimization problem:

$$\min \frac{1}{2} \|w\|^2 \text{ s.t. } y_i (w \cdot x_i + b) \geq 1. \quad (4)$$

By the Lagrange multiplier method and the introduction of Lagrange multiplier, the constrained extremum problem can be transformed into a simple duality problem. By seeking the optimal solution of the duality problem, the optimal solution of the original problem can be obtained. Finally, the decision function of the classifier is obtained:

$$f(x) = \text{sgn} \left(\sum_i a_i y_i (w \cdot x_i + b) \right). \quad (5)$$

For the linear inseparable case, the R^N is mapped to R^F space ($F > N$). The dimensions of the generated attribute space vary greatly when different mapping functions are used. As direct dimension mapping is difficult to carry out, in order to avoid this problem, SVM adopts the kernel function mechanism, making the final discriminant function become

$$f(x) = \text{sgn} \left(\sum_i a_i y_i K(x_i, x) + b \right), \quad (6)$$

where $K(x_i, x)$ is the inner product form of the attribute space vector. Because the specific form of spatial mapping function does not need to be known, the calculation of classification function coefficients only involves the inner product of the image space vector.

There are four types of kernel functions used in support vector machines:

$$\begin{aligned} K(x_i, x) &= x_i^T x_j, \\ K(x_i, x) &= (\gamma x_i^T x_j + r)^d, \\ K(x_i, x) &= \exp(-\gamma \|x_i - x_j\|^2), \\ K(x_i, x) &= \tanh(\gamma x_i^T x_j + r), \end{aligned} \quad (7)$$

where γ and d are both nuclear parameters.

3.2. Multiclassification Algorithm of Support Vector Machine.

One of the most important topics in SVM research is how to expand the binary classification method to multiclassification. Currently, domestic and international academics have offered a number of different promotion tactics. It is usually possible to alter the SVM algorithm's design or to develop a multiclass classifier by combining numerous two-class classifiers. The 1-R (one-Against-REST) technique, 1-1 (one-Against-One) method, and DAG (directed acyclic graph) method are the latter ways. They are introduced in order as follows.

3.2.1. 1-R Method. One of the first SVM multiclassification approaches proposed was the 1-R method. To create all conceivable dichotomies, a class of training samples is segregated from the rest of the training samples, and then a combination method is used to combine all the trained dichotomies to answer multiclassification issues. This method creates n SVMs for n classification problems, each of which differentiates one category of data from other categories, as shown in Figure 4.

This method is simple and intuitive, but the disadvantage is that there are nonseparable regions, and if the distribution of training samples is not balanced, the accuracy will be affected. In addition, all samples must be trained to construct dichotomous SVM each time, and the computational and time complexity is relatively large.

3.2.2. 1-1 Method. The 1-1 method is also a classification method based on two types of questions, but the two types of questions here are extracted from the original multitype questions. The specific method is to select two different categories to form a SVM subclassifier by the 1-1 method. There are $n(n-1)/2$ SVM subclassifiers in this way.

In the test, the "voting method" is adopted; that is, each test sample will get a possible category number after a dichotomous SVM, that is, one vote. After the sample passes all dichotomous SVM, the votes obtained are counted, and the category with the highest number of votes is the category that the sample is most likely to belong to. Figure 5 shows the

FIGURE 4: 1-R strategy of SVM.

FIGURE 5: 1-1 strategy of SVM.

1-1 voting process. For example, there are 5 types of samples, among which type 1 has 2 votes, type 2 has 4 votes, type 3 has 1 vote, type 4 has 1 vote, and type 5 has 2 votes; then, the final classification result is 2 types.

3.3. The Nature and Implementation of Support Vector Machines. This section briefly discusses the basic mathematical properties of SVM and its implementation:

- (1) The statistical learning theory underpins SVM. Traditional nonparametric approaches, such as the closest domain or neural network, aim to reduce classification error in the training sample set as much as possible. SVM reduces structural risk, which is the chance of misclassifying data points collected at random from fixed data sets with uncertain probability distributions. When classifying an arbitrarily distributed test data set, the SVM theory provides the upper limit of the chance of misclassification.
- (2) SVM concentrates classification-related information contained in the training data set into support vector, which greatly reduces the subset of training data that is effective for classification, thus improving the efficiency of classification.
- (3) By selecting different kernel functions, different classifier structures can be established. It can be p-order polynomial classifier, Gaussian RBF

FIGURE 6: Structure of SVM algorithm.

classifier, or multilayer perceptron MLP and other different structures. Structure of SVM algorithm is shown in Figure 6.

- (4) SVM is usually classified in high-dimensional space, and the relationship between computational efficiency and training success rate must be comprehensively considered in practical application.

The main ideas of SVM training are as follows:

- (1) Replace the original proposition with a uniformly increasing objective function dual proposition.
- (2) Find decomposition algorithm so that only a small part of the data in the training sample set can be processed.
- (3) The best solution must meet the QP problem's KKT condition. The QP issue in SVM is defined as the search for the global minimum of a basin-like objective function. The minimal value can only be found by using a hypercube and a hyperplane. The objective function of the D-matrix in the QP problem is usually basin-shaped (positive definite) or flat-bottle-shaped (positive semidefinite) and cannot be saddle-shaped (nondefinite). As a result, either a single global optimum solution or a continuous equivalent optimal solution exists for SVM.

4. Simulation Experiment Analysis

The experimental platform of this paper is Windows 7 system PC, and the simulation experiment environment is MATLAB 2012. Multiple types of mixed sports video data sets were used for classifying experiments. The storage size was 11.2 GB, the duration was 3 000 min, and it contained up to 46 videos. Video content includes seven categories: basketball, badminton, football, table tennis, snooker, tennis, and volleyball.

FIGURE 7: Comparison results of the recall rate.

In addition, recall rate and precision rate were used as evaluation indexes. The calculation formula of recall R and precision P are as follows:

$$R = \frac{\sum_{i=1}^M a_i}{\sum_i \text{count}(G_i)}, \quad (8)$$

$$P = \frac{\sum_{i=1}^M a_i}{\sum_i \text{count}(E_i)},$$

where M represents the number of mixed videos in the input sports videos; G_i represents the positive example type of mixed video H_i ; E_i represents the positive example type of mixed video H_i classification; a_i indicates the number of positive examples of correct classification in video H_i .

In order to further illustrate the effectiveness of the proposed algorithm, the video classification results obtained by the proposed algorithm on the sports video data set are compared with those obtained in the [6-7], and the results are shown in Figures 7 and 8.

Figures 7 and 8 show that, when compared to the other two approaches, the proposed technique's classification accuracy for various types of sports films is relatively high, indicating that the recall rate and precision rate remain high, indicating that the suggested method is effective.

The classification accuracy of single descriptors is compared in this work, which indicates each descriptor's contribution to the total classification accuracy of videos to some extent, as well as its usefulness in identifying a specific class of films. Some useful and required descriptors are chosen as the feature description for the secondary prediction based on the results of the experiment. The original 1-1 SVM was chosen for this experiment. Table 1 shows a comparison of the categorization accuracy of a single descriptor.

FIGURE 8: Comparison results of the precision ratio.

As can be seen from Table 1, among the color descriptors, the overall classification accuracy of GoP is the highest, reaching 81.8%, which indicates that the preferred GoP descriptor can distinguish the five categories of videos to the maximum extent. Other color descriptors also make corresponding contributions, among which CSD has the highest classification accuracy of sports videos, up to 91.6%, while CCD has a good classification accuracy of sports videos, which may be related to the relatively fixed color information of sports videos. As a result, in terms of quadratic prediction descriptors, this paper uses CSD for sports videos and GoP for other films. Both HTD and EHD show

TABLE 1: Comparison of classification accuracy of a single descriptor.

Descriptor	DCD	CLD	CSD	GoP	HTD	EHD	RSD	CCD	MID
Cartoon	44.0	63.3	78.7	76.0	79.7	84.0	58.7	70.0	30.7
Advertising	36.4	62.1	60.5	73.3	60.5	67.3	53.0	58.3	23.5
Music	73.3	68.7	81.3	85.3	70.3	75.0	68.7	72.0	77.7
News	71.7	70.1	82.5	87.3	77.9	82.7	59.4	72.3	82.7
Sports	82.5	82.5	91.6	87.3	82.2	82.9	73.0	83.9	54.6
The overall	61.6	69.3	78.9	81.8	74.1	78.4	62.6	71.3	53.8

good classification accuracy for cartoon video in the texture descriptor, indicating that cartoon video frequency is the easiest to differentiate from texture. Because cartoon video is an artificial video, its texture is smoother than nonartificial video, making it possible to recognize cartoon video by texture. In this study, HTD and EHD are utilized to describe quadratic prediction using cartoon video.

5. Conclusion

This paper proposes a sports video automatic classification method based on support vector machine and conducts classification experiments using mixed sports video data sets of various types, based on the analysis and research of existing content-based video retrieval algorithms and the theory of support vector machine. The experimental results show that the sports video classification algorithm described in this research can efficiently manage large amounts of sports video with complicated samples and classify them quickly and effectively with high accuracy.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] B. Tu Truong, S. Venkatesh, and C. Dorai, "Automatic genre identification for content-based video categorization," in *Proceedings of the 15th International Conference on Pattern Recognition. ICPR-2000*, pp. 230–233, Barcelona, Spain, September 2000.
- [2] M. K. Geetha and S. Palanivel, "HMM based automatic video classification using static and dynamic features," in *Proceedings of the International Conference on Computational Intelligence and Multimedia Applications*, pp. 277–281, Sivakasi, India, December 2007.
- [3] L. Q. Xu and Y. Li, "Video classification using spatial-temporal features and PCA," in *Proceedings of the 2003 International Conference on Multimedia and Expo (ICME'03)*, pp. 485–488, Baltimore, MD, USA, July 2003.
- [4] Y. Zhu and M. Zhong, "SVM-based video scene classification and segmentation," in *Proceedings of the International Conference on Multimedia and Ubiquitous Engineering*, pp. 407–412, Busan, Korea (South), April 2008.
- [5] H. Lee, S. Lee, and T. Nam, "Implementation of high performance objectionable video classification system," in *Proceedings of the 8th International Conference Advanced Communication Technology/ICACT 2006*, p. 4, Phoenix Park, South Korea, February 2006.
- [6] Y. Yan, H. Shen, G. Liu, Z. Ma, C. Gao, and N. Sebe, "GLocal tells you more: c," *Computer Vision and Image Understanding*, vol. 124, pp. 99–109, 2014.
- [7] Y. H. Ng, M. Hausknecht, S. Vijayanarasimhan, V. Oriol, M. Rajat, and T. George, "Beyond short snippets: deep networks for video classification," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4694–4702, IEEE, Boston, MA, USA, June 2015.
- [8] Y. Yuan, Q.-B. Song, and J.-Y. Shen, "Automatic video classification using decision tree method. Machine Learning and Cybernetics," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 1153–1157, Beijing, China, November 2002.
- [9] Y. Yuan, J. Shen, and Q. Song, "An automatic video classification method based on data mining," in *Proceedings of the ICMLC'03*, Xi'an China, November 2003.
- [10] Y. Zheng, L. Duan, T. Qi, and J. S. Jin, "TV commercial classification by using multi-modal textual information. Multimedia and expo," in *Proceedings of the 2006 IEEE International Conference on Multimedia and Expo*, pp. 497–500, Toronto, ON, Canada, July 2006.
- [11] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.
- [12] E. Spyrou, H. Le Borgne, T. Mailis, E. Cooke, Y. Avrithis, and N. O'Connor, "Fusing MPEG-7 visual descriptors for image classification," *Lecture Notes in Computer Science*, vol. 3697, pp. 847–852, 2005.
- [13] A. Mittal and L.-H. Cheong, "Addressing the problems of Bayesian network classification of video using high-dimensional features. Knowledge and Data Engineering," *IEEE Transactions on*, vol. 16, no. 2, pp. 230–244, 2004.
- [14] H. J. Zhang, J. Wu, D. Zhong, and S. W. Smoliar, "An integrated system for content-based video retrieval and browsing," *Pattern Recognition*, vol. 30, no. 4, pp. 643–658, 1997.
- [15] M. Flickner, H. Sawhney, and W. Q. Niblack, "Query by image and video content: the QBIC system," *Computer*, vol. 28, no. 9, pp. 23–32, 1995.
- [16] N. V. Patel and I. K. Sethi, "Video shot detection and characterization for video databases," *Pattern Recognition*, vol. 30, no. 4, pp. 583–592, 1997.
- [17] H. G. Barrow, "Parametric correspondence and chamfer matching," in *Proceedings of the 5th Int. Joint Conf, Artificial Intelligence*, pp. 659–663, Cambridge, MA, USA, August 1977.