International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM)
ISSN: 2150-7988 Vol.1 (2009), pp.214-221
http://www.mirlabs.org/ijcisim

DWT-based Shot Boundary Detection Using Support Vector Machine

Jun Li[1], Youdong Ding[1], and Wei Li[2], Yunyu Shi[1]

[1]School of Computer Engineering and Science, Shanghai University,
149, Yanchang Road, Zhabei District, Shanghai,China
vbli@163.com , ydding@shu.edu.cn , and syy2004042@163.com

[2] School of Communication and Information Engineering, Shanghai University,
149, Yanchang Road, Zhabei District, Shanghai,China
varlee@126.com

Abstract: Shot boundary detection is an important fundamental process in video data access, indexing, search, and retrieval. After a critical review of most approaches seeking to solve this problem, a novel shot boundary detection using discrete wavelet transform (DWT) and support vector machine (SVM) is proposed in this paper. To improve the performance of the algorithm and reduce the computational cost, shot boundary detection algorithms work by extracting the color and the edge features from wavelet transition coefficients. After that, multiple features are extracted from all frames within a temporal window. Finally, a multi-class support vector machine classifier is used to classify the video shot into three categories: cut transition (CT), gradual transition (GT) and normal sequences (NF). Numerical experiments in a variety of videos demonstrate that our method is capable of accurately detecting and discriminating shot transitions in videos with different characteristics.

Keywords: video retrieval, shot boundary detection, discrete wavelet transform, support vector machine**.**

1. Introduction

Due to the recent progress in the decreasing storage costs and the growing availability of broadband data connection, digital videos are becoming widely used. However, the increasing availability of digital video has not been accompanied by an increase in its accessibility. This is due to the nature of video data, which is unsuitable for traditional forms of data access, indexing, browsing, and retrieval. Traditional forms of data retrieval are either text based or based on the query-by-example paradigm [1]. If we want to find a clip of interest, we have to sequentially browse through the video. This is an extremely time consuming, tedious and labor-intensive process [2]. Therefore, the demand of new technologies and tools for effective and efficient indexing, browsing and retrieval of video data has been exacerbated by recent trends.

The area of content based video retrieval, aiming to automate the indexing, retrieval and management of video, has attracted extensive research during the last years [3]. The first step of content based video retrieval is shot boundary detection. A shot is defined as a sequence of frames taken by a single camera with no major changes in the visual content. Shot boundaries can be broadly classified into two types: abrupt transition and gradual transitions. Abrupt transition is instantaneous transition from one shot to the subsequent shot. Gradual transition occurs over multiple frames, which is generated via the application of more elaborated editing effects involving several frames, so that f_i frame belongs to one shot, frame f_{i+N} to the second, and the *N-1* frames in between represent a gradual transformation of f_i into f_{i+N}[4]. Gradual transition can be further classified into fade out/in(FOI) transition, dissolve transition, wipe transition, and others transition, according to the characteristics of the different editing effects [1][3]. Fig.1 shows a shot transition illustration.

1. Cut transition: This is instantaneous transition, where frame f_i belongs to one shot and f_{i+1} to the next shot, a clear discontinuity therefore existing [4].
2. Fade transition: This is a shot transition with the first shot gradually disappearing (fade out) before the second shot gradually appears (fade in). During the FOI, two shots are spatially and temporally well separated by some monochrome frames [2].
3. Dissolve transition: This is a shot transition with the first shot gradually disappearing while the second shot gradually appears. In this case, the last few frames of the disappearing shot temporally overlap with the first few frames of the appearing shot [1][2][3].
4. Wipe transition: This is actually a set of shot change techniques, where the appearing and disappearing shots coexist in different spatial regions of the intermediate video frames. One scene gradually enters across the view while another gradually leaves [1] [2].
5. Other transition types: There is a multitude of inventive special effects techniques used in motion pictures. They are very rare and difficult to detect [1].

The task of shot boundary detection is to identify the shot boundaries with their location and type in the given video clip(s) [5].

In this paper, we propose a new algorithm for shot transition detection. We choose the color and the edges in the vertical, the horizontal and the diagonal direction as the feature vector in shot detection system. To reduce the detection time, the algorithm uses the wavelet to extract the features. Traditionally, video shot segmentation approaches rely on threshold method, which do not generalize well since characteristics is different for the different videos. In this paper, a multi-class support vector machine (SVM) classifier is constructed to classify the video frames within a sliding window into cut transition, gradual transition and normal sequences.

Figure 1. Shot transition illustration. (a) Cut transition, (b) Fade, (c) Dissolve, (d) Wipe.

2. Review of previous works

To shot transition detection, many efforts have been devoted into this area for the past years, and many different methods have been proposed. Almost all shot change detection algorithms reduce the large dimensionality of the video domain by extracting a small number of features from one or more regions of interest in each video frame [1], such as the pixel-by-pixel comparison, luminance/color histogram differences, and the edge change ration. Transform coefficients and motion information have also been proposed for shot transitions. In recent years, methods that combine multiple features in order to increase the detection accuracy have been proposed [6] [7].

1. The initial way to check whether two frames are significantly different is the direct comparison of the pixels in the consecutive frames [8] [9] [10]. If the number of different pixels is large enough, the two processed frames are declared to belong to different shots. The pixel-based method is easy and fast. But it is extremely sensitive, since it has captured any details of the frame, such as highly sensitive to local motion, camera motion and minor changes in illumination [1] [3]. To handle these drawbacks, several ameliorative methods have been proposed, for example luminance/color histogram-based method and edge-based method.

2. Histogram-based method uses the statistics of the luminance and color. Xue L et al. [11] proposed a shot boundary detection measure that the features are obtained from the color histogram of the hue and saturation image of the video frame. Vasileios Chasanis et al [12] chose normalized RGB histogram as the feature vector in his shot boundary detection system. The advantage of the histogram-based shot change detection is that it is quite discriminant, easy to compute, and mostly insensitive to translational, rotational, and zooming camera motions. For these reasons, it is widely used. The weakness of the histogram-based shot boundary detection is that it does not incorporate the spatial distribution information of various color, hence it will fail in the case which similar histograms but different structures [1]. A better tradeoff between pixel and global color histogram methods can be achieved by block-matching methods [13] [14], in which each frame is divided into several nonoverlapping blocks and luminance/color histogram feature of each block are extracted.

3. The edge information is an obvious choice for characterizing image [1] [11] [15]. The advantage of this feature is that it is sufficiently invariant to illumination changes and several types of motion, and it is related to the human visual perception of a scene. Its main disadvantage is computational cost and noise sensitivity [1].

4. Most of the methods have used the motion information. The motions, either object motions or camera motions, exist in almost all video sequences. In order to distinguish shot

changes due to motion from those due to a shot boundary, differential motion based algorithms have been proposed. Park [16] presents a shot boundary detection based on the combination of two motion features: the modified displaced frame difference (DFD) and the blockwise motion similarity. T.LU tong [17] proposed a algorithm using motion compensation. As the main source of difference due to the motions can be eliminated by motion compensation, the remaining difference is due to likely to be a shot boundary.

5. Another important methods have been used is transform coefficients. Dai Xiaowen [18] proposed an algorithm of video shot detection based on partition in image wavelet entropy. Feng *et al.* [19] adopted the wavelet coefficient vectors within a sliding window as the features of the shot boundary detection system. Yu Wolf [2] presented a hierarchical multiresolution video shot transition detection scheme, which used wavelet to decompose every frame into low-resolution and high-resolution components.

Having defined a feature or a set of features for each frame, a shot change detection algorithm needs to detect where these exhibit shot change. The common practice of identifying the shot change between shots is to calculate the continuity (similarity) or discontinuity (distance) values of adjacent features. In the ideal situation, the continuity signal within the same shot always keeps large magnitudes, while drops to low values surrounding the positions of shot transitions[3]. Discontinuity is just an inverse of the continuity. The continuity (discontinuity) decision method can be done in the following ways.

1. Static threshold: This is the simplest decision method, which can be defined as :

$$SBD = \begin{cases} 1, & if \ sim(f_i, f_{i+1}) < T \\ 0, & toherwise \end{cases} \tag{1}$$

where T is a predefined threshold, $sim(f_i, f_{i+1})$ is the similarity value between adjacent frames. If $sim(f_i, f_{i+1})$ is less than threshold T, the classifier outputs 1 to declare the occurrence of shot transition. In contrast, if $sim(f_i, f_{i+1})$ is above the threshold, the classifier yields 0 to indicate no transition occurs between f_i and f_{i+1}.

2. Adaptive threshold: Different video content have different characteristics. It is hard to use one single threshold to detect all kinds of video efficiently. The obvious solution to the problems of static threshold is to vary the threshold according to the feature metrics within a temporal window [1].

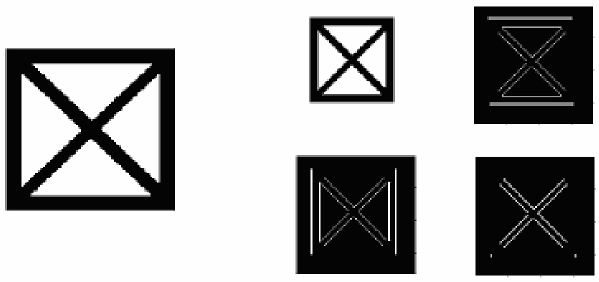
3. Statistical Machine Learning: Thanks to the good function learning and generalization capability, there have been some recent efforts turning to the tools of machine learning (e.g., a neural network, SVM, KNN). For example, in [3], [12], and [20], the shot boundary was classified by SVM. In [21], the shot boundary was detected by K-means.

From above discussion, we can see that many efforts have been undertaken to detect different kinds of shot boundaries, and a lot success has been achieved. However, in real applications, so far, no techniques of shot detection have been able to achieve the very ideal result. The main challenge is that conventional methods are not robust against camera operations, object motions, and the illumination changes, which often mistaken for shot boundary detection. Another challenge is the high computational cost of calculating all the features for each frame.

In this paper, we propose a new algorithm for shot transition detection. The approaches which mainly rely on single feature and single signal processing algorithms to try to capture all the characteristics of all kinds of shot transitions efficiently to detect shot transitions is very hard because different shot transitions have different characteristics. In this paper, we propose an algorithm for shot transition detection with multi-feature. We choose the color and the edges in the vertical, the horizontal and the diagonal direction as the feature vector in shot transition detection system. To reduce the detection time, the algorithm uses the wavelet to extract the features. Traditionally, video shot transition detection approaches rely on threshold method, which do not generalize well since characteristics is different for the different videos. In this paper, a multi-class SVM classifier is constructed to classify the video frames within a sliding window into cut transition, gradual transition and normal sequences. The testing result of the experiment shows that the method has good accuracy for shot boundary detection.

3. Feature Selection

3.1 Wavelet



(a) (b)

Figure 2. (a) Original Image, (b) Image DWT-detail

Wavelet is a nice tool to decompose an image signal into subbands. Not only can it give the desired low frequency and high frequency information we need to extract our feature parameters, but also it is fast and easy to compute and require only linear time in the size of the image [2]. DWT is a relatively transform and many mother wavelet functions are there [16] [22]. Hence we can choose an appropriate mother function for the problem under consideration. Here we utilize a Haar function as a mother wavelet because the Haar wavelet is the fastest to compute and the simplest to implement. Fig.2 shows a two dimensional DWT with a Haar mother wavelet function. The Fig.2(a) is an original image. The Fig.2(b) left upper quadrant is low frequency component of the image, denoted by LL. This looks very similar to the original image but the size is a half of the original one through a down sampling. The Fig.2 (b) right upper, left lower and right lower quadrant is called the vertical, the horizontal and the diagonal high frequency respectively. The vertical high frequency, denoted by LH, includes the information about a vertical element of the original image, which depicts the vertical edge of the original image. Similarly the horizontal DWT high frequency, denoted by HL, is the horizontal directional edge of the original image. The diagonal DWT high frequency, denoted by HH, corresponds to the diagonal directional difference of the original image. In this case, large color and edge features can be captured from the DWT image.

3.2 Feature mining based on wavelet

To define whether two frames are separated with an abrupt cut or a gradual transition we have to look for a difference measure between frames. It is hard to use one single feature to capture all the characteristics of all kinds of shot transitions efficiently because different shot transitions have different characteristics. In our approach we use multi features. The color and the edges in the vertical, the horizontal and the diagonal direction are chose as the feature vector in my shot detection system. The color and edges discontinuity between consecutive frames is very large in the cut transition. There are a series of local larger discontinuity where gradual change is occurring. The local motion, camera motion and the illumination change often cause some misses on shot change detection. The color is sufficiently variant to illumination changes and motion, but the edge is not. The edge is related to the human visual perception of a scene. The color difference between consecutive frames has a maximum value but the edge difference value is lesser where the illumination changes or motion is happening. Thus we can eliminate false detections that are caused by illumination changes and motion using the features of color and edge.

We partitioned a frame into $n \times n$ equal size blocks, and each block is decomposed through DWT in one levels. DWT decompose each block into four subbands, which is LL, HL, LH, HH bands. The variation of the color feature between frames can be calculated according to the low frequency coefficients of each decomposed block. The difference of the edge in the vertical, the horizontal and the diagonal direction between frames can be computed depending on the HL, LH, and HH bands coefficient of each decomposed block respectively. The detail algorithm is depicted as follows:

1. Color Difference

Let $E(l)$ denote total low frequency coefficient of the block L, which can be computed by following formula:

$$E(l) = \sum_{k=1}^{M} C_l(k) \qquad (2)$$

Where $C_l(k)$ is the coefficient value, M represents the total number of low frequency coefficient of the block l. Color difference of block l between two frames can be calculated by the following formula:

$$d_t(l) = E_t(l) - E_{t+1} \qquad (3)$$

In order to eliminate smooth intervals, produce a new sequence named $d'_t(l)$ according to the following formul:

$$d'_t(l)\begin{cases} 0 & d_t(l) < T_d \\ 1 & otherwise \end{cases} \qquad (4)$$

Finally the difference between frames based on their color feature is given from the following equation:

$$D_t^{LL} = \sum_{l=1}^{N} d'_t(l) \qquad (5)$$

In which, N represents the total number of block in one frame.

2. Edge Difference

The same as the above algorithm, the difference between frames based on their edge feature in the vertical, the horizontal and the diagonal direction can be obtained according to formula (2)-(5), which denote by D_t^{HL}, D_t^{LH}, D_t^{HH} respectively.

3.3 Sliding Window

It is intuitively clear that comparison with neighborhoods in temporal dimension plays an important role in deciding the class of a particular frame. To fully account for this, a sliding window is applied to a number of past and future frames of a frame. The feature vector of the current frame is then constructed with the feature data of the frames in the sliding window.

Figure 3 displays clearly the cut transition in the sliding window. The variance curve of the color and the edge difference in the Cut region forms a clear peak.

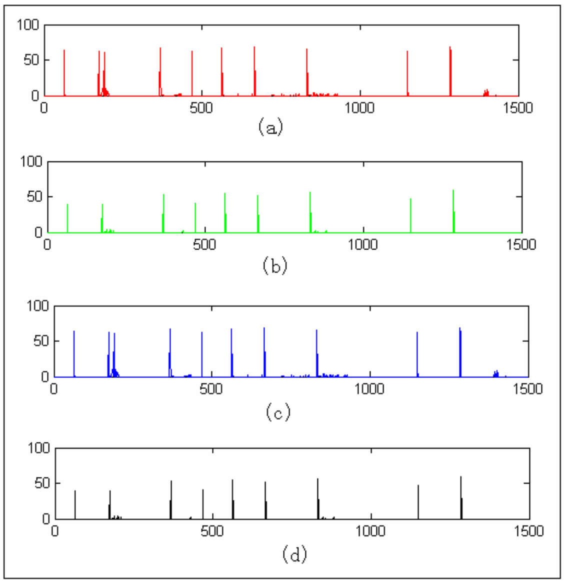


Figure 3. a sequence consists only cuts (a). variance curve in the color difference; (b). variance curve in the vertical direction difference; (c). variance curve in the horizontal direction difference ;(d). variance curve in the diagonal direction.

Figure 4 shows the gradual transition in the sliding window. There are a series of local larger value where gradual change is occurring.

In *Figure 3* and *Figure 4*, the smooth beeline intervals depict constant sequence.

The false detections that are caused by illumination changes or motion can be eliminated. The color difference between consecutive frames has a maximum value but the edge difference value is lesser where the illumination changes or motion is happening. Thus we can eliminate false detections that are caused by illumination changes and motion (seen in *Fig.5*).

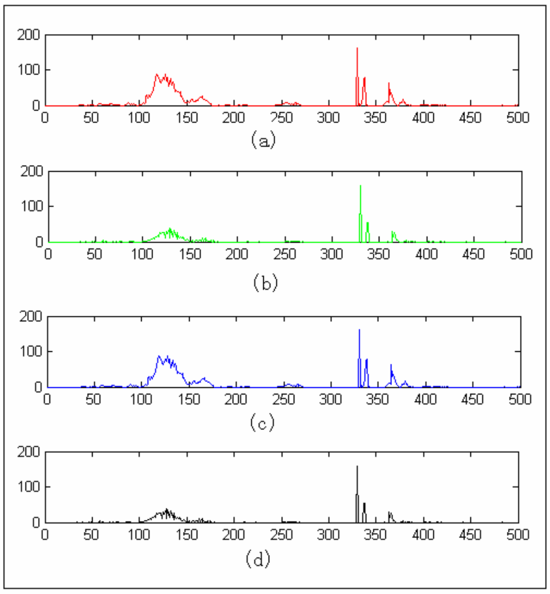


Figure 4. a sequence with gradual transitions. (a). variance curve in the color difference; (b). variance curve in the vertical direction difference; (c). variance curve in the horizontal direction difference ;(d). variance curve in the diagonal direction.

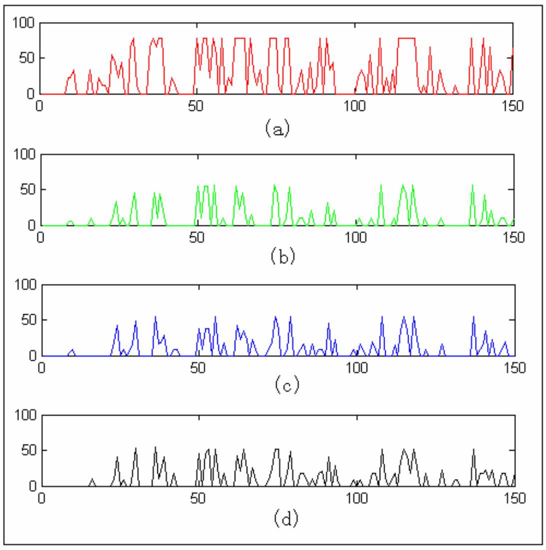


Figure 5. a sequence with illumination changes (a)variance curve in the color difference; (b)variance curve in the vertical direction difference; (c)variance curve in the horizontal direction difference;(d)variance curve in the diagonal direction.

4. Support Vector Machine Classifier

Having obtained the features of the continuity signal, the shot boundaries are detected by threshold scheme as what most of the existing methods do. However, the threshold method has several difficulties in achieving satisfactory results. First, the chosen threshold usually highly depends on the genres of videos. Second, a single threshold can not make full use of the contextual information of videos [3]. Therefore, in order to improve the performance of the algorithm we selected a classifier to identify the shot boundary detection. As for the selection of classifiers, SVM is preferred, not only for its solid theoretical foundations but also for its various empirical success. In the following, we will introduce how to apply SVM to implement the shot boundary detection.

4.1 SVM

SVM (Support Vector Machine) is a useful technique for data classification, which based on the concept of the structural risk minimization using the Vapnik-Chervonenkis (VC) dimension. A classification task usually involves with training and testing data which consist of some data instances. Each instance in the training set contains one "target value" (class labels) and several "attributes" (features). The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are given only the attributes [23]. The basic algorithm is:

Given a training set of instance-label pairs:

$$F = \{(x_1, y_1), \cdots, (x_l, y_l)\} \in (X, Y)^l \tag{6}$$

where, $x_i \in R^n$ is input vector. $y_i \in (1, -1)^l$ is the output vector.

The support vector machines (SVM) require the solution of the following optimization problem:

$$\begin{aligned} Min \quad & \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i \quad , \\ s.t. \quad & y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, \ldots, n, \end{aligned} \tag{7}$$

Here training vectors x_i are mapped into a higher (maybe infinite) dimensional space by the function Φ. Then SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. $C>0$ is the penalty parameter of the error term. Furthermore, $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ is called the kernel function. Then the SVM classifier function is obtained by solving the primal problem.

$$f(x) = sign(\sum_{i=1}^{l} w_i K(x_i, x) + b) \tag{8}$$

4.2 SVM Classifier

SVM is adopted to as the classifier to implement the shot boundary detection. We categorize shot boundary in three categories: normal sequences, abrupt cuts and gradual transitions. Nevertheless, the SVM is originally designed for binary classification. In our application, we have a three-class problem, thus we used the "one-against-one" approach in which for a k-class problem, *k(k-1)/2* binary classifiers are constructed and each one is trained to discriminate data from two classes. In the following, we will describe how to apply "one-against-one" SVM to implement the shot boundary detection.

(1). Let a set of training sample which can be obtained by human observer identifying.

$$F = \{(x_1, y_1), \cdots, (x_i, y_i), \cdots, (x_l, y_l)\} \in (X, Y)^l \tag{9}$$

where x_i is input feature vector. The feature vector of the current frame is constructed with the feature data of the frames in the window. Suppose that the size of the window is n. The total number of features for a particular frame is then *4(n-1)* since there are four features per frame. Here we select n as 15. There are fifty five- dimensional feature vector as the input vector of SVM.

$$x_i = (D_{t-7}^{LL}, \cdots, D_t^{LL}, D_t^{LH}, D_t^{HH} \\ D_t^{HL}, \cdots, D_{t+7}^{HH}) \in R^{56} \quad (10)$$

In formula (9), y_i is output vector, $y_i \in (NF,CT,GT)$. If we assume that class label 1 corresponds to normal sequences, class label 2 to hard cuts and class label *3* to gradual transitions, three binary classifiers discriminating between pairs of classes (1, 2), (1,3) and (2,3) are constructed. *Figure 6* shows the theory of a 3-class SVM classification.

The decision function of classifier *(i, j)* is:

$$Min \quad \frac{1}{2}(w^{ij})^T w^{ij} + C\sum_{jt}^{l} \xi_t^{ij}$$

$$s.t. \quad \begin{cases} (w^{ij})^T \phi(x_t) + b^{ij} \geq 1 - \xi_t^{ij}, & if\ y_t = i \\ (w^{ij})^T \phi(x_t) + b^{ij} \geq -1 + \xi_t^{ij}, & if\ y_t = j \\ \xi_t^{ij} \geq 0 \end{cases} \quad (11)$$

In the final classification we use a voting strategy where the decision of each binary classifier is considered as a vote for its proposed class and the class with the maximum number of votes is selected.

We join the result of SVM classification and get a series of frame labels which indicate NF, CF and GF, separately. Then the cut transition, gradual transition and normal sequences can be detected.

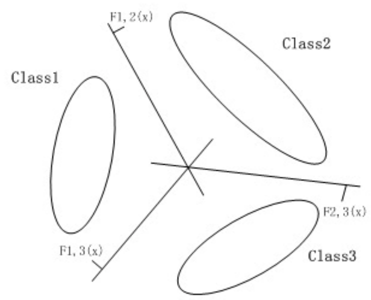


Figure 6. one against one method.

5. Experimental Results and Discussions

In this section, we will carry out several experiments on the platform of TRECVID test data [24] and the four action movies. Totally, there are 8 MPEG clips, 238709 frames, 1454 shots, which include cut and dissolve, fade out/in, wipe gradual transition. Table 1 lists all the videos for the experiment. For all sequences, a human observer identifies the shot boundaries as the ground truth. All of the experiments are conducted with Matlab.

The performance of a shot transition detection algorithm is usually measured with terms of recall and precision. The recall and precision are defined as following:

$$recall = \frac{N_c}{N_c + N_m} \times 100\% \quad (12)$$

$$precision = \frac{N_c}{N_c + N_f} \times 100\% \quad (13)$$

where, N_c is the number of correct detections, N_m is the number of missed detections, N_f is the number of false detections. A good shot transition detector should have both high precision and high recall.

Firstly, we should decide some parameter in the experiment. For SVR, we use the software Libsvm provided by the National Science Council of Taiwan to do SVM classification [23]. We have tested the performance on different kernels and find that the "RBF" kernel outperforms others. There are two parameters while using RBF kernels: the penalty parameter *C* and *r*. It is not known beforehand which *C* and *r* are the best for one problem. We adopt the cross-validation method to obtain the *C* and *r* in this paper.

Video	Frame	Cut	Graudal
V_1	914	0	9
V_2	2494	12	1
V_3	8391	35	15
V_4	32370	156	48
V_5	59540	276	92
V_6	45000	253	6
V_7	45000	242	11
V_8	45000	291	7
Total	**238709**	**1265**	**189**

*Table 1.*Video for the experiment.

Y. Kawai [6] has proposed a method of shot boundary detection based on multiple features, which is the more successful one in lots of algorithms. Zhang algorithm [9] used the twin comparison for shot transition detection. We used Y. Kawai and Zhang algorithm for our comparison study. Table 3 and 4 lists the performance of the proposed algorithm compared with Y. Kawai and Zhang algorithms. For cut, the recall rate of our method is 92.4%. The recall rate of Y. Kawai method is 91.161%, and the recall rate of Zhang method is 79.7%. The precision rate of our method is 95.4%. The precision rate of Y. Kawai method is 93.183, and the precision rate of Zhang method is 79.2%. In our system, the multi-feature and multi-class SVM is chosen to calculate the characteristics of the shot boundary detection, which is robust to luminance changes, object motion, and camera operations, so the proposed algorithm produces better performance compared with Y. Kawai and Zhang algorithms.

For gradual transitions, the recall rate of our method is 67.619%. The recall rate of Y. Kawai method is 63.376%, and the recall rate of Zhang method is 57.34%. The precision rate of our method is 63.2%. The precision rate of Y. Kawai method is 62.043%, and the precision rate of Zhang method is 59.7%. From the experimental results, we can find that the proposed algorithm produces better performance compared with Y. Kawai and Zhang algorithms.

Method	Recall (%)	Precision (%)
Our method	92.4	95.4
Zhang method	75.7	79.2
Y. Kawai method	91.161	93.183

Table 2. CT comparison with the Zhang and Y. Kawai algorithms.

We have investigated the incidence of missed detections and false detections in our system. For cut, many missed detections occurred because the successive abrupt shot changes with similar backgrounds, as shown in Figure 7. False detections were mostly caused by movements of the camera or object during the transition (Figure 8).

Method	Recall (%)	Precision (%)
Our method	67.619	63.2
Zhang method	57.34	59.7
Y. Kawai method	63.376	62.043

Table 3. GT comparison with the Zhang and Y. Kawai algorithms.

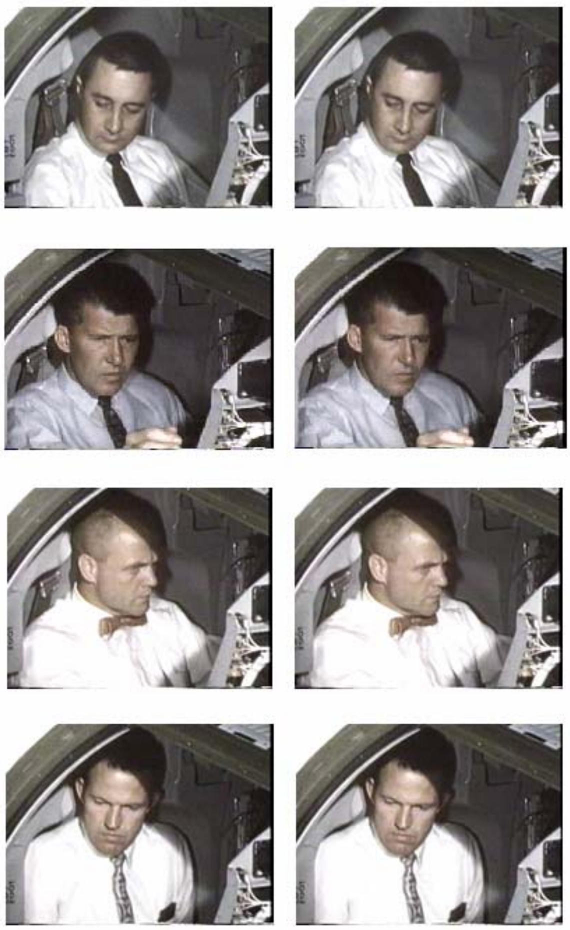


Figure7. Example of missed detections.

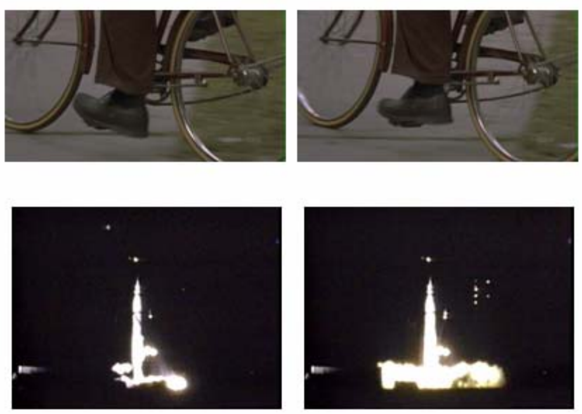


Figure 8. Example of false detections.

6. Conclusions

We have presented our complete framework of a video shot transition detection methodology. Unlike previous approaches which mainly rely on single feature and single signal processing algorithms to try to detect shot transitions, our method detects and identifies shot transitions with different features. By "one-against-one" SVM classifier, the videos shot are classified into hard cuts, gradual transitions and normal sequences. The testing result of the experiment shows that the method has good accuracy for shot boundary detection.

Acknowledgment

This work is supported by National Science and Technology Support Plan(No.2006BAK13B10), National Natural Science Foundation of China(No.60605012), Shanghai Educational Committee Leading Academic Discipline Project(No.J50103), Shanghai Educational Committee Innovation Research Project(No.09YZ10) and Shanghai University Graduate Student Innovation Plan

References

[1] C. Cotsaces, N. Nikolaidis, and I. Pitas. “Video Shot Detection and Condensed Representation”. *Journal of IEEE Signal Processing Magazine*, March, pp. 28--37, 2006.

[2] H. H. YU, W. WOLF, “A hierarchical multiresolution video shot Transition detection scheme[J]”, *Journal of Computer Vision and Image Understanding,* 75(1/2), pp. 196--213, 1999.

[3] J. H. Yuan, H. Y. Wang, and B. Zhang. “A formal study of shot boundary detection”. *Journal of Transactions on Circuits and Systems for Video Technology*, 17(2), pp. 168--186, February 2007.

[4] A. Hanjalic. “Shot-Boundary Detection: Unraveled and Resolved?”. *Journal of IEEE Transactions on Circuits and Systems for Video Technology*, 12(2), pp.90--105, 2002.

[5] TRECVID-2007: “Shot Boundary Detection Task Summary”. In TRECVID 2007 Workshop, Gaithersburg.

[6] Y. Kawai, H. Sumiyoshi, and N. Yagi. “Shot Boundary Detection at TRECVID 2007”. In TRECVID 2007 Workshop, Gaithersburg.

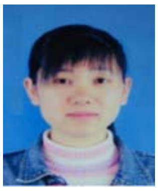
[7] C. Grana, and R. Cucchiara. “Linear Transition Detection as a Unified Shot Detection Approach”. *Journal of IEEE Transactions on Circuits and Systems for Video Technology*, 17(4), pp. 483—489, April 2007.

[8] A. Nagasnlm, Y.Tanaka. “Automatic video indexing and full-motion search for object appearance”. *In Proc. IFIP 2nd Working Conference on Visual Databases Systems*, Budapest, Hungary, 30(3), pp. 113--127, September 1991

[9] H. Zhang, A. Kankanhalli, and S.W. Smoliar. “Automatic partitioning of full-motion video”. *Journal of Multimedia Systems on ACM*, 1(1), pp. 10--28, 1993

[10] S.~K. Choubey and V.~V. Raghavan. "Generic and fully automatic content-based image retrieval using color". *Journal of Pattern Recog*, 18(11-13), pp. 1233--1240, 1997.
[11] L. Xue, C. Li, H. Li, and Z. Xiong. "A general method for shot boundary detection". *In Proceedings of the 2008 International Conference on Multimedia and Ubiquitous Engineering*, PP.394--397, 2008.
[12] V. Chasanis, A. Likas, and N. Galatsanos. "Simultaneous detection of abrupt cuts and dissolves in videos using support vector machines". *Journal of LSEVIER Pattern Recognition Letters*, 30, pp.55--65, 2009
[13] Jesús Bescós, Guillermo Cisneros, José M. Martínez, José M. Menéndez, and Julián Cabrera. "A Unified Model for Techniques on Video-Shot Transition Detection". *Journal of IEEE TRANSACTIONS ON MULTIMEDIA*, 7(2), pp. 293—306, April 2005.
[14] Z.~p. Zong, K. Liu, and J.~H. Peng. "Shot Boundary Detection Based on Histogram of Mismatching-Pixel Count of FMB". *In Proceedings of ICIEA 2006*, pp. 24--26, 2006
[15] H.~ZHAO, X.~H. LI. "Shot Boundary Detection Based on Mutual Information and Canny Edge Detector". *In Proceedings of 2008 International Conference on Computer Science and Software*, pp:1124--1128, 2008.
[16] M.~H. Park, R.~Hong Park, and S.~W Lee. "Efficient Shot Boundary Detection for Action Movies Using Blockwise Motion-Based Features". *In Proceedings of LNCS Advances in visual computing*, pp. 478—489, 2005.
[17] T. LU. "An Accumulation Algorithm for Video Shot Boundary Detection. Multimedia Tools and Applications". *Journal of Multimedia Tools and Applications*, 22, pp. 89--106, 2004.
[18] X.~W. DAI, Z.~P. CAI. "An Algorithm of Video Shot Detection Based on Partition in Image Wavelet Entropy". *Journal of ACTA PHOTONICA SINICA*, 37(7), pp. 1492--1496, July 2008
[19] H. Feng, W. Fang, S. Liu, and Y. Fang, "A new general framework for shot boundary detection and key-frame extraction," *In Proceedings of 7th ACM SIGMM Int. Workshop Multimedia Inf. Retrieval*, pp. 121–126, 2005.
[20] Jianrong Cao, Anni Cai. "A robust shot transition detection method based on support vector machine in compressed domain". *Journal of LSEVIER Pattern Recognition Letters*, 28, pp. 1534–1540, 2007.
[21] M. R. Naphade, R. Mehrotra, A. Ferman, J.Warnick, T. S. Huang, and A. M. Tekalp, "A high-performance shot boundary detection algorithm using multiple cues," *In Proceedings of the IEEE Image Process*, pp.884–887 1998.
[22] Tao Li, Qi li, Shenghuo Zhu, Mitsunori Ogihara. "A survey on wavelet applications in data mining". *Journal of ACM SIGKDD Explorations Newsletter*, 4(2), pp. 49 – 68, December 2002.
[23] C.~W. Hsu, C.~C. Chang, and C.~J. Lin. A Practical Guide to Support Vector Classification. Available at http://www.csie.ntu.edu.tw/~cjlin}.
[24] http://www.open-video.org

Author Biographi



Jun Li received the B.E. degree in computer science and technology from Jiamusi University, Jiamusi, China, in 2003. She is currently working toward the Ph.D degree in the school of Computer Engineering and Science, Shanghai University, Shanghai, China. Her major research interests include content based video retrieval, pattern recognition, and machine learning.

Youdong Ding received the B.E degree and the Ph.D. degree from University of Science and Technology of China, Hefei, Anhui, China, in 1990 and 1997, respectively. He is a Professor in the school of Computer Engineering and Science, Shanghai University. His major research interests include multimedia information processing, content based video retrieval, Image Quality Assessment.

Wei Li received M.S degree from the School of Journalism and Communication, Wuhan University, Wuhan, China. She is currently pursuing the Ph.D. degree in the School of Communication and Information Engineering, Shanghai University, Shanghai, China. Her research interests include machine learning for video content analysis, and integration of video and music information into music video information retrieval systems.

Yunyu Shi received the B.S. and M.S. degrees from East China Institue of Technology in 2004 and 2007, respectively. She is working toward the Ph.D degree in Shanghai University. Her current research interests are image/video quality assessment.