ISSN : 0974 - 7435 Volume 8 Issue 4

BioTechnology
An Indian Journal

FULL PAPER

BTAIJ, 8(4), 2013 [502-511]

A novel method of shot boundary detection based on SVM of optimal parameters

XiaoyuLv, Xuemei Sun*
School of Computer Science and Software Engineering, Tianjin Polytechnic University Tianjin, 300387, (CHINA)
E-mail : seesea_sun@163.com

ABSTRACT

Shot boundary detection (SBD) is the first and prerequisite step for content-based video indexing and retrieval. A novel shot boundary detection algorithm based on modified SVM model which is improved by simulated annealing algorithm and culture algorithm is proposed. The formation of classifying model adopts culture algorithm for its strong evolutional ability and simulated annealing algorithm embedded into culture framework. Simulated annealing algorithm is employed to carry on the local search process in which individual solution changes towards neighborhood, addressing the optimal solution to belief space of culture algorithm. After updating the belief space, it directs the individual evolution, further improves the SVM model parameters. The feature indicators of brightness means, brightness variances, edge changing edges, block histogram and DC coefficients are extracted from pixel domain and compressed domain, and then the sliding window is used to organize the features into feature vectors. Finally the video frames are classified into the cuts, gradual changes, and non-changes by SVM classifying model. The experiments with sample video data demonstrate the algorithm is effective and robust.

KEYWORDS

Support vector machine;
Shot boundary detection;
Culture algorithm;
Simulated annealing algorithm.

INTRODUCTION

Shot boundary detection is the significant prerequisite and key technology for content-based video retrieval. Video temporal segmentation, that is splitting a video into a few video clips in the time dimension. Each video clip consists of several images shot by camera alone, and it is called as a shot. In the field of video retrieval, the key technology is shot boundary detection which includes location of shot boundaries. Namely, shot boundary detection is ascertaining the starting frames and the end frames, and distinguishing the shot transformation types which are also called the transfer mode of shots. The transformation types are mainly including cut changes and gradual changes, and the gradual changes consists of wipes, dissolves, and other types[1].

The common strategy of traditional shot boundary detection method is extracting some certain data from the underlying visual information or video coding, then forming the video feature values. On the basis of the video features extracted, the shot boundaries can be detected according to the differences of characteristic value between two frames. Feature extraction is divided into two steps: the most basic step is to extract the image features of each frame directly; the other step is that new features are formed according to not only the features belonging to the current frame, also the features between the current frame and the frames with some certain distances. Up to now, threshold selection problem is still involved in most shot boundary detection algorithms, and it directly affects the accuracy of detection results. However, the threshold is often derived from experiences or experiments, and it is not with good versatility. In recent years, many scholars have studied the machine learning algorithms to detect the shot boundaries. Typical methods are including the methods based on unsupervised learning, the decision tree, neural network and support vector machine (SVM) and so on[3,4].

Support vector machine (SVM) is proposed as a machine learning method based on statistical learning theory, it solves some practical problems of small sample, non-linear and high dimension as a kind of training method based on structural risk minimization. However, the studying and forecasting capacities of SVM model are up to the parameters selections. At present, the most traditional selection methods of parameters used are mainly through the cross validation, error bounds method and statistical methods to obtain the parameter values. Whereas, the traditional algorithm of parameters optimization combined by web search and cross validation requires a large amount of calculations, especially when the training sample set is quite large, the search process is very time consuming. By contrast, intelligent evolutionary algorithms such as genetic algorithm, culture algorithm are gradually adopted for the SVM parameters optimization. The algorithms above often have better global searching abilities, stronger parallelism and higher efficiency. Consequently, many scholars make use of the intelligent evolutionary algorithms to optimize the parameters of support vector machine[10,11].

Reference[4] focuses on the compressed domain to solve the shot boundary detection problem. The proposed method brings in sliding window algorithm to handle the extracted features including frame types and motion vectors. And then the trained SVM model is used to categorize the video frames into cuts, gradual changes and non-changes. In reference[10] a method that optimises the parameters of least squares support vector machines by using Genetic Algorithm is presented, this method greatly improves the efficiency of SVM's parameters selection, and with the parameters selected, the classification result for the testing samples is the optimum. It avoids the disadvantage of manually specifying the parameters, and also scales down the optimisation time. Reference[11] studies the threshold detection, and presents an adaptive threshold adjusting detection method. It makes use of a combination of brightness differences and mean of values as the feature indicators to detect the cut changes and also uses color differences feature to capture gradual shot changes. The thresholds are generated through the variety of the selected values by itself. Reference[9] proposes a two-stage shot boundary detection method. Based on k-nearest neighbor algorithm and SVM, it classify the video frames into cuts and non-cuts firstly, and then wavelet denoising algorithm is used to distinguish gradual shot changes from non-changes. Reference[8] presents a unified model for detecting different types of video shot transitions. The frame transition parameters and frame estimation errors based on global and local features are used for boundary detection and classification. In the proposed algorithm, the frame estimation scheme are formed with making use of the previous and next frames information, and the local features consist of scatter matrix of edge strength and motion matrix. Furthermore, the multilayer perceptron network is used to classify the video frames into cut changes, gradual changes and non-changes. Reference[12] proposes an efficient approach which is capable to detect various shot boundaries simultaneously in a unified way. The proposed algorithm firstly detects general shot boundaries based on the idea of Fisher criterion, and then classifies them into the cuts and gradual transitions by an SVM classifier. Further computation is performed for the GT shot boundaries to expand rough boundary locations between two frames into the transition interval consisting of all the transitional frames. Finally, a post-processing step is conducted to merge some overlapped transitions. Reference[5] proposes a novel method based on the combination of adaptive threshold and Fourier fitting to detect shot transitions. HSV color histogram on each

FULL PAPER

frame is accumulated to measure a similarity sequence of videos, then the adaptive threshold is used to detect hard cut transitions. Candidate transitions are detected by searching segments with the traits of gradual transition pattern and gradual transition boundaries of different types are collected to train a set of standard templates. The obtained set can be used to judge whether a candidate is a definite gradual transition boundary or other change types.

The classification model of SVM is applied in this paper. The video frames are regarded as the formation of three types of frames which are the abrupt change frames,the gradual change frames and the non-change frames. And the features of pixel domain including mean brightness, brightness variance, marginal rate of change, block histogram and the features of compressed domain, for example DC coefficients and motion vectors, are adopted in this paper. According to the selected sliding window, a multi-dimensional feature vector formed by using the frame features within the window is regard as the feature vector of center frame of window, which is used to train SVM model. Parameters of SVM are optimized by culture-simulated annealing algorithm, the cultural algorithm is used for global optimization, while simulated annealing algorithm is used for local optimization. Video sequence frames is classified by the optimal SVM classifying model obtained by the above steps, and the three types of video frames are detected once, which realizes shot segmentations.

SUPPORT VECTOR MACHINE AND CULTURE-SIMULATED ANNEALING ALGORITHM

Basic principle of support vector machine

Support vector machine is a machine learning method based on the statistical theory. Its principle is introduced as follows:

Given a set of training data $(x_i, y_i)_{1\le i\le N}$,where $\mathbf{x_i \in R^n, y_i \in \{-1,1\}}$ is class code of x_i. And the training data can be partitioned by a hyper plane:

$$\mathbf{y_i(w * x_i + b) > 0, i = 1,..., N} \quad (1)$$

Then the training data is linearly separable. At present, we can adjust the vector w and scalar b which make the below formula set up:

$$\mathbf{y_i(w * x_i + b) > 0, i = 1,..., N} \quad (2)$$

Like this, the distance between the point closest to the hyper plane and hyper plane is $1/\|w\|$. Then the formula(1) is changed to the formula:

$$\mathbf{y_i(w * x_i + b) \ge 1} \quad (3)$$

If the set of vectors satisfying the formula(3) is partitioned by the hyper plane infallibly, and the distance between the vector closest to the hyper plane and hyper plane, then the hyper plane is called as optimal hyper plane. Because of the distance between the point closest to the hyper plane and hyper plane is $1/\|w\|$,searching for hyper plane is equal to minimize $\|w\|$ under the condition of formula (3).

Because $\|\mathbf{w}\|^2$ is convex, $\|w\|$ is minimized by Lagrange multiplier under linear constraints (3). Remember that $a = (a_1, a_2,..., a_N)$ is N nonnegative Lagrange multiplier related to constraint (3), the optimization problem is fallen into to maximize the below formula:

$$\mathbf{W(a) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j x_i * x_j} \quad (4)$$

Constraint condition is: $\alpha_i \ge 0, \sum_{i=1}^{N} y_i\alpha_i = 0$.

Formula (4) can be obtained by standard quadratic programming method. Once the solution $\alpha^0 = (\alpha_1^0,...,\alpha_N^0)$ of formula (4) is solved, the optimal hyper plane satisfies the following formula::

$$\mathbf{w_0 = \sum_{i=1}^{N}\alpha_i^0 y_i x_i} \quad (5)$$

These points satisfying $\alpha_i^0 > 0$ and formula(3)are referred as support vectors. The judge function of hyper plane can be written to:

$$\mathbf{f(x) = sgn(\sum_{i=1}^{N}\alpha_i^0 y_i x_i * x + b_0)} \quad (6)$$

If the data is linear and impartible, we will introduce slack variable $(\varepsilon_1,..., \varepsilon_N)$ and $\varepsilon_i \ge 0$, that is:

$$\mathbf{y_i(w * x_i + b) \ge 1 - \varepsilon_i, i = 1,..., N} \quad (7)$$

Namely, this allows some error classification points. the problem of solving optimal hyper plane is boiled down to solve the following problem:

$$\min_{\varepsilon,w,b} \frac{1}{2} w * w + C\sum_{i=1}^{N} \varepsilon_i \quad (8)$$

The constraint condition is formula(7), and $\varepsilon_i \geq 0$, C is a constant.

The performance of SVM depends on multiple parameters to a great degree. The traditional selection methods of parameters need a large amount of calculation, especially when encountering a larger training sample set, the search process will become more time-consuming. In recent years, some research with the method of intelligent algorithms are adopted to optimize parameters, such as genetic algorithm, culture algorithm and so on.

The description of culture algorithm

As a strategy of double evolution method, cultural algorithm provides evolutionary framework on the macro level[6]. Cultural algorithm has two spaces which are belief space and population space. Belief space is on behalf of the social culture and is used to guarantee evolution direction in the field of macro, and the population space is the collection of human individuals. The mutual relationships are shown in Figure 1.

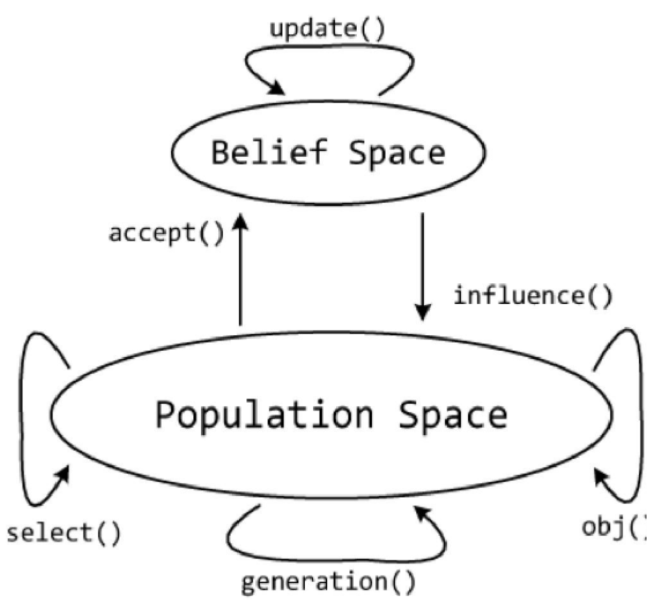

Figure 1 : The framework of culture algorithm

Belief space and population space interact with each other through the accept function and influence function, the accept function exerts population's influence on culture. After accepting the influence, belief space will be renewed by the updating function, which represents the development of culture. Influence function places the guiding function of belief space for the population, individual species complete evolution guided by belief space. The function generate represents the evolution process of each generation of population space. When each generation implements the evolution, population space adopts the function obj() to compute the fitness values of individuals and invokes the function select to choose a certain number of individuals, and these individuals' specific circumstances will be addressed to the function accept, which represents the influences of individual experience, knowledge and other information for social culture.

Principle of simulated annealing algorithm

Simulated annealing (SA) is a kind of heuristic random search algorithm presented by Metropolis, it is widely used to solve a variety of combinatorial optimization problems through the simulation of thermodynamic process of metal cooling, including a lot of NP - complete problems. Additionally, its core principle is to simulate the cooling scheduling theory during the process of making some material, which is the relationship of temperature and the state of the object. Statistical mechanics research shows that, the probability of molecules remaining in the state r satisfies the boltzmann distribution under the temperature T:

$$P_r\{\overline{E} = E(r)\} = \frac{1}{Z(T)} \exp\left(-\frac{E(r)}{k_B T}\right) \quad (9)$$

In the formula 9, $\overline{E}$ is the random variable of molecular energy, $E(r)$ is the energy of a molecule in state r, T is temperature, k_b is for the boltzmann constant, $Z(T)$ is as the normalization factor. When the initial temperature T_0 is enough high, the probability of staying on each state for molecular is close. When the temperature drops to zero, all molecules will stay in the lowest energy state with probability 1, while if the molecules in the lowest energy state r' has energy 0, when $T \rightarrow 0$, there is:

$$P_r\{\overline{E} = E(r)\} = \begin{cases} 1, & r = r' \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

Thereby the algorithm will be convergent to optimal solution in theory.

FULL PAPER

SHOT BOUNDARY DETECTION METHOD BASED ON SVM OF PARAMETER OPTIMIZATION

The construction and training of the support vector machine (SVM)

Radial Basis Function is applied in this paper, its express is:

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\sigma |\mathbf{x}_i - \mathbf{x}_j|^2), \sigma > 0 \quad (11)$$

$k*(k-1)/2$ classifiers are constructed by the one-against-one method for classifying multi-class. And k is the number of classificatory, there are three types shear frame, gradient frame and non-shot transformation frame. This method is first used for SVM in literature[17]. Every classifier trains the data from different classificatory, the training data from the classificatory i and j can solve the following binary classification problem:

$$\min \frac{1}{2}(\mathbf{w}^{ij})^T \mathbf{w}^{ij} + C(\sum_t (\xi^{ij})_t) \quad (12)$$

Its constraint condition is:

$$(\mathbf{w}^{ij})^T * \Phi(\mathbf{x}_t) + \mathbf{b}^{ij} \le 1 - \xi_t^{ij} \quad (13)$$

$$(\mathbf{w}^{ij})^T * \Phi(\mathbf{x}_t) + \mathbf{b}^{ij} \ge -1 + \xi_t^{ij} \quad (14)$$

Voting strategy is used during classifying, namely an output of every binary classifier can be called as a vote, and the data point decides which classificatory it belongs to according to the max votes. If most classifiers regard the data point belong to i classificatory, then the data point will belong to i classificatory.

Optimization of SVM parameters by the simulated annealing - culture algorithm

We use simulated annealing - culture algorithm to optimize two parameters of RBF in SVM, When the parameters of SVM are optimized, global optimization ability of culture algorithm is applied in the macroscopic field, on the other hand, local search is completed by using simulated annealing algorithm in the micro field, and simulated annealing algorithm is used to replace the function generate () in the framework of culture algorithm, which is to complete the evolution of population of each generation. As we can see in Figure 2, the flow diagram shows the forming process of SVM model. This strategy can reduce the search time effectively, get high efficiency and avoid falling into local optimal solution, consequently realize the optimization of SVM parameters preferably.

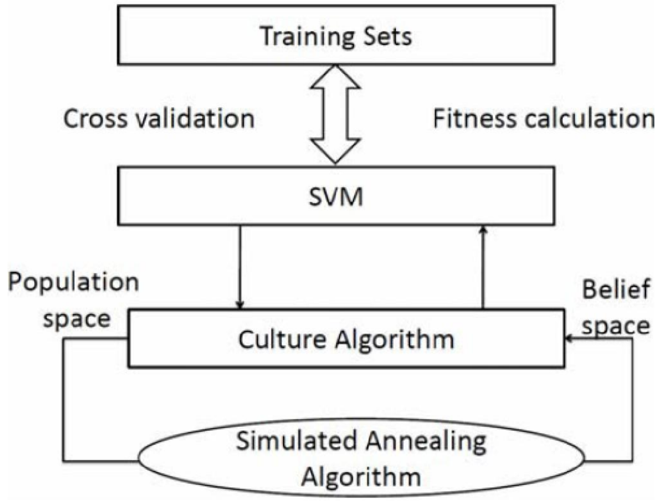

Figure 2 : Cross Validation for SVM Training

1) The construction of the population space

The code of an individual X is a n-dimensional binary vector initially, each dimension of the vector is initialized to 0 or 1 randomly, which represents the parameter. The fitness of individual X is defined as shots detection accuracy of support vector machine.

Population space is $PS = (X_1, X_2, ..., X_n)$, n is for population quantity.

2) The accept function

The function accept() is used to choose the individuals that can directly affect the knowledge and experience of current belief space. It can be defined as follows:

$$\text{Accept}() = \beta \cdot P + \lfloor (\beta \cdot P)/t \rfloor \quad (15)$$

In the formula (15), P is population size, t is for the evolution iteration, β is for a given proportion constant.

3) The update function

Updating the knowledge of belief space is completed by the function update(), it receives the excellent individuals delivered by function accept(), which is used to update situation and specification knowledge.

• update of situation knowledge For individual x_i, the fitness $f(x_i)$ is prescribed by (37). First definition is as follow:

$$\Delta f = f(\bar{x}_{best}^{t}) - f(\bar{s}^{t}), \quad (16)$$

$$p = \min\{1, \exp(-\Delta f T_k)\}, \quad (17)$$

$$r = \mathrm{random}(0,1). \quad (18)$$

If $\bar{x}_{best}^{t}$ is the optimal individual of t'th generation, $\bar{s}^{t}$ is the current optimal individual. Then the situation knowledge S is updated according to the following formula:

$$\bar{s}^{t} = \begin{cases} \bar{x}_{best}^{t} & \Delta f \geq 0 \text{ and } r < p \\ \bar{s}^{t}, & \Delta f < 0 \end{cases} \quad (19)$$

If contemporary individual is superior to the individual in situation knowledge, then replace the corresponding individual; On the other hand, replace according to a certain probability.

• the update of constraint knowledge follows:

$$l_i^{t+1} = \begin{cases} x_{j,i}, & x_{j,i} \leq l_i^t \text{ or } f(\bar{x}_j) < L_i^t \\ l_i^t, & \text{otherwise} \end{cases} \quad (20)$$

$$u_i^{t+1} = \begin{cases} x_{k,i}, & x_{k,i} \geq u_i^t \text{ or } f(\bar{x}_k) < U_i^t \\ u_i^t, & \text{otherwise} \end{cases} \quad (21)$$

$$L_i^{t+1} = \begin{cases} f(\bar{x}_j), & X_{j,i} \leq l_i^t \text{ or } f(\bar{x}_j) < L_i^t \\ L_i^t, & \text{otherwise} \end{cases} \quad (22)$$

$$U_i^{t+1} = \begin{cases} f(\bar{x}_k), & X_{k,i} \geq u_i^t \text{ or } f(\bar{x}_k) < U_i^t \\ U_i^t, & \text{otherwise} \end{cases} \quad (23)$$

Among them, the subscript i represents the i'th constraint variable, j and k represent the j'th and k'th individuals separately, t is iteration. Then $x_{j,i}^{t}$ is j'th constraint variable of i'th individual of t'th iteration.

• the modification of step size is conducted according to the following formula:

$$\delta = \sqrt{\sum_{j=1}^{n}(x_j - x_j^*)^2} \quad (24)$$

$$\delta_i^{t+1} = \delta_i^t + N(0,1) \cdot \max(|\delta_j^t - p|, |q - \delta_j^t|) \quad (25)$$

4) The influence function

This paper adopts specification knowledge to guide the evolution of the offsprings, i.e:

$$x_{j,i}^{t+1} = \begin{cases} x_{j,i}^t + |\delta_j \cdot N(0,1)| & x_{j,i}^t < l_i^t \\ x_{j,i}^t - |\delta_j \cdot N(0,1)| & x_{j,i}^t > u_i^t \\ x_{j,i}^t + \lambda \cdot \delta_j \cdot N(0,1) & \text{otherwise} \end{cases} \quad (26)$$

Where N(0,1) is the random number between 0 and 1, and λ is a constant

5) The procedure of culture-simulated annealing algorithm

1. To carry on the initializations of population space and belief space, t = 1.
   - Initialize population space:
   Initialize each individual x_i: the fitness function of an individual is accuracy detection based on SVM, fitness f_i is calculated according to the formula (37).
   - Initialize belief space:
   Set
   $S = \{\varnothing\}, N_i = < I_i, -\infty, \infty >, h_i = [0, min[size(I_i)]/2]$
   temperature T of initialization is a const t_0.
2. Choose good individuals according to formula (37), $t = t + 1$.
3. Update belief space, update specification knowledge according to the formulas (20)-(23).
4. Update step length, temperature $T = \eta \cdot T$, and η is a constant.
5. The evolution of the next generation in population space is carried on according to the influence function in formula (26).
6. To judge whether meet the end conditions, if yes, then output the results, else turn to the step 2.

Feature extraction

Video features of pixel and compression domain are used synthetically in this algorithm of this paper. Extract mean brightness, brightness variance, marginal rate of change and block histogram from pixel domain, and extract DC coefficients from compression domain respectively, and then compose the training and test data.

a) Brightness

The mean average luminance brightness and brightness variance curve is mainly focused on detecting the cut changes. As is shown in Figure 2, we can observe the drastic changes of video sequences??mean luminance curve under the abrupt shot. Extract the mean brightness of the image frames according to the following formula:

$$M_i = \sum_{(R,G,B)} \sum_{i=1}^{M} \sum_{j=1}^{N} p_{i,j} \tag{27}$$

The $p_{i,j}$ is an pixel brightness value at the point (i, j) in the image having $M * N$ pixels. We define a variable is equal to the difference between the average brightness value of frame before and after:

$$S_i = M_i - M_{i-1} \tag{28}$$

A threshold T usually be defined to compare with S as the measure of shot boundary detection in traditional method. For we adopt the technique of SVM so that S is just entered as a variable of the input vector and obviously its meaning is the luminance difference between the current frame and the former frame.

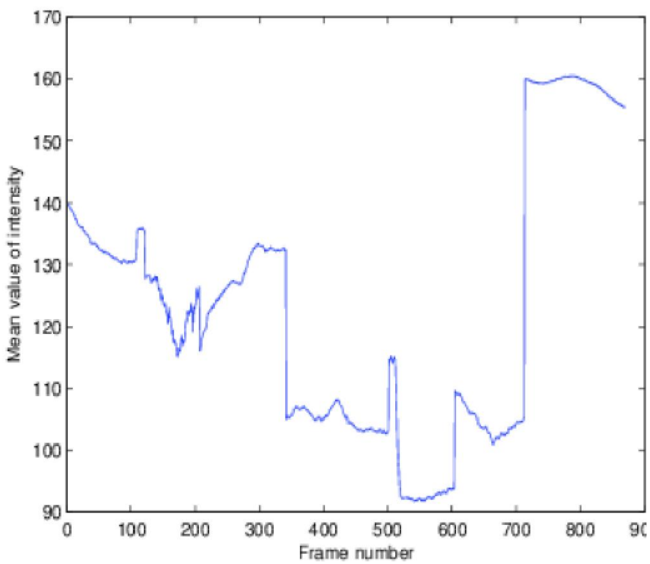

Figure 3 : Mean Luminance

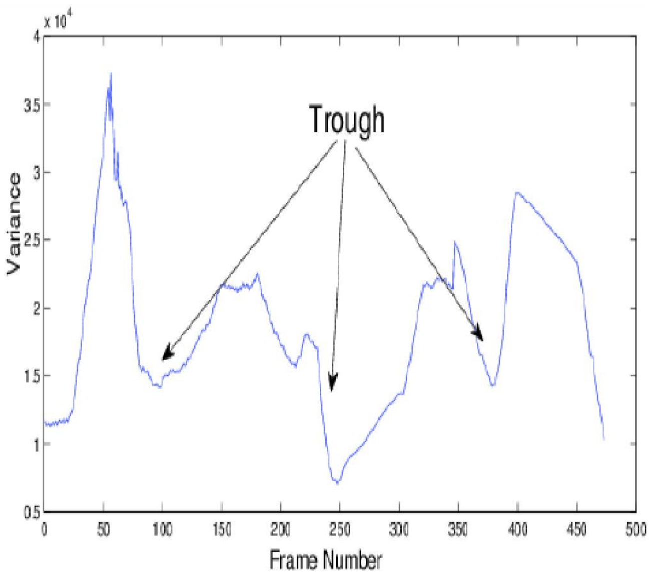

Figure 4 : Brightness Variance

The curve of the brightness variance is capable of distinguishing dissolve. In Figure 3 reference[7] analysis the brightness variance curve often shows a trough state during the process of dissolve. The process, during which the curve will present trends of increasing or decreasing with taking the trough as center, is usually called as dissolve when corresponding to the shot boundary detection. Therefore, we define the variance ratio between the current frame and the former frame.

$$V_i = \sum_{R,G,B} \sum_{j}^{M} \sum_{i}^{N} (\overline{X} - x_{i,j})^2 \tag{29}$$

b) Edge gradient rate:

Basic idea of edge features is : while the change of shots occurs, a new edge should be far away from the old edge, the same that the position old edge disappears should be far away from the position of new edge. First, edge graphs of E_k and E_{k+1} of before and after two video images k and k+1 will be extracted, the difference between two video images is calculated according to the below formula:

$$diff = \max(d_{in}, d_{out}) \tag{30}$$

In the formula, d_{in} is the proportion of entering into pixels (pixels are newly emerging and far away from the exiting edge), d_{out} shows the percentage out of pixels (pixels are newly disappearing and far away from the new edge). If p_1 is defined as the total number of edge pixels of E_{k+1}, which satisfy that the distances between the closest edge pixels and edge pixels in are more than r, PM is the total number of edge pixels in E_{k+1}; p_2 is defined as the total number of edge pixels of E_k, which satisfy that the distances between the closest edge pixels and edge pixels in are more than r, p_n is the total number of edge pixels in E_k, then there are as following:

$$d_{in} = \frac{p_1}{p_m} \tag{31}$$

$$d_{out} = \frac{p_2}{p_n} \tag{32}$$

c) Block histogram:

Consecutive frames of a shot contain similar global visual properties,which performance is the difference between histograms of two frames should less equal to the difference between histograms of two frames of shot boundary. Because the histogram is only representing the global distribution, this feature is not sensitive to local object motion. But when there is global movement within one shot, great changes will happen in histogram.

The basic principle is divided the color space into some discrete short intervals, and then computing the number of pixels falling into each short interval. A color space is divided into n intervals, $H_{k,i}$ is the number of pixels the k'th frame falling into the shot interval of i'th color. The difference between frames is expressed by the following formula:

$$\mathbf{D_{k,k+1}(I) = \sum_{i=1}^{n} | H_{k,i} - H_{k+1,i} |} \quad (33)$$

The calculation of histogram can also regard image block as unit, respectively every frame image is divided into 2^n pieces, straight variance between two frames is calculated for the corresponding blocks. Block division makes the algorithm more sensitive to shot shear.

d) DC coefficient:

Because human visual is more sensitive to image brightness than color, brightness block DC coefficient can be extracted from each macro block, and the change tendency of shots can be gotten by analyzing and calculating of luminance change. By MPEG coding standards, the brightness of a frame is conducted DCT transform, then DC coefficient is obtained through quantizing. Therefore luminance difference between frames has good consistency with the difference of DCT luminance coefficient. If D (x, y) is the pixel value of pixel point (x, y), D (u, v) is the DCT coefficient located at the position (u, v), Then it is easy to show difference of DCT luminance coefficient:

$$\mathbf{DCT = D(u,v,t) - D(u,v,t-1)} \quad (34)$$

In the formula, t is the time serial number, and refers to the frame number in video stream. DC luminance coefficient is the image luminance corresponding to the macro block after decoding, so the change of image luminance of ith frame is expressed by the difference D (i)of DC luminance coefficient between two continuous frames. When some shot change occurs between two consecutive frames, the difference of DC luminance coefficient has greater change. From this perspective, the histogram of difference curve of a video DC luminance coefficient contains a shot change information.

Constructing video Features

In the shot conversion of a video, the type of each frame is related to its front and rear frame. Therefore, a sliding window is established, the feature data at window middle position corresponding to the frame is combined by two parts: feature data of front and rear frame in window and owning feature data, set the width of the window is n, if there are m features, the total number of features of each frame after combining is m * n. Change the width of the window, you can change the total number of features of each frame, namely change the dimension of features each frame. Features of each frame described earlier are extracted from videos, after the normalized processing, set window width 5, the data of vectors are formed, for SVM training and testing.

EXPERIMENT

We use the sample video data set to assess the shot boundary detection algorithm proposed in this pa-

TABLE 1 : VIDEO CLIPS

Name	Total Frames	Cut	Gradual
Goodfellas	1916	20	14
Inception	2023	24	16
Forrest Gump	2996	29	28
Astro Boy	2960	12	9
Ninja Assassin	2045	12	13
Armored	2300	14	14
Defendor	2190	13	16
Brothers	1916	20	14
Leap Year	345	3	3
City of God	1400	15	9
Psycho	9893	120	76
The Matrix	3540	33	29
Citizen Kane	4613	48	40
Taxi Driver	2048	22	18
Fight Club	3924	42	24
Toy Story 3	2636	30	24
The Godfather	4573	45	40
Sunset Blvd.	1514	16	15

per, shown in TABLE 1. It is divided into three groups randomly. A set of video is regarded as the training sample of parameters optimization and the final classification, a set of video is test sample of parameters optimization, the rest is test sample of final classification. Evaluation index of the shot detection algorithm often uses two parameters, Recall and Precision, their definitions are as follows:

$$\textbf{Precision} = \frac{\textbf{Hit}}{\textbf{Hit} + \textbf{False}} \tag{35}$$

$$\textbf{Recal} = \frac{\textbf{Hit}}{\textbf{Hit} + \textbf{Miss}} \tag{36}$$

The higher the ratios of Recall and Precision are, the better the detection is. A good shot boundary detector should have high recall and precision ratio at the same time, F1 is used to comprehensively measure recall and precision ratio commonly, F1 is high only when precision and recall ratios are high. That is:

$$\textbf{F1} = \frac{2 \times \textbf{Precision} \times \textbf{Recall}}{\textbf{Precision} + \textbf{Recall}} \tag{37}$$

Shot boundaries of videos are detected by the SVM classifier after training according to the data, and the test results of the algorithm in this paper are obtained. And the detection effect of algorithm in this paper is compared with three algorithms proposed in[2], the results are as shown in TABLE 2 and 3.

TABLE 2 : CUT CHANGE

Runs	Recall	Precision	F1
Run-1	0.7748	0.5122	0.6178
Run-2	0.6121	0.8992	0.7278
Run-3	0.9332	0.9836	0.9577
Run-4	0.9487	0.9894	0.9781

TABLE 3 : GRADUAL CHANGE

Runs	Recall	Precision	F1
Run-1	0.3949	0.3924	0.3936
Run-2	0.7261	0.5000	0.5922
Run-3	0.7626	0.9330	0.8392
Run-4	0.8235	0.9576	0.8510

As is shown, compared with the above three methods, the performance of the algorithm in this paper is best. Run -1 uses the histogram difference to detect shot boundaries, for the video of more objects and camera movement, larger difference between frames will occur according to the operations of objects and camera movement. Run - 2 and Run - 3 also use SVM, but use traditional methods in the choose of parameters such as cross validation and optimization. In this paper, the parameters of the optimization method (Runs - 4) have achieved quite a good detection effect.

CONCLUSION

A new shot boundary detection method based on parameters optimization of SVM is proposed in this paper. For SVM's parameters selection, culture - simulated annealing algorithm is used to optimize SVM parameters. Firstly the SVM model is trained with the optimal parameters, and then it is used to classify the video frames, and the shot boundary detection is implemented once. By extracting features of pixel domain and compressed domain and using a sliding window to form the feature vectors, it uses SVM for training. By comparing with other three related algorithms, this algorithm has the highest recall and precision performance.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their insightful suggestions. They also are grateful to the anonymous referees for their valuable comments and suggestions to improve the presentation of this paper.

REFERENCES

[1] J.S.Boreczky, L.A.Rowe; Comparison of video shot boundary detection techniques, in Proceedings of SPIE - The International Society for Optical Engineering, San Jose, CA, USA, **2670**, 170–179 **(1996)**.

[2] J.R.Cao, A.N.Cai; Algorithm for shot boundary detection based on support vector machine in compressed domain, Acta Electronica Sinica,**36(1)**, 203–208 **(2008)**.

[3] A.Hanjalic; Shot-boundary detection: Unraveled and resolved? IEEE Transactions on Circuits and Systems for Video Technology, **12(2)**, 90–105 **(2002)**.

[4] C.Jianrong, C.Anni; Shot transition detection method based on support vector machine in compressed domain, Pattern Recognition Letters, **28**, 1534–1540 **(2007)**.

[5] L.Feng, G.K.Saha, Yan Wen, Y.KunYan; Shot boundary detection using adaptive threshold and fourier fitting, Journal of Software, **23(zk2)**, 105–114 **(2012)**.

[6] R.G.Reynolds; An introduction to cultural algorithms, in Third Annual Conference on Evolutionary Programming, ser. Third Annual Conference on Evolutionary Programming, Singapore, Singapore, 131–9 **(1994)**.

[7] Z.Nan, X.Guoqiang, J.Jianmin, Q.Kaijin; Bpnn algorithm towards shot boundary detection, Journal of Computer Applications, May, **29(5)**, 1369–1372. **(2009)**.

[8] P.P.Mohanta, S.K.Saha, B.Chanda; A model-based shot boundary detection technique using frame transition parameters, Multimedia, IEEE Transactions on, **14(1)**, 223–233, **(2012)**.

[9] Y.Qi, A.Hauptmann, T.Liu, Supervised classification for video shot segmentation, in Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on, 1em plus 0.5em minus 0.4em IEEE, **2**, 2–689 **(2003)**.

[10] K.Wang, S.Yang, T.Dai; Method of optimizing parameter of least squares support vector machines by genetic algorithm, ComputerApplications and Software, **26(7)**, **(2009)**.

[11] Y.Feng, S.Luo, L.Wang; A new method of adaptive shot boundary detection, Transactions of Beijing Institute of Technology, **30(1)**, **(2010)**.

[12] C.Zhang, W.Wang; A robust and efficient shot boundary detection approach based on fisher criterion, in Proceedings of the 20th ACM international conference on Multimedia. 1em plus 0.5em minus 0.4em ACM, 701–704 **(2012)**.