SHOT DETECTION IN VIDEO SEQUENCES USING ENTROPY-BASED METRICS

Z. Černeková *C. Nikou* *I. Pitas*

Department of Informatics
University of Thessaloniki
Box 451, Thessaloniki 540 06
GREECE
(E-mail: (zuzana,nikou,pitas)@zeus.csd.auth.gr)

ABSTRACT

A new method for detecting shot boundaries in video sequences using metrics based on information theory is proposed. The method relies on the mutual information and the joint entropy between frames and can detect cuts, fade-ins and fade-outs. The detection technique was tested on TV video sequences having different types of shots and significant object and camera motion inside the shots. It was favorably compared to other recently proposed shot cut detection techniques. The method is proven to detect both fades and abrupt cuts very effectively.

1. INTRODUCTION

The indexing and retrieval of digital video is an active research area. Shot boundary detection is an important task in managing video databases due to the structure of video into units (shots) for indexing, browsing, searching, summarization and other content-based operations.

Early work on shot detection mainly focused on abrupt cuts. A comparison of existing methods is presented in [1, 2]. The standard color histogram-based algorithm and its variations are widely used for detecting cuts [3, 4]. These algorithms detect changes between the frames by comparing the differences of the consecutive video frame intensity histograms.

Gradual transitions such as dissolves, fade-ins, fade-outs and wipes are examined in [5, 6, 7]. These transitions are generally more difficult to be detected, due to camera and object motion within a shot. A *fade* is a transition of gradual diminishing (fade-out) or heightening (fade-in) of visual intensity. Fades are widely used in TV and their appearance generally signals a shot change. Therefore, their detection is a very powerful tool for shot classification and story summarization. Existing techniques for fade detection proposed in the literature relies on twin thresholding [8] or grey level statistics [1] and have a relatively high false detection rate. Moreover, standard methods based on histograms [9], even if they correctly detect scene changes, they cannot distinguish between fades and other transitions.

In this paper, we propose a new approach for shot boundary detection in the uncompressed image domain, based on the mutual information and the joint entropy between consecutive frames. The mutual information is a measure of transported information from one frame to another. Mutual information is used for detecting abrupt cuts, where the image intensity or color is abruptly changed.

In the case of a fade-out, where visual intensity is usually decreasing to a black image, the decreasing inter-frame joint entropy is used as a metric. The opposite stands for a fade-in. The application of these entropy-based techniques for shot cut detection was experimentally proven to be very efficient producing false acceptance rates and false rejection rates very close to zero.

The proposed method was also favorably compared to other recently proposed shot cut detection techniques. At first, we compared the joint entropy metric to the technique relying on the average frame grey level descent (AD) for fade detection [1]. Finally, we compared our algorithm to the technique proposed in [9]. This approach combines two shot boundary detection schemes based on color frame differences and color vector histogram differences between successive frames.

2. SHOT DETECTION

In our approach, the mutual information and the joint entropy [10, 11] between two successive frames is calculated separately for each of the RGB components. Let us consider that the sequence grey levels vary from 0 to $N-1$. At frame f_t three $N \times N$ matrices $\mathbf{C}^R_{t,t+1}$, $\mathbf{C}^G_{t,t+1}$ and $\mathbf{C}^B_{t,t+1}$ are created carrying information on the grey level transitions between frames f_t and f_{t+1}.

In other words, considering only the R component, the matrix $\mathbf{C}^R_{t,t+1}(i,j)$, with $0 \leq i \leq N-1$ and $0 \leq j \leq N-1$, corresponds to the probability: *a pixel with grey level i in frame f_t has grey level j in frame f_{t+1}*. The mutual information $I^R_{t,t+1}$ of the transition from frame f_t to frame f_{t+1} for the R component is expressed by:

$$I^R_{t,t+1} = -\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \mathbf{C}^R_{t,t+1}(i,j) \log \frac{\mathbf{C}^R_{t,t+1}(i,j)}{\mathbf{C}^R_{t,t+1}(i)\mathbf{C}^R_{t,t+1}(j)} \quad (1)$$

and the total mutual information is given by:

$$I_{t,t+1} = I^R_{t,t+1} + I^G_{t,t+1} + I^B_{t,t+1} \quad (2)$$

By the same considerations, the joint entropy $H^R_{t,t+1}$ of the transition from frame f_t to frame f_{t+1}, for the R component, is given by:

$$H^R_{t,t+1} = -\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \mathbf{C}^R_{t,t+1}(i,j) \log \mathbf{C}^R_{t,t+1}(i,j) \quad (3)$$

This study has been supported by the Commission of the European Communities in the framework of the Methods for Unified Multimedia Information Retrieval project MOUMIR.

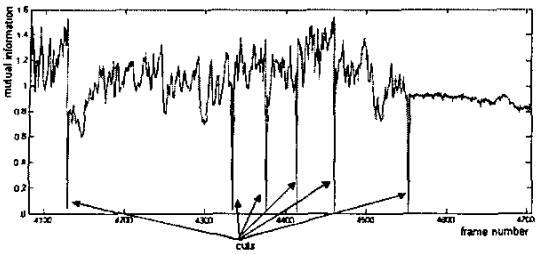

Fig. 1. *Time series of the mutual information from "star" video sequence showing detection of abrupt cuts. X-axis: frame number. Y-axis: mutual information.*

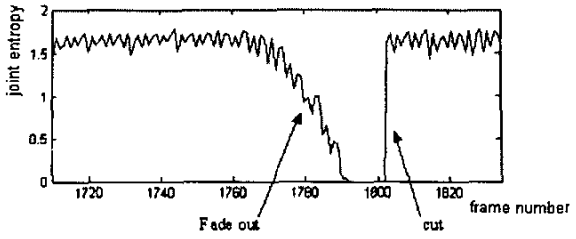

Fig. 2. *Joint entropy pattern from "basketball" video sequence showing a fade-out and an abrupt cut from a black frame to the next shot. X-axis: frame number. Y-axis: joint entropy.*

Fig. 3. *Consecutive frames from "football" video sequence showing an abrupt cut between two shots coupled with high movement.*

and the total joint entropy is obtained by:

$$H_{t,t+1} = H^R_{t,t+1} + H^G_{t,t+1} + H^B_{t,t+1} \tag{4}$$

2.1. Abrupt cut detection

A small value of the mutual information $I_{t,t+1}$ leads to a high probability of having a cut between frames f_t and f_{t+1}. Basically, in this context abrupt cut detection is an outlier detection in an one-dimensional signal [12]. Several algorithms exist for outlier detection, notably trimmed means and trimmed medians [12]. In order to detect possible shot cuts, an adaptive thresholding approach was employed. Trimmed local mutual information mean values on an one-dimensional temporal window W of size N_W are obtained at each time instant t_c by trimming the current value I_{t_c,t_c+1} at the current window center t_c [12]:

$$\bar{I}_{t_c} = E[I_{t,t+1}], \quad t \in W, \quad t \neq t_c \tag{5}$$

The quantity $\bar{I}_{t_c}/I_{t_c,t_c+1}$ is then compared to a threshold ϵ_c. An examples of abrupt cut detection using mutual information is illustrated in Figure 1.

Assuming that the video sequence has a length of N_L frames, the overall abrupt cut detection algorithm may be summarized as follows:

- calculate the mutual information time series $I_{t,t+1}$ (eq. 2) with $0 \leq t \leq N_L - 2$.
- calculate the trimmed average mutual information time series $\bar{I}_{t_c}$ at instant t_c (eq. 2) over a window N_W without taking into account the value I_{t_c,t_c+1}.
- if $\frac{\bar{I}_{t_c}}{I_{t_c,t_c+1}} \geq \epsilon_c$ then a cut is detected at instant t_c.

2.2. Fade detection

In order to get high precision in the detection of start and end points of fade-outs and fade-ins and to efficiently distinguish fades from cuts, the joint entropy (4) is employed. The joint entropy measures the amount of information carried between frames. Therefore, its value decreases during fades, where a weak amount of inter-frame information is present.

Thus, only the values of $H_{t,t+1}$ below a threshold T, set up near zero are examined. The instant, where the joint entropy is at a local minimum, is detected and is characterized as the end time instant t_e of the fade-out. The next step consists in searching for the fade-out start point t_s in the previous frames using the criterion:

$$\frac{H_{t_s,t_s+1} - H_{t_s-1,t_s}}{H_{t_s-1,t_s} - H_{t_s-2,t_s-1}} \geq \epsilon_f \tag{6}$$

where ϵ_f is a predefined threshold. The same procedure also applies for fade-in detection (with t_s being detected at first). Finally, the segment is considered as a fade only if $t_e - t_s \geq 2$, otherwise it is labeled as a cut. An example of joint entropy pattern showing a fade-out detection is presented in Figure 2.

The overall fade-in detection algorithm may be summarized as follows:

- calculate the joint entropy time series $H_{t,t+1}$ (eq. 4) with $0 \leq t \leq N_L - 2$.
- if, at instant t_e, the joint entropy H_{t_e,t_e+1} has a local minimum and is below a threshold, caracterize t_e as a fade-in ending point.
- if equation (6) is satisfied at instant t_s, and $t_e - t_s \geq 2$ then t_s is characterized as the fade-in starting point.

3. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed method was tested on several real TV sequences having many commercials in-between, characterized by significant camera effects like zoom-ins/outs and pans, abrupt camera movement and significant object and camera motion inside single shots (e.g. "football" video, Figure 3). For each video sequence, the human observer has determined the precise locations and duration of the edits to be used as ground truth.

In order to evaluate the performance of the segmentation method presented in section 2, the following measures, inspired by receiver operating characteristics in statistical detection theory, were used [1, 13]. Let GT denote the ground truth, Seg the segmented (correct and false) shots using our methods and $|E|$ the number of elements (frames) of a set E. The following measures have been considered:

Fig. 4. *Consecutive frames from "football" video sequence showing an occlusion during panning.*

- the *Recall* measure, also called true positives function or sensitivity, corresponding to the probability of detection:

$$Recall = \frac{|Seg \bigcap GT|}{|GT|} \quad (7)$$

- the *Precision* corresponding to the accuracy of the method considering false detections:

$$Precision = \frac{|Seg \bigcap GT|}{|Seg|} \quad (8)$$

- the *Overlap* measure defined as:

$$Overlap = \frac{|Seg \bigcap GT|}{|Seg \bigcup GT|} \quad (9)$$

  It is considered as a strong test for detection accuracy, since for example a shot of length N_L shifted by one frame results in only $\frac{N_L-1}{N_L}$ overlap.

At first, experimental tests were performed using a common prefixed threshold for all video sequences in order to detect shot boundaries. The results are summarized in Table 1. The large majority of the cuts were correctly detected even in the case of the "basketball" video sequence, which contains fast object and camera movements. Compared to histogram-based methods, the mutual information and joint entropy metrics are not sensitive to shot illumination changes even in the RGB color model. This comes from the fact that both (joint entropy and mutual information) operate with coocurrence matrices. Therefore, no false positive appeared due to camera flashes (Table 1). A snapshot of the "football" sequence is shown in Figure 4, where a big object appears in front of the camera. This case is generally characterized by standard methods as a transition, while our method correctly did not characterize it so.

A second experiment consists in applying our algorithms to the same sequences with an adaptive threshold chosen individually for each video sequence. As can be observed in Table 2, the results illustrate slightly better shot boundary detection rates compared to the fixed threshold.

In both experimental setups, the boundaries of the fades were detected within a precision of ± 2 frames. In most cases the boundaries towards black frames were recognized with no error. The robustness of the joint entropy measure in fade detection and especially in avoiding false fade detections is illustrated in Figures 5and 6.

Our method was also compared with two different approaches proposed in the literature. At first, we compared the joint entropy metric to the technique relying on the average frame grey level descent (AD) for fade detection [1]. The AD method is based on the observation that the average frame grey level time series of a video sequence is a decreasing function towards zero in the case of a fade-out. The opposite holds for fade-ins. As can be seen in Table 3, several fades were not correctly detected by AD showing a weaker performance of AD than our approach (Tables 1 and 2).

Grey level-based fade detection evaluation

video	fade-ins		
	Recall	Precision	Overlap
basketball	0.85	1.00	0.41
news I	1.00	0.86	0.54
video	**fade-outs**		
	Recall	Precision	Overlap
basketball	1.00	1.00	0.85
news I	1.00	0.86	0.65

Table 3. *Fade detection results using the AD method.*

Finally, we compared our algorithm to the technique proposed in [9]. This approach combines two shot boundary detection schemes based on color frame differences and color vector histogram differences between successive frames. It is claimed to efficiently detect shot boundaries even under strong edit effects and camera movement. In order to overcome the possible drawback of histogram sensitivity to shot illumination changes the method operates in the HLS color space and ignores luminance information. The results of this algorithm applied on the same video sequences are summarized in Table 4. Several false shot cut detections were performed due to camera flushes. Although this approach has a high shot cut detection rate, its accuracy is generally lower compared to the mutual information measure (Tables 1 and 2).

4. CONCLUSION

A new technique for shot transitions detection using the mutual information and the joint entropy measures was presented. The accuracy of our approach was experimentally shown to be very high. Experiments illustrated that fade detection using the joint entropy can efficiently differentiate fades from cuts, pans, object or camera motion and other types of video scene transitions, while most of the methods reported in the current literature fail to characterize these kinds of transitions.

5. REFERENCES

[1] R. Lienhart. Comparison of automatic shot boundary detection algorithms. In *Proc. of SPIE Storage and Retrieval for Image and Video Databases VII, San Jose, CA, U.S.A.*, volume 3656, pages 290–301, January 1999.

[2] A. Dailianas, R. B. Allen, and P. England. Comparison of automatic video segmentation algorithms. In *Proceedings, SPIE Photonics East'95: Integration Issues in Large Commercial Media Delivery Systems, Oct. 1995, Philadelphia*, volume 2615, pages 2–16, 1995.

[3] G. Ahanger and T.D.C. Little. A survey of technologies for parsing and indexing digital video. *Journal of visual Communication and image representation*, 7(1):28–43, 1996.

[4] S. Tsekeridou and I. Pitas. Content-based video parsing and indexing based on audio-visual interaction. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(4):522–535, 2001.

[5] R. Lienhart. Reliable dissolve detection. In *Proc. of SPIE Storage and Retrieval for Media Databases 2001*, volume 4315, pages 219–230, January 2001.

Fixed threshold shot detection evaluation

video	cuts		fade-ins			fade-outs		
	Recall	Precision	Recall	Precision	Overlap	Recall	Precision	Overlap
basketball	1.00	1.00	1.00	1.00	0.78	1.00	1.00	0.90
news	0.96	1.00	1.00	1.00	0.71	1.00	1.00	0.85
football	0.93	1.00	-	-	-	-	-	-
star	1.00	1.00	-	-	-	-	-	-

Table 1. *Shot detection results using a fixed threshold. See text for measures explanation.*

Adaptive threshold shot detection evaluation

video	cuts		fade-ins			fade-outs		
	Recall	Precision	Recall	Precision	Overlap	Recall	Precision	Overlap
basketball	1.00	1.00	1.00	1.00	0.78	1.00	1.00	0.90
news	1.00	1.00	1.00	1.00	0.71	1.00	1.00	0.85
football	0.93	1.00	-	-	-	-	-	-
star	1.00	1.00	-	-	-	-	-	-

Table 2. *Shot detection results using an adaptive threshold. See text for measures explanation.*

Color-based shot detection evaluation

video	cuts	
	Recall	Precision
basketball	0.91	0.97
news	0.96	0.98
football	0.96	1.00
star	0.93	0.98

Table 4. *Shot detection results using the method presented in [9]. See text for measures explanation.*

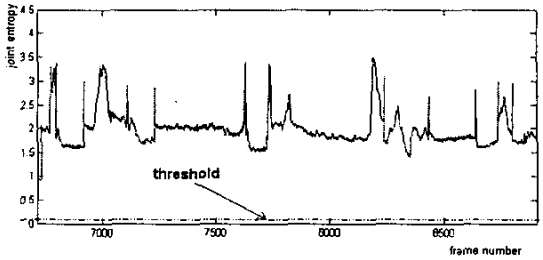

Fig. 5. *A joint entropy pattern from "star" video sequence presenting no fades. The high values of the joint entropy measure enable the method to avoid false detections. X-axis: frame number. Y-axis: joint entropy.*

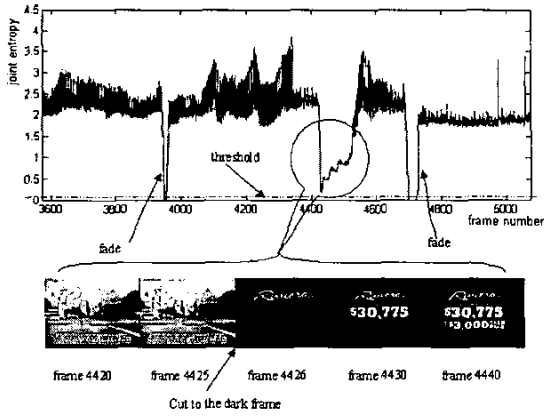

Fig. 6. *A joint entropy pattern from "news" video sequence presenting fades. The very low local minima of the joint entropy function represent fades. X-axis: frame number. Y-axis: joint entropy.*

[6] M. S. Drew, Z.-N. Li, and X. Zhong. Video dissolve and wipe detection via spatio-temporal images of chromatic histogram differences. In *Proceeding of IEEE Int. Conf. on Image Processing (ICIP 2000)*, volume 3, pages 929–932, 2000.

[7] Y. Wang, Z. Liu, and J.-Ch. Huang. Multimedia content analysis using both audio and visual clues. *IEEE Signal Processing Magazine*, 17(6):12–36, November 2000.

[8] A. Del Bimbo. *Visual Information Retrieval.* Morgan Kaufmann Publishers, Inc, San Francisco, California, 1999.

[9] S. Tsekeridou, S. Krinidis, and I. Pitas. Scene change detection based on audio-visual analysis and interaction. In *2000 Multi-Image Search and Analysis Workshop, accepted for publication, Schloss Dagstuhl, Germany, 12-17 March 2001*, March 2001.

[10] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley and Sons, New York, 1991.

[11] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* New York: McGraw-Hill, Inc., 1991.

[12] I. Pitas and A.N. Venetsanopoulos. *Nonlinear Digital Filters: Principles and Applications.* Kluwer Academic, 1990.

[13] C. E. Metz. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8:283–298, 1978.