

# Video Multiple Classification Algorithm Based on SVM

Chao Jiang<sup>1</sup> and Shuguang Wang<sup>2</sup>

<sup>1,2</sup>*Jilin Communications Polytechnic, Changchun 130012, china*  
<sup>1</sup>*jch0323@sina.com.cn* and <sup>2</sup>*Wang\_shuguang@sohu.com*

## Abstract

*The performance of video automatic classification algorithm depends largely on the extraction of video features and selection of classification algorithm. From the perspective of video contents and video style type, the paper presents a new feature representation scheme, i.e. MPEG-7 visual description sub-combination model, a new method based on support vector machine (SVM) to solve problems with existing algorithms, by analyzing visual differences between five types of videos. Also we improve the classifier decision scheme and then propose the secondary prediction mechanism based on SVM 1-1 approach, improving the accuracy of SVM multi-classification method. The experimental results indicate that the proposed method manifests differences of different videos about feature selection, enhances the discrimination ability of videos pending for classification and increases the effectiveness of SVM multi-video classification.*

**Keywords:** video classification, feature, support vector machine

## 1. Introduction

The automatic classification method of video contents is an important research topic in the computer vision field [1-3]. It provides convenience for the management of increasing video data. The content-based video automatic classification, as one key technology in video spreading control, plays a critical role in the orderly management of network media [4-6]. The application based on video automatic classification technology is helpful for media website to classify automatically enormous video contents, for more effective organization, storage and retrieval; also it can realize automatically the initial screen of bad video information, e.g. terror and violence videos [7-8].

What the classification needs to solve is to assign observed values to one category according to those values represented by objects pending for recognition [9-10]. As a starting point, we seek one kind of classification discrimination function or decision criterion and design decision classification model [11]; then determine parameters of the model based on the training set; next, use the model for discrimination decision and use the decision function to discern unknown sample point belongs to which type. Firstly the paper gives the whole classifying framework and steps of classification algorithm; then regarding problems in the discrimination decision algorithm in the design of SVM-based multi-classifiers, we discuss and make certain improvements. Finally, simulation test is conducted for the proposed method.

## 2. SVM-Based Video Automatic Classification System

### 2.1. Video Automatic Classification System Framework

The video automatic classification system framework proposed here includes two key parts: feature extraction and classifier design, dividing videos automatically into five types like variety, education, automobile, life and technology. On the feature side, we

fetch nine kinds of MPEG-7 visual descriptors as video characteristics. Experimental findings suggest that using nine features can achieve ideal classifying effects. On the classifier side, we adopt support vector machine as classifier.

The existing common classifiers are: Bayes classifier based on Bayes decision theory, K nearest classifier, decision tree classifier, principle of minimization risk, various neural network classifiers of self-organization mapping, and burgeoning SVM classifiers based on the statistical learning theory [12]. The statistical learning theory builds on a set of solid theoretical basis, especially offering a unified framework to solve learning problems of limited samples. Why we choose SVM as classifier is because SVM is more suitable for the problem of limited sample classification; and to a great degree, it overcomes shortcomings with traditional classifying methods such as model selection, over-learning, nonlinearity, multi-dimension, local minimum point, along with powerful generalization ability.

## 2.2. Process of Video Automatic Classification Algorithm

Based on fusion of MPEG-7 visual descriptors and SVM classifier, we put forward the process of video content automatic classification algorithm:

- (1) Input video samples and pre-treat original video data after sample selection, including shot segmentation, video frame extraction;
- (2) Based on the analysis of video contents and styles, we use the method introduced in part II to extract nine kinds of MPEG-7 visual descriptors and fuse them as video's overall feature;
- (3) Combine video features to characteristic vector space and classify them into five training data, making preparations for learning training of SVM;
- (4) Use the improved classification algorithm to construct SVM multi-classification model;
- (5) Utilize training data in Step (3) for learning training; with the method of cross validation, we can get the optimal parameters of classifier;
- (6) Use the optimal parameter  $c$  and  $\gamma$  to train the whole training set as to obtain SVM model;
- (7) Apply the obtained classification model to predict samples whose class is unknown, as to determine video type.

## 2.3. Feature Selection and Extraction

In order to classify automatically videos, we withdraw nine MPEG-7 visual descriptors from various input videos; then, fuse those descriptors to the video overall feature and input SVM classifier for training and testing to get final classification results. MPEG-7 visual description is shown in Table 1.

# 3. Multi-Classifer Design of Support Vector Machine

## 3.1. Classification Decision Algorithm of SVM

The basic method of classification decision is to define decision rules on the basis of sample training set and minimize the false recognition rate or loss caused by such rules to the mode object. The machine for classification decision is classifier discussed here. A complete classifier system is composed of two parts: learning and implementation. Learning means the design of classifier with certain number of samples, *i.e.* training sample set; implementation means using designed classifier to make classification decision on samples waiting for recognition (testing sample set).

**Table 1. MPEG-7 Visual Descriptor Extraction**

Descriptor	Brief introduction	Dimension
Dominant color descriptor (DCD)	describes the dominant color information	8
Color layout descriptor (CLD)	reflects the image color distribution	13
Color structure descriptor (CSD)	reflects the local color image structure feature	256
Group frame / map color descriptor (GoP)	reflect the overall color feature of video	256
Parallel structure descriptor (HTD)	provides a quantitative description of texture energy and deviation	67
Edge histogram descriptor (EHD)	to describe the distribution of texture information space	90
Regional shape descriptor (RSD)	depicting regional objects based on the shape of	37
Block color descriptor (BCD)	reflect the different regions of color distribution	478
Movement intensity descriptors (MID)	reflect the movement intensity of video	4

To use SVM for classification decision, it has the following basic steps:

(1) Feature normalization

Due to different features with diversified physical significances, the range of values often differs greatly. This may lead to smaller value of one feature, which has little or bigger effect on the training process, leading to the feature dominating classification results. To eliminate the effect, it's necessity to make normalized treatment of each sample's various characteristic components.

(2) Choose kernel function for the nonlinear transformation and penalty factor  $c$  for punishment of errors

The most important parameter of SVM model is kernel function. Choosing what kernel function means mapping training sample to what kind of space for linear division. It's a pending problem that whether the selection of kernel function can get correct recognition result and how to guarantee theoretically to choose the best kernel function. We chose the well-known best RBF kernel as kernel function of support vector machine.

Penalty factor  $c$  is important controlling parameter of SVM classifier. It relates to regularized parameters. Its selection relies on the noise quantity in data.  $c=\infty$  implies that training samples must be classified accurately, but its generalization will be affected. Hence, it's necessary to consider as per actual situation. We choose bigger  $c$  to ensure that training samples can be classified accurately; then, diminish the value of  $c$ ; with cross validation method, we seek sufficiently little  $c$  in order to get classification face with favorable generalized ability.

The cross validation is a technology which trains learning models under the condition of guaranteeing good generalization ability of learning model with existing data. It's frequently used to solve machine learning problems. The common approach is  $n$ -dimensional cross validation, *i.e.* separate randomly data into non-intersecting  $n$  portions and choose a group of model parameters, from which choose  $n-1$  portions as training sample to create a training model; then, use the left data for testing. For each portion of data, the process repeats to achieve  $n$  different accuracy rates. If some accuracy rates can't meet requirements, it needs to change model parameters to repeat the above procedure till

n accuracy rates suffices the condition. With the method, a set of better model parameters can be obtained to make learning model with better generalized capability.

(3) Form quadratic optimization problem; use optimization method to solve the problem to get SVM classifier;

The algorithm is finally transformed into a constrained optimization problem for the two time. In theory, it will be the global best. The local extreme value problem cannot be avoided in the neural network method is solved.

(4) Utilize multi-classification SVM classifiers got after the training to predict unknown video types.

When various features of unknown videos are obtained, the SVM classifier model got in the above steps is applied for pattern matching of features to forecast the type of predictable videos, *i.e.* classify unknown videos.

Since video content classification is a complicated multi-class discriminant problem, and SVM itself is a method for two-kind problem, the construction of multi-class classifier can be implemented through combination of multiple binary sub-classifiers when multi-class problem is dealt with. So far, the common algorithms to construct multi-class SVM classifier have three kinds: 1-r (one-against-rest) method, 1-1 (one-against-one) method and directed acyclic graph (DAG) method. Here we adopt 1-1 method to build multi-class SVM classifier; and on that basis we propose secondary prediction mechanism to increase the classification accuracy rate of 1-1 method.

### 3.2. Secondary Prediction Mechanism Based on 1-1 Method

Because of the support vector machine is a kind of two value classifier. Therefore, when multi classification is performed, multiple support vector machines are needed to construct a multi valued classifier. The common construction methods are 1-r, 1-1 and DAG methods. 1-1 method is one of the most commonly used methods.

In 1-1 method. Regarding n kinds of samples, the method each time selects two kinds of samples to constitute SVM of all possible two kinds. It requires totally  $n(n-1)/2$  SVMs. Next, train and predict each SVM to get its binary classification result. Finally, train and predict each SVM to have respective binary classification result; lastly, cast a vote of each category as per such results, the class with the most votes as ultimate classification result. 1-1 method has feature of high accuracy of classification; however indiscernible regions exist when the vote equals.

The specific procedure of the prediction mechanism involves:

- (1) Construct SVM of any possible classes, which is totally  $n(n-1)/2$ ;
- (2) Put any two kinds of sample in relevant SVM for training;
- (3) For prediction, put testing sample in all constructed SVM and finally  $n(n-1)/2$  results reached;
- (4) Sum up prediction results and make vote to check if there's two classes with the highest and equal votes;
- (5) If not, the class with the most votes as the final prediction result;
- (6) If yes, perform second training of SVM constructed in accordance to descriptors chosen by the two kinds of classes;
- (7) In the new built SVMs, carry out second prediction and use the obtained result as ultimate predictive result.

The specific descriptor every two kinds of video selection, it is shown in Table2. Here we choose specific descriptors of every two kinds of videos. Through comparing the classifying accuracy of single descriptor for different classes, we find the selection of specific descriptors can discriminate better two types of particular videos, so as to improve the classifying accuracy of second prediction.

**Table 2. According to the Specific Two Categories Selected Descriptors**

Descriptor	Variety	Education	Automobile	Life
Technology	CSD,HTD,EHD,RSD	CSD,GoP,RSD,CCD	CSD,RSD,MID	CSD,RSD,MID
Life	GoP,HTD,EHD,MID	GoP,CSD,MID	CSD,GoP,MID	(-)
Automobile	GoP,HTD,EHD,MID	GoP,CSD,MID	(-)	(-)
Education	GoP,HTD,EHD	(-)	(-)	(-)

Since there're five kinds of videos, in the second prediction, there would be 10 kinds of possible combination of classes. In actual practice, we train SVM models of them all at one time. When it requires for re-prediction, we can pick up proper SVM for prediction, with no need of re-training, which saves the time to be consumed by initiating quadratic prediction mechanism.

## 4.Experiment Design and Discussion

### 4.1. Experimental Environments

This experiment will be divided into 5 categories, variety, education, automobile, life and technology. Five types of videos are downloaded from the Internet MPEG-4 video, ranged from 320 \* 240 to 640 \* 480 resolution, frame rate is 25 FPS or 30fps, the frame rate of the cartoon to 15fps.The total length of the five categories of video more than 8 hours, of which 1 hours 40 hours cartoon, advertising 1 hours. Music 1 hours 30 minutes, 1 hours 49 minutes, 1 hours 20 minutes

This experiment in the training and testing phase of the total number of samples is 5658, of which 2500 cartoon sample, ad sample is 1831, music sample is 1800, news sample is 1690, sport sample is 2588. The training sample and test sample accounted for half of the total sample.

The tool and software used in the experiment is ffmpeg [13], MPEG-7 eXperimentation Model [14], MATLAB [15], libsvm [16],. Before training, it's required to make shot segmentation and frame extraction of video segments; then making feature extraction of each shot and fetched video frames for normalized processing to regard the feature vector as the video's feature data; then, divide training data to five different training datasets.

The multi-classification SVM algorithm is implemented based on libsvm. SVM type is C-SVC. The chosen kernel function is radius basis function RBF, in the expression like:

$$K(x, x') = \exp(-\gamma \|x - x'\|^2), \quad \gamma > 0 \quad (1)$$

When using 1-1 method for classification, we create ten binary classifiers; then, use SVM training data to form five training datasets according to different types to make feature data of each set the same. Finally, each training dataset is used for cross-comparison training to acquire the best parameters C and  $\gamma$ , finishing the construction of each SVM.

### 4.2. Comparison of Classifying Accuracy for Single Descriptor

We compared in this part the accuracy of single descriptor. Such accuracy, to a certain degree, suggests the contribution of each descriptor to the overall classification rate of videos and also the effectiveness of them in distinguishing particular kind of video. From the experimental results, we extract some effective and necessary descriptors for feature description in the second prediction. For the experiment here we chose 1-1 SVM.

Among color descriptors, GoP's overall classifying accuracy rate is the highest, reaching 81.8%. It implies that the first choice of GoP descriptor can differentiate farthest

five kinds of videos. The other color descriptors made relative contributions, of which CSD got the highest classifying accuracy rate of scientific videos, *i.e.* 91.6%; yet, CCD realized good rate of scientific videos, which may be because the block color information of scientific videos is fixed. Hence, we used CSD in descriptors for the re-prediction of scientific videos and GoP for other videos. It is shown in Table3.

**Table 3. Comparison of Classifying Accuracy for Single Descriptor**

Descriptor	DCD	CLD	CSD	GoP	HTD	EHD	RSD	CCD	MID
Variety	43.0	63.5	75.8	75.0	79.8	84.0	57.8	69.0	29.9
Education	36.2	61.3	60.5	73.6	60.2	67.2	52.0	57.3	24.2
Automobile	74.4	80.3	80.3	88.6	70.5	76.0	67.8	71.0	78.7
Life	70.8	81.5	80.5	85.3	77.8	81.5	58.4	71.0	81.5
Technology	81.5	82.7	90.6	85.3	82.5	81.9	72.0	81.9	52.3
Totality	61.6	69.3	78.9	81.8	74.1	78.4	62.6	71.3	53.8

In the texture descriptors, HTD and EHD have higher classifying accuracy rate of variety videos, indicating that those videos are easily differentiated in texture. Since variety videos are fabricated manually, its texture looks smoother than common un-artificial videos. So good effects can be achieved to discern variety videos by texture. In descriptors for the second prediction of those videos, we adopt HTD and EHD.

In shape descriptors, RSD has the highest classifying accuracy rate of scientific videos, probably because of relatively fixed location and shape of scientific video ground and stand, resulting in the easy distinction of such videos by RSD. Thus we adopt RSD in the re-prediction of descriptors for scientific videos.

In movement descriptors, MID can better discriminate car and life videos, whose accuracy rate is respectively 77.7% and 82.7%. MID reflects the intensity of video motions. Cars especially concerts often have complicated backgrounds like flashing lights and moving settings, which leads to rather big MID value. On the other way around, life videos mainly refer to still image, even though there's slow-moving image or constrained in local range, MID value of life videos is small. MID can discriminate effectively two kinds of videos. In descriptors for the second prediction, we use MID for car and life videos.

Figure 1 shows a single descriptor respectively on the five video overall classification accuracy rate of contribution. It can see a single descriptor in video classification accuracy is not ideal. The accuracy of the classification is 53.8%, the highest classification accuracy is 81.8%.

This shows that not a single performance descriptors in video classification meet the need. Therefore, a large number of fusion MPEG-7 descriptors and optimized combination is a key problem of automatic video classification.

In conclusion, single descriptor can discriminate well one or two types of videos; however, it's not enough to make the whole difference of five videos. To get favorable classifying effect of five videos, it's required to fuse all these descriptors.

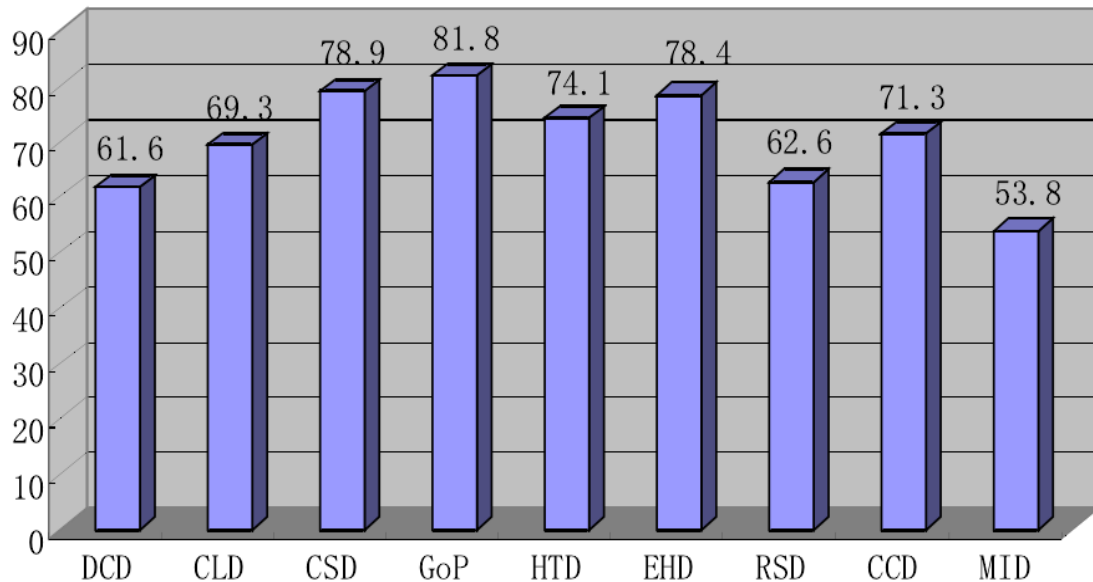
#### 4.3. Accuracy Comparison of Different Multi classification Algorithms

In this paper, we use Python language to achieve the above described several SVM multi classification algorithm and compare the experiment. It is shown in Table4.

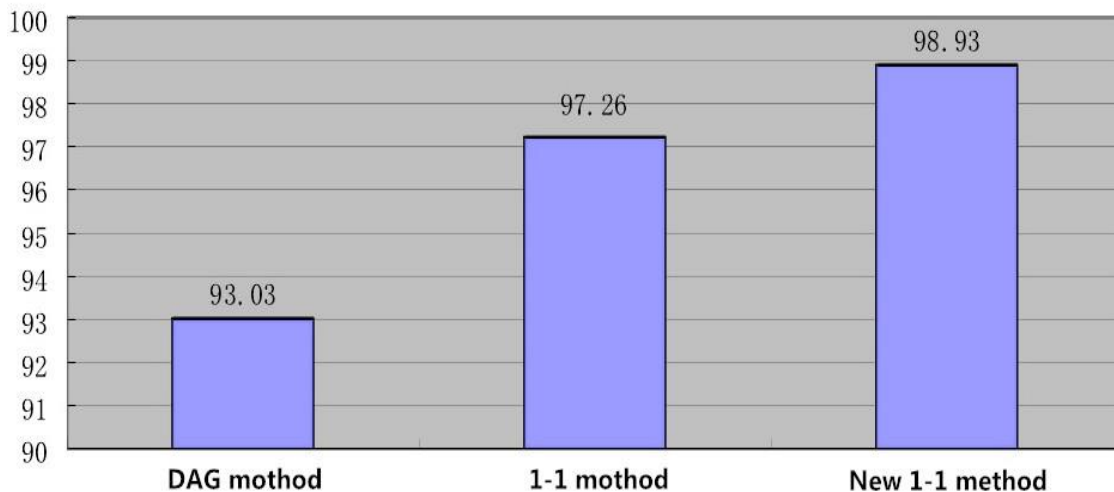
Because of the low accuracy of 1-r method, it is not common in practice, so it is not to compare the 1-r method.

**Table 4. Accuracy Comparison of Different Multi Classification Algorithms**

Multi classification algorithm	DAG method	1-1 method	New 1-1 method
Error instance / total instance	258/2879	205/2879	40/2879
Accuracy rate (%)	93.03	97.26	98.93



**Figure 1. Overall Video Classification Accuracy on Single Descriptors**



**Figure 2. Accuracy Comparison on Different Multi-class Classification Algorithms**

It is shown in Table 4. There are 2879 test cases, the DAG method is 258, and the 1-1 method is 205. New 1-1 method, the error prediction of 1-1 method is 40, and the classification accuracy is 98.93%. The classification accuracy rate of 5.90% and 1.67% respectively, compared with new 1-1 method. Thus, the classification accuracy of SVM method is improved new 1-1 method is enabled.

After analyzing the data, it is found that the 1-1 method appears to have 150 instances in the 2879 instances. 21 of them are not the same as one of the two categories of video,

so it is bound to be classified;

The remaining 129 instances are correctly 119 classified by new 1-1 method .And the original 1-1SVM by random selection of correctly classified 55. Therefore, after new 1-1 method is enabled, the 1-1SVM is less than 64 of the original algorithm, and the accuracy of video classification is improved.

## 5. Conclusion

In the paper, the problem of automatic classification for different types of video is studied, the automatic classification model of video content MPEG-7 visual descriptors and support vector machine are proposed, and gives the system frame design and the concrete process design. In this chapter, the design method of SVM multi classifier is improved.

The new 1-1 method based on 1-1 method, in order to improve the classification accuracy of the system. The classification experiments of the common five kinds (variety, education, automobile, life and technology) of video are compared with the experimental results of the same type. The effectiveness and feasibility of the proposed algorithm are proved.

## Acknowledgement

This work was supported by Research on Education and Teaching Reform of Jilin Vocational and Adult Education.

## References

- [1] Q. Dan, "Study on the algorithm for automatic classification of video content and multi feature combination based on SVM", Shanghai Jiao Tong University, (2009).
- [2] R. Zhou, "Based on support vector machine based video watermarking algorithm research", Wuhan University of Technology.
- [3] W. Shuang, "Research on video object extraction technology based on support vector machine", Jilin University, (2013).
- [4] D. Dan, "Based on a on a support vector machine automatic video classification algorithm", Computer applications and software, vol. 1, (2010), pp.3-5.
- [5] F. Wen, D. Q. Jeong and J. Yan, "Moving object classification algorithm based on CKPCA-HOG and support vector machine", Journal of Harbin Engineering University, vol. 5, (2011), pp. 643-649.
- [6] L. Kai, "Research on tracking algorithm for ranking support vector machine based on target", Xiamen University, (2014).
- [7] M. Xiangyi, "Tracking and recognition algorithm for moving target", Jilin University, (2011).
- [8] L. Lili, "Study on the method of moving target classification based on shape features", Hunan University, (2006).
- [9] Z. Haiyong, "Research on the identification of human behavior based on video stream", Xi'an Electronic and Science University, (2011).
- [10] K. Jia, "Research on the method of video event detection and analysis based on semantic", Jiangsu University, (2013).
- [11] L. Wen, "Based on clustering algorithm and support vector machine algorithm of text classification algorithm research", Jiangsu University of Science and Technology, (2012).
- [12] S. Theodoridis and L. J. Jing Jiao, "Pattern recognition. Third edition. Beijing", publishing house of electronics industry, vol. 12, (2006), pp. 296-298.
- [13] FFmpeg Documentation. <http://www.ffmpeg.org/ffmpeg-doc.html>.
- [14] MPEG-7 Reference Software: eXperimentation Model(XM). [http://www.lis.ei.tum.de/research/bv/topics/mmdb/e\\_mpeg7.html](http://www.lis.ei.tum.de/research/bv/topics/mmdb/e_mpeg7.html)
- [15] Y. Shuying, "Pattern recognition and Intelligent Computing: Matlab technology", Beijing, electronics industry press, (2008).
- [16] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machine", <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.



## Authors



**Chao Jiang**, He received his B.S degree from Jilin Animation Institute. He is an engineer in Jilin Communications Polytechnic. His research interests include data recovery, cloud computing.



**Shuguang Wang**, He received his M.S degree from Changchun University of Technology. He is a senior engineer in Jilin Communications Polytechnic. His research interests include data recovery, cloud computing.

