A Novel Technique for Automatic Abrupt Shot Transition Detection

Malay Kumar Kundu
Machine Intelligence Unit
Indian Statistical Institute
Kolkata, India
e-mail: malay@isical.ac.in

Jaydeb Mondal
Machine Intelligence Unit
Indian Statistical Institute
Kolkata, India
e-mail: jaydebmondal2011@gmail.com

Abstract—**In this article, we propose a novel automatic thresholding method for the detection of abrupt shot transition (AST). based on statistics of the pixel difference value of the consecutive frames, from a given video sequence. An outlier removal and a false alarm elimination scheme are also introduced to counteract the disturbances from illumination variations, object and camera movements. Experimental results and comparisons with state-of-the-art SBD schemes show the effectiveness of the proposed method and having average superior accuracy.**

Keywords-component; Difference value; outlier removal; threshold computation; false alarm elimination;

I. Introduction

Video content processing and analysis have become an important research topic in recent times, due to its wide range of applications in different areas which include education/distance communication, online medical diagnostics, video on demand and surveillance etc. Partitioning of video into shots is the first step toward VCA and CBVR. Efficient and effective partitioning of video data provides a basis for nearly all types of video abstraction and higher level video segmentation.

A video shot can be defined as a set of consecutive frames taken from a single non-stop camera operation. Shot-boundary is the transition between two shots. The shot transitions are of two types, abrupt shot transitions (AST) and gradual shot transitions (GST). AST is resulted from a sudden change of the content between two consecutive frames due to the restarting or stopping of the camera as per directives. On the other hand, GST is a gradual change that spans for several consecutive frames, and is used to maintain the continuity between two shots. GST is further subdivided into three categories, fade in/fade out, dissolve and wipe.

Detection of shot boundary requires an appropriate feature and distance metric to define the changes between two consecutive frames and a suitable threshold to determine whether this difference is acceptable for the frame to be identified as a shot boundary or not. The value of threshold may be either pre-determined or computed from the video characteristics itself. Various methods for SBD have been developed in the past [1, 3, 4, 5, 8, 10, 11]. The feature and distance metric used for ideal SBD should be invariant to the disturbances from illumination variation, object movements as well as camera movements [4]. However, the existing features and distance metrics are more or less suffer from a major problem due to moving objects of either large size or of high speed and result in the detection of false scene cut. To reduce the influence of object movements called outlier, various methods exist in the literature [6, 11]. In this article, we propose a new robust technique based on the statistical properties of the difference values to remove this shortcoming. Another most important aspect of SBD is the determination of suitable threshold, which must be automatic. Value of the threshold should be computed from the characteristics of a given video sequence which means it should be more or less sequence independent. However, most of the works reported in the literature are either rely on heuristically chosen global thresholds or semi-automatic due to the presence of experimentally given parameters for accuracy [2, 9, 10, 11]. To overcome from this shortcoming, we also propose an automatic thresholding approach for the detection of AST based on statistical properties of the corresponding frame-to-frame differences. Here value of the threshold is computed adaptively in an iterative fashion from the entire video sequence. The proposed scheme is free from user defined parameters, independent of video sequence, robust and computationally less expensive. A false alarm elimination scheme is also suggested based on the inherent characteristics of video to reduce the false AST while keeping the correct ones.

The rest of the paper is organized as follows. Section 2 describes the proposed method. In Section 3, Experimental results and comparisons are presented. Finally Conclusion and Future Work are discussed in Section 4.

II. Proposed method

The basic requirements of a SBD scheme are to define the changes between two consecutive frames and a suitable threshold to determine the presence of shot boundary. The overall structure of a SBD scheme is shown by the block diagram in figure1. The details of the various components of our proposed scheme for the detection of AST are described in the following subsections.

A. Computation of DifferenceValue

At first, we measure the dissimilarity between two consecutive frames of a given video sequence, for the detection of AST. In our algorithm we use conventional [9, 12] pixel based comparison approach to compute difference value. To do this, each frame is divided into 16 equal sized non-overlapping blocks of size $M \times N$.

The difference value $D(i, i+1)$ between two consecutive frames, is defined as follows

$$D(i,i+1) = \frac{\sum_{k=1}^{r} DB(i,i+1,k) - \sum_{k=1}^{l} DB(i,i+1,k)}{r-l} \quad (1)$$

$$DB(i,i+1,k) = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} |f_{i,k}(x,y) - f_{i+1,k}(x,y)|}{M \times N} \quad (2)$$

where $DB(i,i+1,k)$ is the difference value for the k^{th} block of two consecutive frames f_i and f_{i+1}. Total number of blocks is represented by r and l is the total number of outliers which is described in the following section. The expression $f_{i,k}(x,y)$ denotes the intensity value at co-ordinate (x, y) for k^{th} block of the frame f_i.

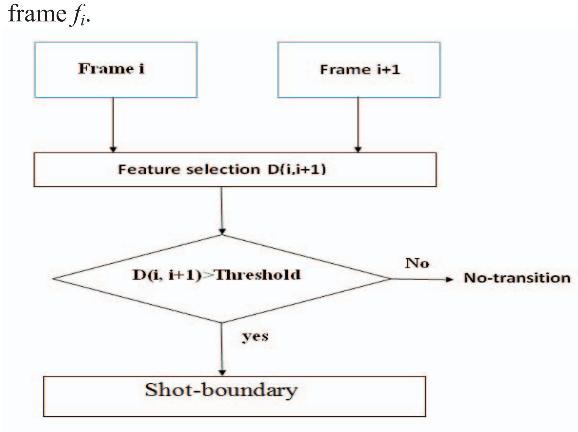

Figure 1: Block diagram of SBD

B. Outlier Removal

To detect shot-boundary correctly, it is necessary to reduce the effect of object motion. Due to the movement of objects, abrupt changes occur in the difference values which are considered as outlier. To remove this shortcoming, *Nagasaka et al*. break the entire frame into 16 equal sized non-overlapping blocks and instead of taking the total difference between two consecutive frames; the differences of the corresponding blocks are taken. This produces 16 $DB(i,i+1,k)$ and the detection of AST is based on the eight lowest $DB(i,i+1,k)$ values [6]. This is based on the assumption that such disturbances could not affect more than half of the entire frame. *Zhang et al.* have proposed an improvement over this by taking the sum of the median eight $DB(i,i+1,k)$ values instead of taking the lowest eight ones [11].

Here, we suggest an alternative approach where we calculate mean (μ) and standard deviation (σ) of the $DB(i,i+1,k)$ values and only those values are taken which are within the range of $(\mu \pm 1.5\sigma)$. The basic motivation behind such an approach is that the difference values i.e. $DB(i,i+1,k)$ follow Gaussian distribution and more than 80% $DB(i,i+1,k)$ space lies within this range. The values outside of this range are considered as outlier and should not be taken into the computation of AST. The mean (μ) and standard deviation (σ) of $DB(i,i+1,k)$ are calculated as follows and the total process for calculating $D(i,i+1)$ is described in the algorithm 1.

$$\mu = \frac{\sum_{k=1}^{r} DB(i,i+1,k)}{r} \quad (3)$$

$$\sigma = \sqrt{\frac{\sum_{k=1}^{r} (DB(i,i+1,k) - \mu)^2}{r}} \quad (4)$$

Algorithm 1:

1. Divide each frame into r number of equal sized non-overlapping blocks. In our experiment we take r=16.
2. Calculate $DB(i,i+1,k)$ between two consecutive frames f_i and f_{i+1}, for k=1,2…r .
3. Compute mean (μ) and standard deviation (σ) of $DB(i,i+1,k)$.
4. If any value of $DB(i,i+1,k)$ is beyond $(\mu \pm 1.5\sigma)$, remove it. We call it *outlier*.
5. Compute mean of $DB(i,i+1,k)$, for the remaining blocks. This is the desired difference value $D(i,i+1)$, between two consecutive frames f_i and f_{i+1}.
6. Repeat steps 1-5, for all $i=1,2,\dots N-1$, where N is the total number of frames in a video.

C. Computation of Automatic Threshold for AST

Another key issue for SBD is the selection of a suitable threshold, which actually controls the overall accuracy of the system. Moreover, the value of threshold should be computed from the given video sequence in an adaptive manner.

There is reasonable number of threshold selection algorithm available in the literature which includes application in image segmentation. One such method found in [7] has been utilized in the proposed method for the detection of AST.

Suppose the histogram of an image is bi-modal, i.e., the image contains only two primary regions, object and background. Assuming the distribution of the pixels in each region follow Gaussian distribution, the probability of a pixel value is given by

$$p(x) = \frac{P_1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{\sigma_1^2}} + \frac{P_2}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x-\mu_2)^2}{\sigma_2^2}} \quad (5)$$

Where P_1 and P_2 are the priori probabilities of object and background pixels, μ_1 and μ_2 are the mean of object and background pixels, σ_1 and σ_2 are the standard deviations of object and background pixels, respectively.

Optimal threshold can be found by minimizing the following error function:

$$E(T) = P_2 \int_{-\alpha}^{T} p_2(x)dx + P_1 \int_{T}^{\alpha} p_1(x)dx \quad (6)$$

with respect to threshold T. If $\sigma_1=\sigma_2=\sigma$, then the value of optimal threshold is given by minimizing $E(T)$ and T becomes

$$T = \frac{\mu_1+\mu_2}{2} + \frac{\sigma^2}{\mu_1-\mu_2}\ln\frac{P_2}{P_1} \tag{7}$$

If $P_1=P_2$, then value of the optimal threshold becomes simply the average of two class means.

Based on this basic idea, we have proposed an iterative method for computation of optimal threshold for the detection of AST. The initial threshold is set by the mean of the corresponding frame to frame difference value and is defined as

$$T_{initial} = \frac{\sum_{i=1}^{N-1} D(i,i+1)}{N-1} \tag{8}$$

where, N is the total number of frames in a given video sequence. Depending on this threshold, $D(i, i+1)$ is classified into two classes, AST and non-AST where,

$$D(i,i+1) = \begin{cases} AST & if \;\; D(i,i+1) > T_{initial} \quad 1 \le i \le N-1 \\ non-AST & else \quad\quad\quad\quad\quad\quad\quad 1 \le i \le N-1 \end{cases} \tag{9}$$

Now, let AST contain X number of $D(i, i+1)$ and non-AST contain Y number of $D(i, i+1)$. The mean value of both classes (μ_1, μ_2) are computed using the following equation

$$\mu_1 = \frac{\sum_{i=1}^{X} D(i,i+1)}{X} \qquad D(i,i+1) \in AST \tag{10}$$

$$\mu_2 = \frac{\sum_{i=1}^{Y} D(i,i+1)}{Y} \qquad D(i,i+1) \in non-AST \tag{11}$$

Then, the new threshold becomes

$$T_{new} = \frac{\mu_1 + \mu_2}{2} \tag{12}$$

If it is found that the new value of the threshold T_{new} is same as $T_{initial}$, then the initial value of the threshold is considered as the optimal threshold value. Otherwise we proceed for the computation of new threshold value until the value of $T_{initial}$ and T_{new} become same. In that case, T_{new} is considered as $T_{initial}$ and the procedure is repeated for the convergence of the $T_{initial}$ and T_{new}. The method is described by the following block-diagram in figure 2. The symbols used in the block diagram are as: T_1 is the initial threshold, T_2 is the new threshold and $D(i,i+1)$ represents, the difference value between two consecutive frames f_i and $fi+1$, *for* $i=1,2…N-1$, where N is the total number of frames in a given video sequence.

D. False Alarm Elimination

False alarm is the one which is detected by the SBD technique as a true scene-cut, although it is not. Major source of it, is the presence of illumination variation, large object movements and camera movements such as panning, zooming, tilting, etc. as they produce high value like abrupt transitions.

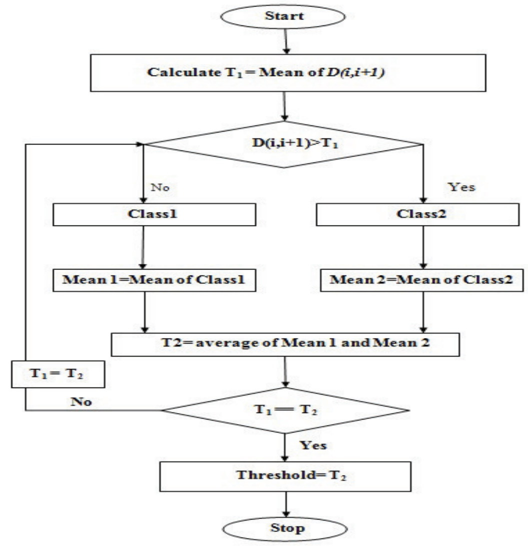

Figure 2: Block diagram of Threshold Computation

From the characteristics of human visual system (HVS) for the perception of change, the duration of a shot should be at least 1/3 of a second.. So it is logical to assume that the frame difference between two perceivable AST should be as least (r/3), where r is the frame rate per second.

Let there be N number of frames and M number of AST in a given video sequence where, N >> M. Also let T_i represents the frame number where AST occurs. Then, from the above fact, it can be written as

$$(T_i-T_{i-1}) \ge (r/3), \text{ for all } i = 1, 2,\ldots,M \tag{13}$$

If, r= 30 frame/second, then above equation becomes

$$(T_i-T_{i-1}) \ge 10, \text{ for all } i = 1, 2,\ldots,M \tag{14}$$

This fact is used in the proposed scheme to eliminate false alarms from the true AST.

III. Experimental Results

The proposed algorithm has been tested on a variety of video sequences which contain sufficient number of ASB along with large amount of object and camera movements. These videos are taken from <TRECVID 2001 test database- http://www.open-video.org> which includes abrupt shot transitions as well as different types of gradual shot transitions of various length. The datasets also contain the object motion, camera motion and illumination variations of various degree and in adequate numbers which makes the detection process even harder. Experimental results and comparisons with some recently published articles are given in Table 1, in terms of Recall (P), Precision (P) and F-measure (F) found in [5]. Higher value indicates better performance. The best results are shown by bold digits which indicate the average best performance of our proposed scheme. The method in [1] give best results in terms of R measure for the sequence anni005 and Nad53 while the performance of method in [5] is best in terms of P and F measures for the sequence anni005 and anni006.

The result of our proposed scheme for few frames of the

TABLE I. PERFORMANCE OF VARIOUS SBD SYSTEMS

Sequence	Number of shots	Adjeroh et. al. [1]			Mohanta et. al. [5]			Proposed Scheme with false alarms rejections		
		P	R	F	P	R	F	P	R	F
anni005	38	0.87	**0.91**	0.89	**0.97**	0.90	**0.93**	**0.97**	0.89	**0.93**
anni006	41	0.82	0.89	0.85	**0.95**	0.93	**0.94**	0.84	**0.98**	0.91
anni009	38	0.87	0.93	0.90	0.94	0.92	0.93	**0.95**	**0.95**	**0.95**
BOR08	197	0.86	0.91	0.88	0.92	0.93	0.92	**0.95**	**0.95**	**0.95**
Nad53	83	0.81	**0.97**	0.88	0.86	0.93	0.89	**0.96**	0.90	**0.93**
	Average	0.85	0.92	0.88	0.92	0.92	0.92	**0.94**	**0.93**	**0.94**

video sequence anni006 is shown in figure 3. The detected frames where AST occur are marked by the red boxes. In figure 4, we show the results in the same sequence ann006, using the outlier removal scheme proposed in [11]. The results are shown by the blue boxes. It gives 5 correctly detected frames and 3 falsely detected frames which shows the robustness of our proposed outlier removal scheme over the method used in [11].

IV. CONCLUSION

Partitioning video data into its constituent shots is essential for VCA and CBVR. We have proposed a novel scheme for the detection of AST. Although, low level pixel features has some shortcomings in the field of CBVR. it is simple to implement and computationally less expensive and still in use [12]. Experimental results and comparison with two recently reported articles show the robustness of our proposed scheme against variety of video sequences. In future we will experiment this technique using more robust features and also try to identify different types of gradual transitions like fade in/fade out, dissolve and wipe.

References

[1] D. Adjeroh, M. C. Lee, N. Banda and U. Kandaswamy, “Adaptive edge-oriented shot boundary detection,” EURASHIP J. Image Video Process., vol. 2009, June 2009.

[2] F. Arman, A. Hsu and M. Chiu, “Feature management for large video databases,” In Proc. IS&T/SPIE storage and retrieval for image and video databases, pages 2-12, 1993.

[3] R. Kasturi, S. Strayer, J. Yu and U. Gargi, “Performance characterisation of video-shot-change detection methods,” IEEE Trans. CSVT, vol. 10, pp. 1–13, 2000.

[4] A. Hanjalic, “Shot-boundary detection: unravelled and resolved?,” IEEE Trans. CSVT, vol. 12, pp. 90-105, 2002.

[5] P. P. Mohanta, S. K. Saha and B. Chanda, “A model-based shot boundary detection technique using frame transition parameters,” IEEE Trans. Multimedia, vol. 14, pp. 223–233, 2012.

[6] A. Nagasaka and Y. Tanaka, “ Automatic video indexing and full video search for object appearences,” In proc. of the second working conference on visual database systems , pp. 119-133, 1991.

[7] A. Rosenfeld and A. C. Kak, Digital Picture Processing, Academic press, 1982.

[8] K. K. Warhade, K. K. Merchant and U. B. Desai, “Performance evaluation of shot boundary detection metrics in the presence of object and camera motion,” IETE Journal of Research, vol. 57, pp. 461-466, 2011.

[9] B. L. Yeo and B. Liu, “Rapid scene analysis on compressed video,” IEEE Trans. CSVT, vol. 5, pp. 533-544, 1995.

[10] Y. Yusoff, W. Christmas and J. Kittler, “Video shot cut detection using adaptive thresholding,” In Proc of the International Conference on British Machine Vision, 2000.

[11] HJ. Zhang, A. Kankanhalli and W. Smolier, “Automatic partitioning of full-motion video,” Multimedia Systems, vol 1,pp. 10-28, 1993.

[12] C. Cuevas and N. Garcia, "Temporal Segmentation Tool for High-Quality Real-Time Video Editing Software," IEEE Trans. on Consumer Electronics, vol. 58, No. 3, 2012.

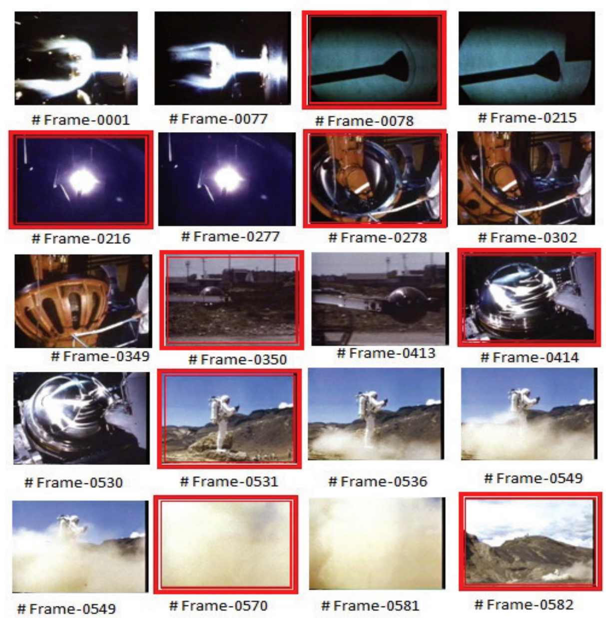

Figure 3. Result of the proposed AST scheme. The frames marked by red boxes are the output frames.

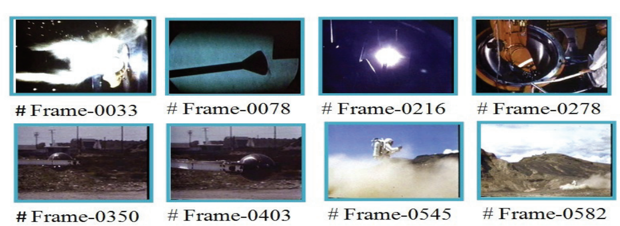

Figure 4. Result of the AST scheme using outlier removal method proposed in [11].