Shot Boundary Detection in MPEG Videos using Local and Global Indicators

Jinchang Ren, Jianmin Jiang and Juan Chen

***Abstract*— Shot boundary detection (SBD) plays important roles in many video applications. In this paper, we describe a novel method on SBD operating directly in compressed domain. Firstly, several local indicators are extracted from MPEG macroblocks, and AdaBoost is employed for feature selection and fusion. The selected features are then used in classifying candidate cuts into five sub-spaces via pre-filtering and rule-based decision making. Following that, global indicators of frame similarity between boundary frames of cut candidates are examined using phase correlation of DC-images. Gradual transitions like fade, dissolve and combined shot cuts are also identified. Experimental results on the test data from TRECVID'07 have demonstrated the effectiveness and robustness of our proposed methodology.**

***Index Terms*— shot boundary detection, TRECVID, video segmentation, decision making, video signal processing.**

I. Introduction

SHOT boundary detection (SBD) is the fundamental task in content-based analysis, indexing and retrieval of videos, as it helps to provide a hierarchical structure of video and enables extraction of meaningful highlights from such a structure [1-4]. As a result, it has continuously attracted extensive attentions on this topic, which was also one of the motivations for the well-known TREC Video Retrieval Evaluation (TRECVID) activity, providing objective samples as a common platform on SBD and other video processing tasks [5].

In general, there are at least two steps for shot boundary detection, i.e. extracting features in either compressed or uncompressed domain to construct dissimilarity metrics between adjacent frames, and making decisions based on these metrics. In uncompressed domain, frame difference is usually measured using pixel difference [4], histogram [1, 11], texture or edge [11], motion [10, 15], and frame correlation [3]. In compressed domain, the most frequently used features are DC-image [12], macroblock types [14], edges [13] as well as DCT coefficients, motion vectors and bit-rate information [4, 6].

With extracted features, a continuity signal can be constructed using pair-wise comparison or temporal filtering [1]. Afterwards, shot changes are determined in several ways, including thresholding [14], (fuzzy) decision making [11], machine learning, clustering [1, 10], mutual information [9], and model-based approaches [8, 13]. Since modeling and statistical analysis usually needs some prior knowledge and assumptions such as shot length [7, 10, 13], they may produce unsatisfactory results if these assumptions cannot be met.

In this paper, detailed techniques used for our submission to TRECVID 2007 on SBD are presented, in which our main contributions can be highlighted as: (i) By extracting several novel features as local content indicators, robust shot detection is achieved in a very small set of selected features; (ii) By categorizing shot cuts into five classes, abrupt shot changes and several gradual transitions are effectively detected; (iii) A fast implementation of such a system is presented that fully operates in compressed domain. Evaluation results by TRECVID test data indicates that our method achieve the best results on cut detection, sixth best on gradual transition detection, and third best on overall performances among all participation teams worldwide. Such evaluation also supports that our method is effective and robust on a wide range of video sources.

II. Feature Extraction and Selection

Unlike most of the existing techniques working in pixel-domain or directly on DC-images, our proposed method defines features via statistical analysis of the *difference* between two consecutive DC-images as discussed below.

A. Feature Extraction

First of all, DC-images are extracted from each input frame f_i in MPEG videos, which provides a low-resolution version of the original frame for further analysis. Let $Y_{dc}^{(i)}, U_{dc}^{(i)}$ and $V_{dc}^{(i)}$ be the corresponding DC-images of the luminance and chrominance components , a **DC-differencing image** between the i^{th} frame and the $(i+1)^{th}$ frame can be defined below:

$$D(i) = 3^{-1}\sum |\, Ch_{dc}^{(i)} - Ch_{dc}^{(i+1)} \,|, \qquad Ch = Y, U, V \tag{1}$$

For each $D(i)$, its **mean** and **standard derivation** are determined as $\mu(i)$ and $\sigma(i)$. Further, we define $p_1(i)$ and $p_2(i)$ as **two proportions** representing the percentage of pixels in $D(i)$ that are larger than two thresholds $\lambda_1(i)$ and $\lambda_2(i)$, where $\lambda_2(i) = \mu(i)/4$ and $\lambda_1(i) = \lambda_2(i) + 0.5$. As $\lambda_1(i) > \lambda_2(i)$, we have $p_1(i) \le p_2(i)$. Since $\lambda_1(i)$ and

Manuscript received November 8, 2007. This work was supported by the EU IST Framework Research Programme under both HERMES project (Contract No IST-216709) and LIVE project (Contract No IST-4-027312). This paper was recommended by Associate Editor E. Izquierdo.

J. Ren, J. Jiang and Juan Chen are with Digital Media & System Institute, School of Informatics, University of Bradford, U. K. (j.ren@bradford.ac.uk, j.jiang1@bradford.ac.uk, j.chen12@bradford.ac.uk).

$\lambda_2(i)$ are dependent on $\mu(i)$, they present an adaptive thresholding mechanism, which makes $p_1(i)$ and $p_2(i)$ robust to the luminance changes across frames inside shot cuts.

In addition, a **motion prediction error** $err(i)$ is defined for the i^{th} frame as given below, where N_i is the number of non-intra coded blocks. Also we define a normalized energy $E_y(i)$ in which E_{0_y} is the maximum value of energy in Y component and N_y is the number of DC-coefficients in $Y_{dc}^{(i)}$.

$$err(i) = N_i^{-1}\sum Y_{dc}^{(i)}(j), \quad 1 \le j \le N_i \tag{2}$$

$$E_y(i) = E_{0\_y}^{-1}\sum [Y_{dc}^{(i)}(j)]^2, \quad 1 \le j \le N_y \tag{3}$$

B. Feature Selection

To optimally choose a group of features for cut detection, we employed AdaBoost [16] to exploit its power in classification and optimization. In addition to the described features including $\mu(i)$, $\sigma(i)$, $p_1(i)$, $p_2(i)$ and $err(i)$, five traditional features are also extracted including indicators of luminance, color, motion magnitude, edge and inter-frame differences. In our system, a 5-fold cross-validation process is employed, using the test data from TRECVID in 2006 and 2005 with manual ground truth maps for training purposes.

Our test results are summarized into three groups as illustrated in Table 1, where the first two tests use a temporal window of 11 frames and the third uses a window of 3 frames. While our selected features have indeed produced improved results even with much lower dimensions, it is still very difficult to accurately characterize all the cuts, due to the fact that cuts in reality present a wide range of inconsistent visual appearances. To this end, we propose to classify cuts into a number of categories with relatively consistent visual appearances so that more accurate characterization within each individual category of cuts can be achieved.

Table 1. Performance comparison using AdaBoost based cross validation on the data from TRECVID in 2006 and 2005.

Experiments	Test 1	Test 2	Test 3
Feature Dimension	10*11	5*11	5*3
Average recall	0.9709	0.9718	0.9716
Average precision	0.9702	0.9706	0.9705

III. Modeling and Shot Boundary Detection

In this section, cut detection is modeled as a process of decision-making, where cuts are categorized into five sub-spaces according to their visual appearances. As a result, a coarse-to-fine process is employed for the SBD as follows.

A. Modeling

If we take the cut detection problem as a process of decision making, then we have one feature space Ψ and one decision space Ω. Let Θ be the decision-making process, we should simply have $\Theta(\Psi) = \Omega$. Since cuts may have various appearances under different contexts, the feature space Ψ is further divided into K sub-spaces, namely $\Psi = \bigcup \Psi_k$ and $\Psi_m \bigcap \Psi_n = \Phi \quad if \quad m \ne n$. In fact, we have $K = 5$ in our implementation, leading to $\Omega_k \mid k \in [1,5]$, where $\Theta(\Psi_k) = \Omega_k$ and each Ω_k can be taken as one category of cuts which has its own characteristics of visual appearances.

In Ω_1 and Ω_2, two boundary frames of a cut almost share nothing in both background and foreground. The difference between these two is that, in Ω_1, we can find very large change of intensity in frame images while in Ω_2 the intensity change is limited, although some color difference may be apparent. These cuts should appear as a peak of $\mu(i)$ and $\sigma(i)$. Consequently, we expect a large peak for cuts in Ω_1. In Ω_3, there is a relative large part of common background or foreground during shot changes, which will inevitably lead to lower difference of the two boundary images. Therefore, lower peaks of $\mu(i)$ and $\sigma(i)$ are expected.

Ω_4 characterizes those shot boundary changes, where a shot cut is followed by sudden intensity changes such as the effect of flash lighting etc. This will lead to a large peak of $\sigma(i)$, $\mu(i) > \mu(i-1)$, $\mu(i) < \mu(i+1)$, and large prediction errors satisfying $err(i-1) < err(i) < err(i+1)$. Finally, Ω_5 contains shot changes followed by strong motions, which are reflected by large frame differences across several frames, indicated by a peak of $\mu(i)$ and a large $err(i+1)$. Under this circumstance, the value of $\sigma(i)$ does not generally present any apparent peak.

B. Pre-filtering of Cuts

Pre-filtering is to remove those frames with very limited changes, which are considered as non-cuts from its neighbors, in order to achieve high level of efficiency and robustness for shot cut detection. Since a cut often causes an overall change of the visual content inside boundary frames, such changes will inevitably lead to a larger value of $\mu(i)$. Yet due to the fact that certain level of consistency is maintained between differential frame pixels, the value of $\sigma(i)$ is relatively small. As a result, we propose to use the condition, $\sigma(i) < \rho\mu(i)$ where $\rho > 1$, as the first step for the pre-filtering process.

As for $p_1(i)$ and $p_2(i)$, they are mainly used to represent the percentage of active (changed) blocks in frames. When a cut occurs, there should be a large percentage of changed areas across neighboring frames. As a result, we use two condition tests to remove those non-cuts frame differences. The first condition is: $p_2(i) > p_0$ where $p_0 \in (0,1)$, which specifies a minimum requirement of the changed macroblocks. Although $p_1(i) \le p_2(i)$, $p_1(i)$ and $p_2(i)$ should be close to each other and this is constrained by $\rho_c p_1(i) > p_2(i)$ where $\rho_c > 1$.

Further, cuts generally satisfy $p_2(i)$ is greater than both $p_2(i-1)$ and $p_2(i+1)$, which indicate stronger block changes at the current frame, where cut occurred, than its neighboring ones. Considering the effect of noise such as motion and lighting changes, however, some cuts may not necessarily produce sufficiently large block changes. To this end, the condition below is used to complete the pre-filtering.

$$\rho_c p_2(i) > \max(p_2(i-1), p_2(i+1))$$

C. Decision Rules for the Five Categories of Cuts

Most existing work detects cuts via thresholding peak values of certain features such as $\mu(i)$ and $\sigma(i)$. However, such technique often fails to achieve sufficient robustness, especially in cases where cuts do not generate sufficiently strong block changes. Therefore, we propose to measure their relative peak values as a changing ratio with respect to that of its neighboring frames to complete the cut detection as follows.

$$\mu_{\min}(i) = \min(\mu(i)/\mu(i-1), \mu(i)/\mu(i+1)) \quad (4)$$

$$\sigma_{\min}(i) = \min(\sigma(i)/\sigma(i-1), \sigma(i)/\sigma(i+1)) \quad (5)$$

where $\mu_{\min}(i) > 1$ and $\sigma_{\min}(i) > 1$ represent a peak of $\mu(i)$ and $\sigma(i)$, and their values are good indications to show the strength of sudden changes between the current frame and its neighboring frames included in a cut transition.

As larger values of $\mu_{\min}(i)$, $\sigma_{\min}(i)$ and $p_2(i)$ are more likely to indicate a potential cut, three likelihoods of cuts, $\ell_{i,k}(\mu)$, $\ell_{i,k}(\sigma)$ and $\ell_{i,k}(p_2)$, are extracted below:

$$\ell_{i,k}(\mu) = 1 - [\rho_\mu(k)\mu_{\min}(i)]^{-1} \quad (6)$$

$$\ell_{i,k}(\sigma) = 1 - [\rho_\sigma(k)\sigma_{\min}(i)]^{-1} \quad (7)$$

$$\ell_{i,k}(p_2) = \begin{cases} \sqrt{p_2(i)} & k = 1,2,3 \\ p_2(i) & k = 4,5 \end{cases} \quad (8)$$

where parameters $\rho_\mu(k)$ and $\rho_\sigma(k)$ are used to characterize the visual appearances of cuts within each likelihood. In (8), likelihood is designed in consideration of individual features across all five categories, where $\ell_{i,k}(p_2) = p_2(i)$ is required to establish the likelihood for cuts in Ω_4 and Ω_5. This is due to the fact that both Ω_4 and Ω_5 often contain more active blocks to indicate a cut. Yet for other three categories, $\ell_{i,k}(p_2) = \sqrt{p_2(i)}$ is sufficient to yield similar likelihood.

Through weighting of these three separate likelihoods, a combined likelihood $\ell(i,k)$ is obtained below:

$$\ell(i,k) = [\ell_{i,k}(\mu) + \ell_{i,k}(\sigma) + \ell_{i,k}(p_2)]/3 \quad (9)$$

Given the combined likelihood values, $\ell(i,k)$, the category of the **maximum likelihood** is determined as:

$$k_0 = \arg\max_k \ell(i,k) \mid k \in [1,5] \quad (10)$$

A candidate in category Ω_{k0} is detected as a cut if we have $\ell(i,k_0) > \ell_0$, where $\ell_0 \in (0,1)$ is a constant threshold. Otherwise, it is a false alarm. Here we choose a relatively small ℓ_0 to allow most possible cuts to be detected as they will be further validated in the next section.

Regarding $\rho_\mu(k)$, it is determined as follows. Due to the apparent luminance changes, there exists a very large peak of $\mu(i)$ in Ω_1 but relative small peaks in Ω_2, Ω_3 and Ω_5. Therefore, to obtain a high likelihood $\ell_{i,k}(\mu)$, $\rho_\mu(1) = 1$ is sufficient for Ω_1. For Ω_2, Ω_3 and Ω_5, however, ρ_μ is decided as: $\rho_\mu(2) = \rho_\mu(3) = \rho_\mu(5) = t_\mu$ and $t_\mu > 1$.

For cuts in Ω_4, $\rho_\mu(4)$ is determined in a way to exploit the fact that $err(i+1) > err(i) > err(i-1)$, and both $\mu(i)$ and $\mu(i+1)$ are larger than $\mu(i-1)$, as suggested by its definition. Hence we have $\rho_\mu(4) > \rho_{err\_\mu}(i) > 1$ and larger $\rho_\mu(4)$ will lead to higher cut likelihood of $\ell_{i,4}(\mu)$.

$$\rho_\mu(4) = \rho_{err\_\mu}(i)\min[\mu(i), \mu(i+1)]/\mu(i-1)$$
$$\rho_{err\_\mu}(i) = [\frac{err(i+1)}{err(i)} + \frac{err(i)}{err(i-1)}]/2 \quad (11)$$

For $\rho_\sigma(k)$, it is determined to enable $\ell_{i,k}(\sigma)$ to have appropriate likelihood values for all five cut categories below:

$$\rho_\sigma(k) = \begin{cases} t_\sigma, & k = 1,3,4 \\ 1 + \mu(i)/\sigma(i), & k = 2 \\ \frac{err(i+1)}{err(i)} \frac{\min[\sigma(i), \sigma(i+1)]}{\sigma(i-1)}, & k = 5 \end{cases} \quad (12)$$

where $t_\sigma > 1$ for cut category Ω_1, Ω_3 and Ω_4 in order to maintain a high likelihood value in (7); $\rho_\sigma(2)$ is defined as such that a smaller value of $\mu(i)/\sigma(i)$ can be expected. To derive a high likelihood for Ω_5, we expect large differences in both motion prediction error $err(i)$ and the standard derivation $\sigma(i)$, according to the definition of Ω_5.

D. Validation of Detected Cuts

For motion caused scene changes, as an example, although they may have large $\mu(i)$ and $\sigma(i)$ making it like a real cut, the overall similarity of the two frames is still high. In contrast, the boundary frames of a real cut remains to be dissimilar. Therefore, the similarity of boundary frames can be used as a good indicator to validate our detected cuts. Among many

techniques proposed in measuring such a similarity, we propose to use phase correlation on extracted DC-images to validate the detected results for both efficiency and robustness.

E. Detecting Gradual Transitions

After cut detection, we need to identify boundaries of gradual transitions within each pair of neighboring cuts. Techniques on detection of gradual transitions including fades, dissolves and combined shot cuts are discussed below.

From instructions given by TRECVID'07, combined cuts are defined to contain cuts and a series of monochrome (black or white) frames, and they can be classified into two parts, i.e. normal cuts in the boundary and monochrome frames in the middle. To detect such patterns, frame energy is found to be sufficient to complete the detection due to the fact that the energy among all monochrome frames present little changes in the transition, yet in the boundaries, the energy presents dramatic changes with either very large or very small values.

Detection of dissolves mainly relies on identification of a downward-parabolic or a U-shape pattern [1]. In real situations, however, such detection is inaccurate and lack of robustness since the U-shape is often distorted by the noise. Since a dissolve contains information from two different frames, a high motion prediction error $err(i)$ and a large value of $\mu(i)$ in corresponding frames can be expected. Therefore, candidates for dissolves can be detected via thresholding both $err(i)$ and $\mu(i)$. Their validation as detected dissolves can be completed by measuring the similarity between boundary frames.

As for the event of fade, it is detected only if a fade-out event is followed by a fade-in, i.e. fade out/in (FOI). During such a FOI process, one apparent appearance is the change of luminance, where its intensity values present a clear V-shape. As a result, the left and right sides of this V-shape are detected as fade out and fade in, respectively.

F. Determining Parameters

In the following, we will discuss how the parameters in our algorithm are determined. Taking p_0 in pre-filtering of cuts as an example, firstly the probability density functions (PDFs) of $p(p_2 \mid cut)$ and $p(p_2 \mid !cut)$ are extracted from the training set. Then, an overall cost of error classification $\gamma(p_0)$ is obtained where a larger η may help to obtain a smaller p_0.

$$\begin{aligned} \gamma(p_0) &= \eta a_c(p_0) + (1-\eta) a_{\bar{c}}(p_0) \\ a_c(p_0) &= \int_0^{p_0} p(p_2 \mid cut) dp_2 \\ a_{\bar{c}}(p_0) &= \int_{p_0}^{1} p(p_2 \mid !cut) dp_2 \end{aligned} \quad (13)$$

As our strategy is to make all real cuts above the selected threshold, we intend to choose a larger η to generate a smaller p_0 to reduce missed cuts. Hence parameter η is decided by:

$$\eta = \left[\frac{a_c(p'_0)}{a_c(p'_0) + a_{\bar{c}}(p'_0)} \right]^{\alpha} \quad (14)$$

where p_c and $p_{\bar{c}}$ satisfy $a_c(p_c) = a_{\bar{c}}(p_{\bar{c}}) = 1\%$ and $p'_0 = (p_c + p_{\bar{c}})/2$. When $\alpha < 1$, we have smaller missing rate $a_c(p_0)$ but larger false alarm rate $a_{\bar{c}}(p_0)$. In contrast, $\alpha > 1$ leads to a larger missing rate and smaller false alarm rate. To obtain smaller missing rate, we select $\alpha = 0.5$ to determine all relevant parameters.

IV. Results and Discussions

In this section, comprehensive comparisons of experimental results are discussed under TRECVID framework to verify our methodology in detecting both cuts and gradual transitions.

A. Data Preparation

Unlike news video used in previous years, test data in TRECVID'07 covers a wide range of sources including news, documentaries, educational programmes, and archived videos in black and white [5]. The 6-hour data in MPEG-1 format is selected from 400-hour video sources containing 2320 shots in 17 sequences. For quantitative evaluation, manual ground truth (GT) data is provided. It is worth noting that errors are likely in the provided GT data and controversy maybe inevitable, and this is mainly due to (i) unclear boundary of some special editing effects and (ii) time-consuming and labor intensive efforts required in producing such GT.

B. Overall Performance and Evaluation

There are three measurements used in evaluating the results, i.e. recall and precision rate of cut detection, gradual transition, and overall performance. Each approach has up to 10 runs with various parameters to form a precision versus recall curve, and different approaches are compared accordingly. Also, a combined measurement of both precision and recall, F_1, is defined below to rank the performance of different algorithms.

$$F_1(recall, precision) = \frac{2recall \cdot precision}{recall + precision}. \quad (18)$$

In 2007 there are 35 teams registered for SBD task, and eventually only 15 teams have their results submitted as 128 runs. According to the report from TRECVID'07, the performance evaluation is summarized as follows [5]:

- For cut detection, our submission is the best and ranked as number one, where the recall/precision rates achieved are 97.3% and 98.2% respectively;
- For gradual transition detection, our submission is ranked as number six and further analysis is given later;
- Our submission is among the top three in terms of overall evaluation ;
- It is worth noting that our best results in terms of all four measures are delivered in one single run, while most of others have their best results achieved in different runs.

C. Performance Analysis in Details

To analyze the reasons that lead to the best results on cut detection and slightly weaker results in detecting gradual transitions, we provide further discussions as follows.

1) Effectiveness of Important Parameters

To illustrate the effectiveness of parameters, p_0 (for cut pre-filtering) and t_μ (for cut detection) are selected and evaluated in the precision-recall curves as shown in Fig. 1. For each parameter, three curves are plotted according to its three values. Ideally, the curve with a shape closest to the top-right corner presents the best result in terms of F_1 measurement. From Fig. 1, the following observations can be made:

- It can be seen that the overall performance is robust to the value change of parameters in a certain range where the precision-recall curves look similar to each other;
- A larger p_0 and a smaller t_μ can help to deliver a higher precision rate. In contrast, a higher recall rate can be obtained if we choose a smaller p_0 and a larger t_μ.

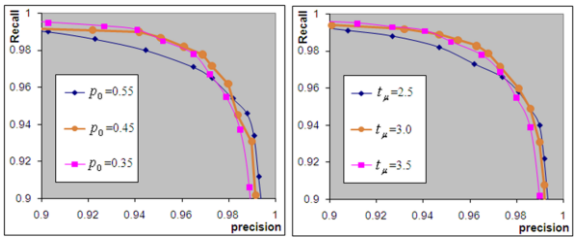

Figure 1. Precision-recall curves of p_0 and t_μ to show effectiveness of selected parameters for cut detection.

2) Effect of Phase-correlation for Post-processing

Further analysis reveals that the post-processing with phase-correlation helps to reduce about 3% of the false alarms in the improved precision rate while degraded recall rate is maintained at 0.2%. In other words, it contributes 1.2% towards the improvement of F_1 measurement.

3) Error Analysis

For abrupt cuts, missed detections are due to the fact that their content change is too small to be identified as any of the five cut categories as defined in Section III. For false alarms, they do present apparent visual differences introduced by strong motion or special editing effect. Some of them can be arguably regarded as a cut but as undefined in the GT.

Regarding gradual transitions, the missed detection is mainly caused by dissolve of small changes in intensity/color and irregular wipe effects. False alarms are primarily caused by motion or change of lighting conditions. Since gradual transitions are detected within each pair of cuts, errors in cut detection is part of the cause. Among all the missed detection and false alarms, some errors are actually due to the ambiguity or even the mistakes in defining shot boundaries inside the GT.

D. Complexity and Speed Analysis

Since it is difficult to theoretically analyze the complexity of these algorithms, a relative comparison can be made according to number of frames processed in one second. As our proposed algorithm operates entirely in compressed-domain, it delivers 123 frames per second (fps) in detecting shot changes from MPEG-1 videos. As a result, it can be well applied to online video segmentation and many other applications.

V. Conclusions

We provided a detailed description of the proposed method for SBD. In comparison with existing work and all other submissions for TRECVID'07, our algorithm features in: (i) compressed domain operation, providing 5 times as fast as real-time video play; (ii) extraction of content differential features and their optimized selection via AdaBoost; (iii) pre-filtering and mapping of the selected features to characterize cuts in five categories; (iv) establishment of corresponding likelihood for decision making in SBD; (v) statistics analysis and determination of key parameters; and finally (vi) validation of the detected results using phase-correlation. As a result, excellent performance results have been achieved, and yet its high speed processing also provides a great potential for many real-time video processing and content-based applications.

References

[1]. J. Yuan, H. Wang, L. Xiao, W. Zheng, J. Li, F. Lin, and B. Zhang, "A formal study of shot boundary detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 2, pp. 168-186, 2007.

[2]. C. Grana and R. Cucchiara, "Linear transition detection as a unified shot detection approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 4, pp. 483-489, 2007.

[3]. Q. Urhan, M. K. Gullu, and S. Erturk, "Modified phase-correlation based robust hard-cut detection with application to archive film," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 6, pp. 753-770, 2006.

[4]. C. Cotsaces, N. Nikolaidis, and I. Pitas, "Video shot detection and condensed representation: a review," *IEEE Signal Proc. Mag.*, vol. 23, no. 2, pp. 28-37, 2006.

[5]. National Institute of Standards and Technology (NIST) [Online]. Available: http://www-nlpir.nist.gov/projects/trecvid/

[6]. J. Bescos, "Real-time shot change detection over online MPEG-2 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 4, pp. 475-484, 2004.

[7]. H. Lu and Y. P. Tan, "An effective post-refinement method for shot boundary detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 11, pp. 1407-1421, 2005.

[8]. G. Boccignone, A. Chianese, V. Moscato, and A. Picariello, A.: "Foveated shot detection for video segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 3, pp. 365-377, 2005.

[9]. Z. Cernekova, I. Pitas, and C. Nikou, "Information theory-based shot cut/fade detection and video summarization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 82-91, 2006.

[10]. L.-Y. Duan, M. Xu, Q. Tian, C.-S. Xu, J. S. Jin, "A unified framework for semantic shot classification in sports video," *IEEE Trans. Multimedia*, vol. 7, no. 6, pp. 1066-1083, 2005.

[11]. H. Fang, J. Jiang, Y. Feng, "A fuzzy logic approach for detection of video shot boundaries," *Pattern Recogn.*, vol. 39, no. 11, pp. 2092-2100, 2006.

[12]. R. A. Joyce and B. Liu, "Temporal segmentation of video using frame and histogram space," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 130-40, 2006.

[13]. A. Hanjalic, "Shot boundary detection: unraveled and resolved?" *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 2, pp. 90-105, 2002.

[14]. S.-C. Pei, and Y.-Z. Chou, "Efficient MPEG compressed video analysis using macroblock type information," *IEEE Trans. Multimedia*, vol. 1, no. 4, pp. 321-333, 1999.

[15]. C.-L. Huang and B.-Y. Liao, "A robust scene-change detection method for video segmentation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 12, pp. 1281-1288, 2001.

[16]. Y. Freund and R. E. Schapire, "A decision-theoretic generalization of online learning and an application to boosting," *J. Computer and System Sciences*, vol. 55, no. 1, pp. 119-139, 1997.