Video Shot Boundary Detection Algorithm

Kyong-Cheol Ko[1], Young- Min Cheon[1], Gye-Young Kim[1], Hyung –Il Choi[1], Seong-Yoon Shin[2], and Yang-Won Rhee [2]

[1, *] Information Media Technology Research Institute, Soongsil University
1-1, Sangdo-Dong, Dongjak-Gu, Seoul 156-743, South Korea
{hic, gykim11}@ssu.ac.kr
[2] Department of Computer Information Science, Kunsan National University
68, Miryong-dong, Kunsan, Chonbuk 573-701, South Korea
{roadkkc, ywrhee}@kunsan.ac.kr

Abstract. We present a newly developed algorithm for automatically segmenting videos into basic shot units. A basic shot unit can be understood as an unbroken sequence of frames taken from one camera. At first we calculate the frame difference by using the local histogram comparison, and then we dynamically scale the frame difference by Log-formula to compress and enhance the frame difference. Finally we detect the shot boundaries by the newly proposed shot boundary detection algorithm which it is more robust to camera or object motion, and many flashlight events. The proposed algorithms are tested on the various video types and experimental results show that the proposed algorithm are effective and reliably detects shot boundaries.

1 Introduction

There are many shot detection methods already proposed in past decades [1], [2]. The common way for shot detection is to evaluate the difference value between consecutive frames represented by a given feature. Although reasonable accuracy can be achieved, there are still problem that limit the robustness of these algorithms [4].

One of the common problems in robust shot detection results from the fact there are many flashlights in news video, which often introduce false detection of shot boundaries. Only some simple solutions to this problem have been proposed in [2], [3]. There are main limitations are that they assume the flashlights just occur during one frame or limited window region. In real world, such as news video, there are many flashlight events occur during a period of time and influence multiple consecutive frames.

Another problem that has not been solved very effectively well is threshold selection when comparison changes between two frames. Most of the existing methods use global pre-defined thresholds, or simple local window based adaptive threshold.

Global threshold is definitely not efficient since the video property could change dramatically when content changes, and it is often impossible to find a universal optimal threshold method also has its limitation because in some situation the local statistics are polluted by strong noises such as big motions or flashlights.

[*] This work was supported by the Korea research Foundation Grant (KRF-2006-005-J03801).

P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 388 – 396, 2006.

The objective of this paper are: 1) to provide the metrics that are robust to camera and object motion, and enough spatial information is retained, 2) to provide the scaled frame difference that are dynamically compressed by log formula and it is more convenient to decide the threshold, 3) to propose a new shot boundary detection algorithm that are robust to camera operation or fast object movement, flashlight events.

The rest of this paper is organized as follows. In the next section 2, we provide a proposed algorithm that gives a detail description of the three new algorithms. Section 3 presents experimental results, and we conclude this paper and discuss the future work in Section 4.

2 The Proposed Algorithm

Firstly, we denote the metrics to extract the frame difference from consecutive frames. And we scale the frame difference by log formula which makes more dynamically robust to any camera or object motion, and many flashlight events. Finally we propose the new shot boundary detection algorithm. Our proposed algorithm works in real time video stream and not sensitive to various video types.

Throughout this paper, we shall treat a shot, defined as a continuous sequence of frames recorded from a single camera, as a fundamental unit in a video sequence.

2.1 Metrics in Shot Detection

To segment the video into shot units, we should first define suitable metrics to extract frame difference; so that a shot boundary is declared whenever that metric exceed a given threshold.

We use the local histogram comparison that are more robust to camera and object motion, and enough spatial information is retained to produce more accurate results [5], [6].

The local histogram comparison metrics are defined as:

$$d(f_i, f_j) = \sum_{bl=1}^{m} d_{x^2}(f_i, f_j, bl) \tag{1}$$

$$\begin{aligned} d_{x^2}(f_i, f_j, bl) = \sum_{k=1}^{N-1} \Bigg(& \frac{(H_i^r(k) - H_j^r(k))^2}{\max(H_i^r(k), H_j^r(k))} \times \alpha \\ & + \frac{(H_i^g(k) - H_j^g(k))^2}{\max(H_i^g(k), H_j^g(k))} \times \beta \\ & + \frac{(H_i^b(k) - H_j^b(k))^2}{\max(H_i^b(k), H_j^b(k))} \times \gamma \Bigg), \end{aligned} \tag{2}$$

where m is the total number of the blocks, and $H_i^r(k)$ denotes the histogram difference at gray level k for the block bl of i'th frame in red channels. α, β and γ are

constants and, according to NTSC standard, we set these constants to 0.299, 0.587, and 0.114, respectively.

The best frame difference can be obtained by breaking the frame into 16 equal sized regions, using weighted x^2-test on color histograms for these regions and discarding the largest differences to reduce the effects of noise, object and camera movements.

2.2 Scaled Frame Difference

Most of video segmentation algorithms rely on suitable threshold of similarities between consecutive frames. However, the thresholds are highly sensitive to the type of input video. This drawback can be overcome by the scaled frame difference.

The scale of frame difference is performed by Log-formula which makes more dynamically compressed frame difference and Log-formula was referenced by digital image processing which was used to image enhancement.

The proposed Log –formula defined as:

$$d_{\log} = c \times \log(1 + d^2)$$
$$c = \frac{\max(d_{\log})}{\max(\log(1 + d^2))} \tag{3}$$

Where d is the frame difference extracted from equation (1) and c is the constant calculated from d.

Figure 1 shows the distribution of total frame differences extracted from news video.

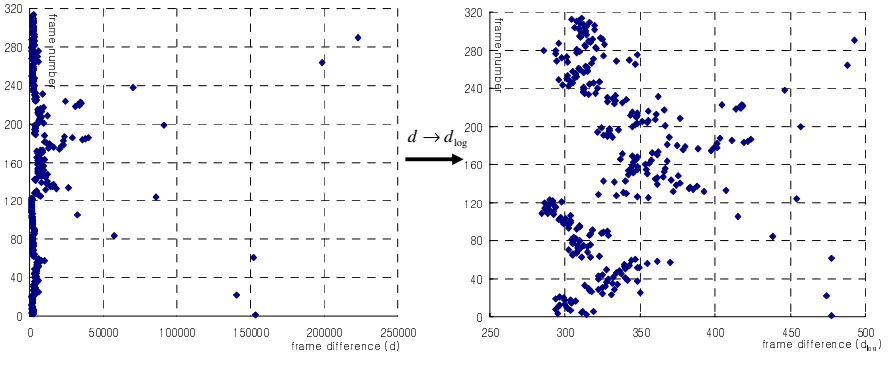


Fig. 1. Distribution of all frame difference 'd' and 'd_{log}'

Distribution of all frame differences d_{log} has widely spread difference values in a scaled region than d and each difference values are enhanced and concatenated each other more closely. So if we apply the simple shot cut rules, we can detect the shot boundaries only using the frame difference.

Table 1 shows the max (maximum), min (minimum), ave (average), and stdev (standard-deviation) represented from three video types(news, sports, adv.). Each of the frame differences d and d_{log} are calculated from the given equations (1) and equation (2).

Table 1. Comparison of difference values 'd' and 'd_{log}'

videos	Max.		Min.		Ave.		Stdev.	
	d	d_{log}	d	d_{log}	d	d_{log}	d	d_{log}
News	223057.9	492.6	1234.2	284.7	9191.3	334.2	23778	38.2
Sports	212168.3	490.6	703.2	262.2	3740.2	308.1	13380.3	25.1
Adv.	216088.1	491.3	3993.2	331.7	26494.2	391.6	30614.5	33.3

As mentioned above it, scaled difference values are more robust and reliable to detect the shot boundaries and are convenient to select the global threshold.

Figure 2 shows the normal graph of Table 1. Scaled frame difference d_{log} are dynamically compressed and more normally distributed under the scaled region than d.

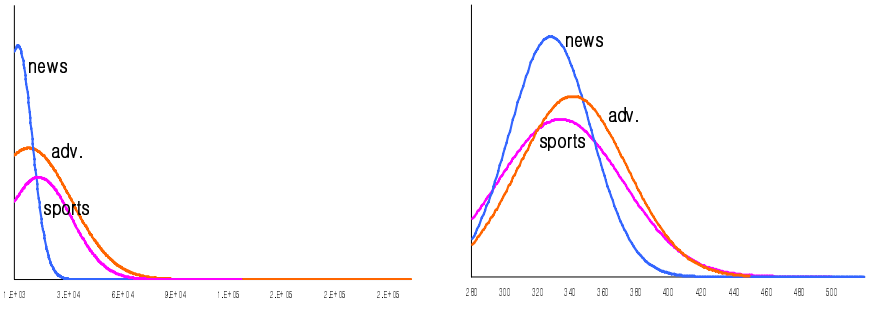


Fig. 2. Normal Distribution of frame difference 'd' and scaled frame difference 'd_{log}'

2.3 Shot Boundary Detection Algorithm

Shot boundary detection is usually the first step in generic video processing. A shot represents a sequence of frames captured from a unique and continuous record from a camera. Therefore adjacent frames of the same shot exhibit temporal continuity. Both the real shot cut and the abrupt cut could cause a great change in frame difference because of the special situations such as flashlight events, sudden lightening variances, and fast camera motion, or large object movements. So each shot corresponds to a single continuous action and no change of content can be detected inside a shot. Change of contents always happen at the boundary between two shots. Partitioning a video sequence into shots is also useful for video summarization and indexing.

We define shot boundary detection algorithm based on the temporal property of shot cut and abrupt cut. If the scaled frame difference of consecutive frames is larger than a max-threshold (th_{max}), and its neighboring difference value of frame difference is larger than a k-threshold (k_{gloval}), and also its Euclidian distance is satisfied with global-threshold (th_{gloval}) , then the shot cut is detected by shot boundary detection algorithm.

Figure 3 shows the proposed shot boundary detection algorithm more details.

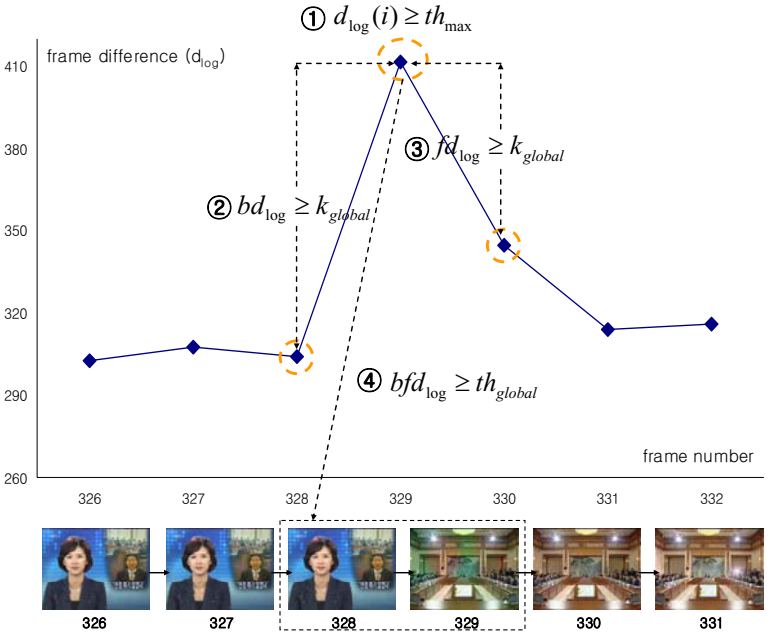


Fig. 3. The illustration graph of proposed shot boundary detection algorithm

As shown in Figure 3, the shot boundary detection algorithm can be summarized as follows:

Step 1. At first, if the scaled frame difference $d_{log}(i)$ is larger than a max-threshold th_{max} then the current frame is selected to candidate shot frame,

$$d_{\log}(i) \geq th_{\max}$$

Step 2. And we calculate the newly defined difference value $bd_{log}(i)$, $fd_{log}(i)$ as follows:

$$\begin{aligned} bd_{\log}(i) &= \left| d_{\log}(i) - d_{\log}(i-1) \right|, \\ fd_{\log}(i) &= \left| d_{\log}(i+1) - d_{\log}(i) \right| \end{aligned} \tag{4}$$

The calculated difference value $bd_{log}(i)$, $fd_{log}(i)$ must be larger than a k-threshold k_{gloval} .

$$bd_{\log}(i) \geq k_{global} \;\&\&\; fd_{\log}(i) \geq k_{global}$$

Setp 3. Finally, the Euclidean distance of each calculated frame difference value bfd_{log} is defined as:

$$bfd_{\log}(i) = \sqrt{(bd_{\log}(i))^2 + (fd_{\log}(i))^2} \tag{5}$$

And it must larger than a global-threshold th_{gloval} .

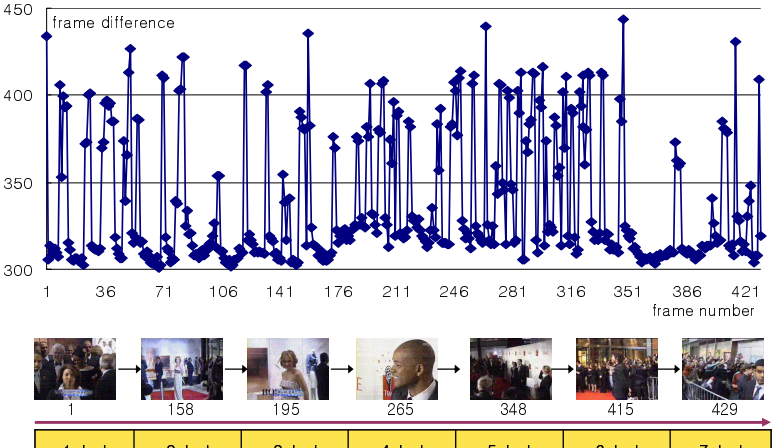
$$bfd_{\log}(i) \geq th_{global}$$

Step 1 is the basic step to check the candidate shot frame. Most of shot frame has a big difference value and we heuristically determine the max-threshold th_{max} from scaled frame difference. In experiments results, the determined max-threshold th_{max} was reliable and robust than previous approaches.

Step 2 is to check whether the current frame is shot cut or abrupt cut. A real shot cut has enough distance between bd_{log} and fd_{log} but abrupt cut has small distance each other. If the distance bd_{log} and fd_{log} is smaller than k-threshold k_{gloval}, then current frame is classified as abrupt cut.

Step 3 is to check the sensibility over the set of threshold bd_{log} and $fd_{log.}$.

Figure 4 shows the illustration of the proposed shot boundary detection algorithm.



	1shot	2shot	3shot	4shot	5shot	6shot	7shot
$d_{\log}(i)$	433.7	435.5	406.9	439.6	443.7	430.7	409.3
$bd_{\log}(i)$	128.5	52.6	74.7	113.7	118.7	100.4	90.3
$fd_{\log}(i)$	433.7	121.9	30.8	123.7	58.9	122.5	100.9
$bfd_{\log}(i)$	452.4	132.8	80.8	168.0	132.6	158.4	135.3

Fig. 4. Distribution of remaining number of frames by the proposed algorithm

As shown in Figure 4, the diagram is the scaled frame difference of consecutive frames in sequence 'interview videos' which has a lot of flashlight events. Detected shot cut frame, and used difference value of each frame difference is shown in Figure 4.

All possible shot cut is detected and flashlight is eliminated in reliable.

3 Experimental Results

We evaluate the performance of our proposed method with DirectX 8.1 SDK, MS-Visual C++ 6.0 on Windows XP.

The proposed method has been tested on several video sequences such as news, interviews, and commercials videos that have a lot of scene changes occurs, as shown in table1. Each video sequence has the various types digitized in 320*240 resolutions at 30frames/sec.

Table 2. Description of the Videos in the experiment dataset

Videos	# of frames	# of abrupt cuts		# of shot cuts (ground truth)
		fast object and camera motion or etc.	flashlights	
news1	2772	2	31	26
news2	2665	2	55	19
Choice	2975	1	14	21
soccer	2167	3	6	22
Flash1	2578	1	52	15
Movie1	1175	2	25	11
Golf	665	12	0	19
Wine	3096	10	0	30

In table 2, 'News2' or 'Flash1' videos contain many flashlights events and 'Golf' or 'Wine' videos contain fast object and camera motions.

We manually identify the ground truth by a user with frame accuracy. In our experiments, the shot cut detection results are compared with the ground truth in terms of precision and recall. Assume N is the ground truth number of shot cuts, M is the number of missed cuts and F is the number of false alarms, the recall and precision are defined as follows:

$$\mathrm{Re}\,call = \frac{N-M}{N}$$
$$\mathrm{Pr}\,ecision = \frac{N-M}{N-M+F} \tag{3}$$

These two measures are both important. We certainly do not want to miss any critical shot changes. On the other hand, too many false alarms will compromise the efficiency of video segmentation.

Table 3 indicates that proposed algorithm can detect not only abrupt cuts but also shot cut with satisfactory accuracy. Approximately 97% of fast camera transitions, fast object motions and flashlight events are detected. The missed abrupt cuts mainly results from the fact that the frame differences between consecutive frames are lower than the given threshold.

Table 3. Experiment Results

Videos	# of abrupt cuts				# of shot cuts			
	# of false	# of missed	recall	precision	# of false	# of missed	recall	precision
news1	0	2	94%	100%	2	0	100%	93%
news2	2	0	100%	97%	0	1	94%	100%
Choice	0	0	100%	100%	0	0	100%	100%
soccer	0	0	100%	100%	0	0	100%	100%
Flash1	0	0	100%	100%	1	1	93%	93%
Movie1	0	3	89%	100%	1	2	82%	90%
Golf	0	1	92%	100%	0	1	95%	100%
Wine	0	0	100%	100%	0	1	97%	100%
TOTAL	2	6	97%	99%	3	5	95%	97%

4 Conclusion

This paper has presented an effective shot boundary detection algorithm, which focus on three difficult problems solutions: To provide the metrics that are robust to camera and object motion, and enough spatial information is retained. To provide the scaled frame difference that are dynamically compressed by log formula and it is more convenient to decide the threshold. To propose a new shot boundary detection algorithm that are robust to camera operation or fast object movement, flashlight events. Experiments show that the proposed algorithm is promising.

However the automatic video partition is still a very challenging research problem especially for detecting gradual transitions or camera fabrication, special events and so on. Further work is still needed.

References

1. I. Koprinska and S. Carrato, “Temporal Video Segmentation: A Survey,” Signal Processing Image Communication, Elsevier Science 2001.
2. G. Ananger, T.D.C. Little, “A survey of technologies for parsing and indexing digital video,” Journal of Visual Communication and Image Representation, 1996, pp. 28-43.
3. D. Zhang, W. Qi, H. J. Zhang, “A News Shot Boundary Detection Algorithm,” IEEE Pacific Rim Conference on Multimedia, pp. 63-70, 2001.
4. U. Gargi, R. Kasturi, and S. H. Strayer, “Performance Characterization of Video-Shot-Change Detection Methods,” IEEE transaction on circuits and systems for video technology, Vol. 10, No. 1, Feb. 2000.
5. A. Nagasaka, Y. Tanaka, “Automatic video indexing and full-video search for object appearances,” in Visual Database Systems II, pp. 113-127, Elsevier, 1995.
6. K. C. Ko, Y. W. Rhee, “Scene Change Detection using the Chi-test and Automated Threshold Decision Algorithm,” ICCSA06, Vol. 3983, pp. 1060-1069, 2006.

7. C. L. Huang and B. Y. Liao, "A Robust Scene Change Detection Method for Video Segmentation," IEEE Trans on CSVT, Vol. 11. No. 12, pp. 1281-1288, December 2001.
8. H. Zhang, A. Kankamhalli, and S. Smoliar, "Automatic partitioning of full-motion video," ACM Multimedia Systems, New York: ACM Press, Vol. 1, 1993, pp. 10-28.
9. U. Gragi, R. Kasturi, S. Antani, "Evaluation of video sequence indexing and hierarchical video indexing," in: Proc. SPIE Conf. Storage and Retrieval in Image and Video Databases, 1995, pp. 1522-1530.
10. Gonzalez, "Digital Image Processing 2/E," Prentice-Hall, 2002.
11. R. M. Ford, C. Robson, D. Temple, M. Gerlach, "Metrics for shot boundary detection in digital video sequences," Multimedia Systems 8: 37-46, 2000.
12. A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," IEEE Trans. On Image Processing, Vol. 12, No. 7, pp. 796-807, July 2003.
13. C. L. Huang and B. Y. Liao, "A Robust Scene Change Detection Method for Video Segmentation," IEEE Trans. Circuit System. Video Technology, Vol. 11, No. 12, December 2001.