Entropy Metrics used for Video Summarization *

Z. Černeková, C. Nikou and I. Pitas †

Department of Informatics
Aristotle University of Thessaloniki
Greece

Abstract

New methods for detecting shot boundaries in video sequences and for extracting key frames using metrics based on information theory are proposed. The method for shot cut detection relies on the mutual information and the joint entropy between the frames. It can detect cuts, fade-ins and fade-outs. The detection technique was tested on TV video sequences having different types of shots and containing significant object and camera motion inside the shots. It is demonstrated that the method detects both fades and abrupt cuts with high accuracy. The method for key frame extraction is using the mutual information. We show that it captures satisfactorily the visual content of the shot.

Keywords: shot boundary detection, entropy, mutual information, detection accuracy, video segmentation, video analysis, key frame extraction

1 Introduction

The indexing and retrieval of digital video is an active research area. Shot boundary detection and key frame extraction are important tasks for analyzing the content of video sequences, indexing, browsing, searching, summarizing and performing other content-based operations of large video databases.

The video shot is a basic structural building block of a video sequence and its boundaries need to be determined possibly automatically to allow for content-based video manipulation. A video shot may be defined as a sequence of frames captured by *one camera in a single continuous action in time and space* [4]. It should be a group of frames that have consistent visual characteristics (including color, texture and motion). After shots are segmented, key frames can be extracted from each shot. *Key frame* is the frame which can represent the salient content of the shot. Depending of the content complexity of the shot, one or more frames can be extracted [22].

Early work on shot detection mainly focused on abrupt cuts. A comparison of existing methods is presented in [3, 9]. The standard color histogram-based algorithm and its variations are widely used for detecting cuts [1, 6, 15, 18]. 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 [7, 10, 11, 19, 21]. 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 summarisation. Existing techniques for fade detection rely on twin thresholding [2] or grey level statistics [9] and have a relatively high false detection rate. Moreover, standard methods based on histograms, even when they correctly detect scene changes, they cannot distinguish between fades and other transitions [17].

Key frames provide a suitable abstraction and framework for video indexing, browsing and retrieval. The use of key frames greatly reduces the amount of data required in video indexing and provides an organizational framework for dealing with video content. Much research work has been done in key frame extraction [8, 13, 20]. The simplest proposed methods are choosing for each shot only one frame usually the first one, regardless of the complexity of visual content. The more complicated approaches take into account visual content, motion analysis and shot activity [22]. These approaches either can not effectively capture the major visual content or are computationally expensive.

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 the information passed from one frame to another. Mutual information is used for detecting abrupt cuts, where the image intensity or color is abruptly changed. A large difference in content between two frames, that shows a weak inter-frame dependency leads to a small value of mutual information.

In the case of a fade-out, where visual intensity usually decreases to a black image, the decreasing inter-frame joint entropy is used as a metric. The opposite stands for a

*This paper has been supported by the Commission of the European Union in the framework of the project Methods for Unified Multimedia Information Retrieval (MOUMIR).

†E-mail: (zuzana,nikou,pitas)@zeus.csd.auth.gr

fade-in. The application of these entropy-based techniques for shot cut detection was experimentally demonstrated to be very efficient yielding false acceptance rate and false rejection rate very close to zero.

The proposed method was 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 [9]. We also compared our algorithm to the technique proposed in [17], which is an approach combining two shot boundary detection schemes based on color frame differences and color vector histogram differences between successive frames.

We propose also a method for extracting key frames from each shot using already calculated mutual information values. The mutual information expresses the content changes and thus, the selected key frames capture well the visual content of the shot.

The remainder of the paper is organized as follows. In Section 2, a brief description of the mutual information and the joint entropy is presented. The description of our approach for shot boundary detection is addressed in Section 3. Our method for key frame extraction is described in Section 4. Experimental results are presented and commented in Section 5 and conclusions are drawn in Section 6.

2 Background and definitions

2.1 Mutual information

Let X be a discrete random variable with a set of possible outcomes $A_X = \{a_1, a_2, \ldots, a_N\}$ having probabilities $\{p_1, p_2, \ldots, p_N\}$, with $p_X(x = a_i) = p_i, p_i \geq 0$ and $\sum_{x \epsilon A_X} p_X(x) = 1$.

The entropy measures the information content or "uncertainty" of X [5, 14] and it is given by:

$$H(X) = -\sum_{x \in A_X} p_X(x) \log p_X(x) \qquad (1)$$

It is a measure of expected information across the all the outcomes of the random variable.

The *joint entropy* of X, Y is expressed as:

$$H(X,Y) = -\sum_{x,y \in A_X, A_Y} p_{XY}(x,y) \log p_{XY}(x,y) \quad (2)$$

where $p_{XY}(x,y)$ is the joint probability density function. The *mutual information* between X and Y is given by:

$$I(X,Y) = -\sum_{x,y \in A_X, A_Y} p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x) p_Y(y)} \qquad (3)$$

and measures the amount of information that X conveys about Y.

If X and Y are independent random variables, some important properties of the mutual information are:

Figure 1: *Consecutive frames from "news" video sequence showing: (a) a fade-in,(b) a fade-out.*

- $I(X,Y) \geq 0$
- for both independent and zero entropy sources X and Y: $I(X,Y) = 0$
- $I(X,Y) = I(Y,X)$
- the relation between the mutual information and the joint entropy of the random variables X and Y is given by:

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \qquad (4)$$

where $H(X)$ and $H(Y)$ are the marginal entropies of X and Y.

In equation (4), the mutual information not only provides us with a measure of association between X and Y but also determines the quantity of information carried by each variable at their overlap. By these means, mutual information decreases because $H(X)$ or $H(Y)$ are weak in their region of overlap. On the other side, the joint entropy simply represents the information shared by X and Y without taking into account their separate contributions in their region of overlap.

2.2 Video Cuts and Fades

A video shot cut (abrupt cut) is an instantaneous content transition from one shot to the next one. It is obtained by simply joining two different shots without the insertion of any other photographic effect. The cut boundaries show an abrupt change in image intensity or color. Cuts between shots with little content or camera motion and constant illumination conditions can be easily detected by looking for sharp brightness changes. However, in presence of continuous fast object motion, camera movements or illumination changes, it is difficult to distinguish if brightness changes are due to these conditions or to the transition from one shot to the other [2].

Fading is an optical process determining the progressive darkening of a shot until the last frame becomes black (fade-out, see Figure 1a). In the opposite, fade-in allows the gradual transition from black frame to the fully illuminated one (see Figure 1b). Fades spread the boundary between two shots across a number of consecutive video frames. They have both starting and ending frames identifying the transition sequence. In both cases (fade-in, fade-out) fades can be mathematically modeled as luminance scaling operations.

If $G(x,y,t)$ is a grey scale sequence and l_s is the length of the transition sequence, a chromatic scaling of $G(x,y,t)$ is modeled as [2]:

$$E(x,y,t) = G(x,y,t) \cdot (1 - \frac{t}{l_s}) \qquad t \in [t_0, t_0 + l_s] \quad (5)$$

Therefore, fade-out is modeled by:

$$E(x,y,t) = G_1(x,y) \cdot (\frac{l_1 - t}{l_1}) \quad (6)$$

and fade-in by:

$$E(x,y,t) = G_2(x,y) \cdot (\frac{t}{l_2}) \quad (7)$$

3 Shot detection

In our approach, the mutual information and the joint entropy between two successive frames is calculated separately for each of the RGB components. Let us consider that grey levels of the image sequence 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, that carry 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 $0 \le i \le N-1$ and $0 \le j \le N-1$, corresponds to the joint probability: *a pixel with grey level i in frame f_t has grey level j in frame f_{t+1}*. $\mathbf{C}^R_{t,t+1}(i,j)$ represent a co-occurrence matrix between frames f_t and f_{t+1}. Following equation (3), 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:

$$I^R_{t,t+1} = -\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \mathbf{C}^R_{t,t+1}(i,j) \log \frac{\mathbf{C}^R_{t,t+1}(i,j)}{\mathbf{C}^R_{t,t+1}(i)\mathbf{C}^R_{t,t+1}(j)} \quad (8)$$

and the total mutual information is given by:

$$I_{t,t+1} = I^R_{t,t+1} + I^G_{t,t+1} + I^B_{t,t+1} \quad (9)$$

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:

$$H^R_{t,t+1} = -\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} \mathbf{C}^R_{t,t+1}(i,j) \log \mathbf{C}^R_{t,t+1}(i,j) \quad (10)$$

and the total joint entropy is obtained by:

$$H_{t,t+1} = H^R_{t,t+1} + H^G_{t,t+1} + H^B_{t,t+1} \quad (11)$$

3.1 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 [16]. 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 [16]:

$$\bar{I}_{t_c} = E[I_{t,t+1}], \quad t \in W, \quad t \neq t_c \quad (12)$$

The quantity $\bar{I}_{t_c}/I_{t_c,t_c+1}$ is then compared to a threshold ϵ_c. Some examples of abrupt cut detection using mutual information are illustrated in Figure 2 and Figure 3.

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. 9) with $0 \le t \le N_L - 2$.
- calculate the trimmed average mutual information time series $\bar{I}_{t_c}$ at instant t_c (eq. 9) over a window N_W without taking into account the value I_{t_c,t_c+1}.
- if $\frac{\bar{I}_{t_c}}{I_{t_c,t_c+1}} \ge \epsilon_c$ then a cut is detected at instant t_c.

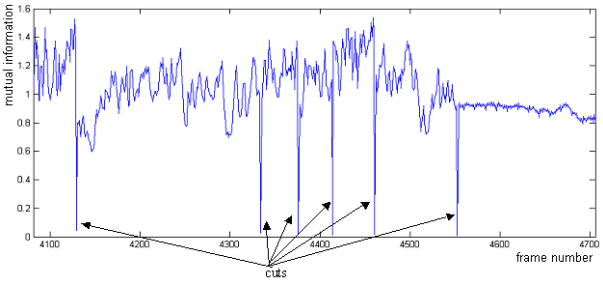

Figure 2: *Time series of the mutual information from "star" video sequence showing detection of abrupt cuts. X-axis: frame number. Y-axis: mutual information.*

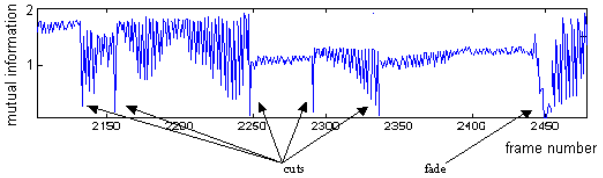

Figure 3: *Time series of the mutual information from "basketball" video sequence showing abrupt cuts and fades. X-axis: frame number. Y-axis: mutual information.*

3.2 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 (11) is employed. The joint entropy measures the amount of information carried between frames. Therefore, its value de-

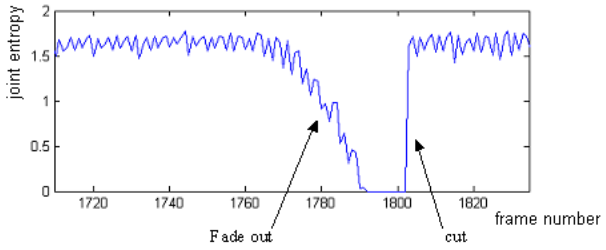

Figure 4: *Joint entropy pattern from "basketball" video sequence showing a fade-out and a transition from a black frame to the next shot. X-axis: frame number. Y-axis: joint entropy.*

creases 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 near to zero are examined. The instant, where the joint entropy presents 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:

$$\frac{H_{t_s,t_s+1} - H_{t_s-1,t_s}}{H_{t_s-1,t_s} - H_{t_s-2,t_s-1}} \geq \epsilon_f \qquad (13)$$

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 4.

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

- calculate the joint entropy time series $H_{t,t+1}$ (eq. 11) 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, characterize t_e as a fade-in ending point.
- if equation (13) is satisfied at instant t_s, and $t_e - t_s \geq 2$ then t_s is characterized as the fade-in starting point.

4 Key frame selection

After video sequence temporal segmentation to the shots, the key frames can be selected from each shot for video indexing. Our approach uses already calculated mutual information values, which provided us information about content changes between consecutive frames in the shot. Let us have a video shot $s = \{f_1, f_2, ..., f_N\}$ obtained by our method for shot cut detection. Let the mutual information values in this shot be $I_s = \{I_{1,2}, I_{2,3}, ..., I_{N-1,N}\}$. In order to find if the content in the shot changes significantly, the standard deviation σ_{I_s} of the values of mutual information in this shot is calculated. The value σ_{I_s} is compared to

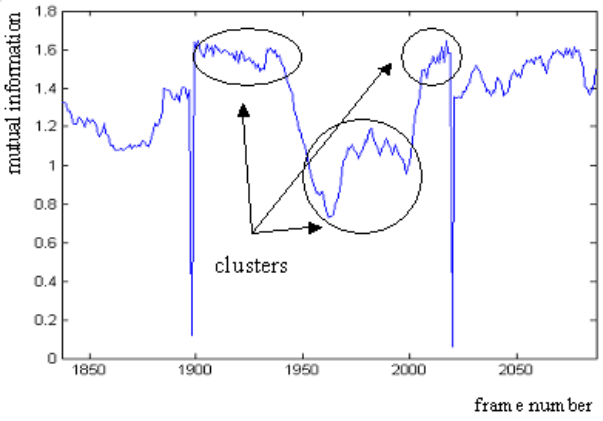

Figure 5: *A mutual information pattern from "star" video sequence presenting the clusters created after using our method. The selected potential key frames from each cluster is shown in Figure 10*

predefined threshold ϵ. If $\sigma_{I_s} < \epsilon$ it means the content during the whole shot changed negligible, so whatever frame can effectively express visual content. In our method for shots with no or small changes in content the first frame is selected as a key frame.

The shots with big changes in content are further processed using a clustering method. The mutual information values in the shot are divided into clusters $\{c_i\}_{i=1}^{K}$, where K is a number of clusters obtained after clustering. A threshold parameter δ provides us a control over the density of classification. Initially, first five mutual information values $\{I_{1,2}, I_{2,3}, ..., I_{5,6}\}$ are assigned to a first cluster c_1. Then at each iteration next five values $\{I_{5t+1,5t+2}, I_{5t+2,5t+3}, ..., I_{5t+5,5t+6}\}$ are added to the cluster. The standard deviation σ_1 is calculated for the cluster and is compared to the threshold δ. If $\sigma_1 > \delta$, it means that content of the video sequence has changed. Therefore, a new cluster is created and the next set of values is assigned to it. An example of such created clusters can be seen in Figure 5. This way all frames from the given shot are splitted to the groups, dependently on the mutual information values included in clusters. In other words, lets suppose $c_i = \{I_{i_1,i_2}, I_{i_2,i_3}, ..., I_{i_{n-1},i_n}\}$ is a obtained cluster, then frames $\{f_{i_1}, f_{i_2}, ..., f_{i_{n-1}}\}$ integrate a group of frames with similar visual content. Finally, from each group the first frame is taken as a potential key frame.

After extracting potential key frames $\{k_i\}_{i=1}^{K}$ from the shot s, we aim to reduce the number of key frames to represent the shot. To do this, these key frames are compared between themselves by calculating their mutual information. If the content of the frames even after changing is similar enough (indicated by a high mutual information value), it can be presented by less number of frames. Therefore, if $I_{k_i,k_{i+1}} > \epsilon$, where ϵ is a predefined threshold, only the frame k_i is considered to be a key frame and is compared to the next potential key frame k_{i+2}. Otherwise, both frames k_i and k_{i+1} are taken as key frames and

Video sequences

video	frames	cuts	fade-ins	fade-outs
basketball	3882	44	7	4
news	9446	40	6	6
football	5589	28	0	0
star	19722	147	0	0

Table 1: *The video set used in our experiments and the respective number of frames, abrupt cuts, fade-ins and fade-outs.*

k_{i+1} is further compared to the others potential key frames $\{k_{i+2}, \ldots, k_K\}$.

5 Experimental results and discussion

The proposed method was tested on several real TV sequences having many commercials in-between (see Table 1), 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. "basketball" video, Figure 3). The video sequences contain sport, studio news, advertisements, political talks and TV series logos. 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 3, the following measures, inspired by receiver operating characteristics in statistical detection theory, were used [3, 12]. 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:

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

$$Recall = \frac{|Seg \bigcap GT|}{|GT|} \tag{14}$$

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

$$Precision = \frac{|Seg \bigcap GT|}{|Seg|} \tag{15}$$

- the *Overlap* measure defined as:

$$Overlap = \frac{|Seg \bigcap GT|}{|Seg \bigcup GT|} \tag{16}$$

  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 $\frac{N_L - 1}{N_L}$ overlap.

Figure 6: *Consecutive frames from "football" video sequence showing an abrupt cut between two shots coupled with high movement.*

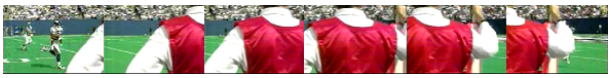

Figure 7: *Consecutive frames from "football" video sequence showing an occlusion during panning.*

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 3. The 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. As both operate with co-occurrence matrices. Therefore, no false positive appeared due to camera flashes (Table 3). A part of the "football" video sequence showing a cut between two shots involving high content motion that was successfully detected by the proposed method is presented in Figure 6. A snapshot of the "football" sequence is shown in Figure 7, 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 4, the results illustrate slightly better shot boundary detection rates with no false positive or true negative fade detections compared to the fixed threshold.

In both experimental setups, the boundaries of the fades

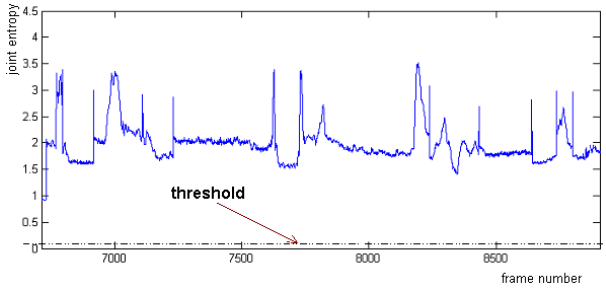

Figure 8: *A joint entropy pattern from "star" video sequence presenting no fades. The high values of the joint entropy measure enable the method to avoid false detections. X-axis: frame number. Y-axis: joint entropy.*

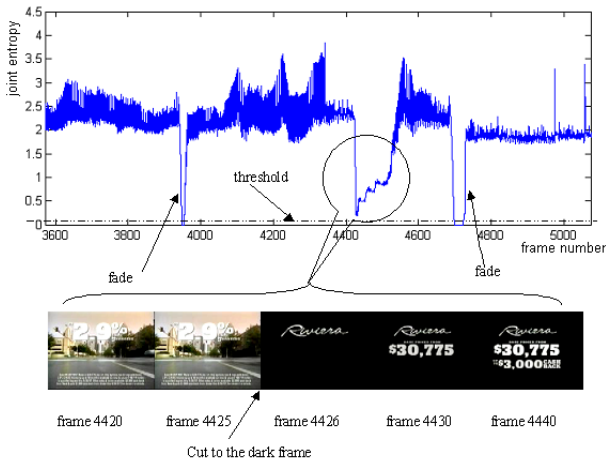

Figure 9: *A joint entropy pattern from "news" video sequence presenting fades. The very low local minima of the joint entropy function represent fades. X-axis: frame number. Y-axis: joint entropy.*

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 8 and 9.

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. The AD method is based on the observation that the average frame grey level time series of a video sequence is a decreasing function in the case of a fade-out. The opposite holds for fade-ins. As can be seen in Table 5, several fades were not correctly detected by AD showing a weaker performance of AD than our approach (Tables 3 and 4).

The above mentioned observations are also confirmed by the detection error statistics provided by the AD technique and the proposed joint entropy (JE) approach and presented in Tables 6 and 7 respectively. For a total number of 23 fades (Table 1), the starting and ending frames were detected by both methods and various statistics on the detection errors were calculated. The JE metric provides a superior performance than the AD method in median and mean error values and presents no errors in fade-out end point detection. Furthermore, the significantly smaller maximum errors of the JE technique with regard to AD, illustrate the robustness of our algorithm.

Finally, we compared our algorithm to the technique proposed in [17]. 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

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 2: *Shot detection results using the method presented in [17]. See text for measures explanation.*

Figure 10: *Potential key frames from "star" video sequence extracted from each cluster of the shot. After calculating mutual information between this frames only first frame (frame number 1904) was selected as a key frame to represent content of the shot.*

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 2. 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 3 and 4). Moreover, our technique revealed more robust to shots with small length, occurring particularly during TV advertisements (tables 3 and 4).

After video segmentation to the shots, we applied our method for key frame selection on the video sequences. In case of shots without significant content changes our method successfully chose only one frame, even if after clustering were extracted more potential key frames. An example can be seen in Figure 10. For shots with big content changes (usually camera or object movements) more key frames were selected, depending on visual complexity of the shot. An example of key frames extracted from one shot with more complicated content can be seen in Figure 11. Our method does not strongly depends on the threshold parameter δ used for creating clusters, due to final comparison of potential key frames. The examples of selected key frames are shown in Figure 12. After using our proposed method for shot boundary detection, this method is efficient to compute, because is using already calculated mutual information values, and as we showed it captures satisfactorily visual content of the shot.

Figure 11: *Examples of key frames selected by our method from "star" video sequence to represent visual content of one shot.*

Figure 12: *Examples of key frames extracted by our method from "star" video sequence.*

6 Conclusions

New methods for shot boundary detection and key frame selection using the mutual information and the joint entropy measures were presented. The accuracy of our approach was experimentally shown to be very high. Experiments have 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.

References

[1] 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.

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

[3] P. Browne, A. F. Smeaton, N. Murphy, N. O'Connor, S. Marlow, and C. Berrut. Evaluation and combining digital video shot boundary detection algorithms. In *Proceedings of the Fourth Irish Machine Vision and Information Processing Conference, Queens University Belfast*, 2000.

[4] X. U. Cabedo and S. K. Bhattacharjee. Shot detection tools in digital video. In *Proceedings of Nonlinear Model Based Image Analysis 1998, Springer Verlag, Glasgow*, pages 121–126, July 1998.

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

[6] 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.

[7] 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 909–932, 2000.

[8] B. Grünsel and A. M. Tekalp. Content-based video abstraction. In *Proceeding of IEEE Int. Conf. on Image Processing (ICIP'98), Chicago IL*, October 1998.

[9] 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.

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

[11] R. Lienhart and A. Zaccarin. A system for reliable dissolve detection in video. In *Proceeding of IEEE Intl. Conf. on Image Processing 2001 (ICIP'01), Thessaloniki, Greece*, Oct. 2003.

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

[13] A. Nagasaka and Y. Tanaka. Automatic video indexing and full-video search for object appearances. In *Visual Database Systems II*, 1992.

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

[15] N. V. Patel and I. K. Sethi. Video shot detection and characterization for video databases. *Pattern Recognition*, 30(4):583–592, April 1997.

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

[17] 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.

[18] 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.

[19] 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.

[20] W. Wolf. Key frame selection by motion analysis. In *Proceeding IEEE Int. Vonf. Acoust., Speech and Signal Proc.*, 1996.

[21] R. Zabih, J. Miller, and K. Mai. A feature-based algorithm for detecting and classifying production effects. *ACM Journal of Multimedia Systems*, 7:119–128, 1999.

[22] Y. Zhuang, Y. Rui, T. S. Huang, and S. Metrotra. Adaptive key frame extraction using unsupervised clustering. In *Proceeding of IEEE Int. Conf. on Image Processing (ICIP'98), Chicago IL*, pages 886–890, October 1998.

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 3: *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 4: *Shot detection results using an adaptive threshold. See text for measures explanation.*

Grey level-based fade detection evaluation

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

Table 5: *Fade detection results using the AD method.*

AD Detection Error Statistics

effects	fade-outs		fade-ins		fades	
frame	f_s	f_e	f_s	f_e	f_s	f_e
median	1	1	1.5	1	1	1
mean $\pm$ s. dev.	1.7 ± 2.3	1.3 ± 0.9	2.5 ± 3.2	2.9 ± 4.0	2.1 ± 2.8	2.2 ± 3.0
max	8	4	9	13	9	13

Table 6: *Fade detection error statistics for the average grey level descent-based technique (AD). The median, mean, standard deviation and maximum values of the starting (f_s) and ending (f_e) frame detection errors are presented. Errors are expressed in terms of frame numbers.*

JE Detection Error Statistics

effects	fade-outs		fade-ins		fades	
frame	f_s	f_e	f_s	f_e	f_s	f_e
median	1	0	0	1	0	1
mean $\pm$ s. dev.	2.1 ± 2.6	0.0 ± 0.0	0.2 ± 0.6	2.7 ± 2.7	1.1 ± 1.9	1.5 ± 2.4
max	8	0	2	8	8	8

Table 7: *Fade detection error statistics for the joint entropy-based technique (JE). The median, mean, standard deviation and maximum values of the starting (f_s) and ending (f_e) frame detection errors are presented. Errors are expressed in terms of frame numbers.*