World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:6, No:3, 2012

An Efficient Method of Shot Cut Detection

Lenka Krulikovská, Jaroslav Polec

***Abstract*—**In this paper we present a method of abrupt cut detection with a novel logic of frames' comparison. Actual frame is compared with its motion estimated prediction instead of comparison with successive frame. Four different similarity metrics were employed to estimate the resemblance of compared frames. Obtained results were evaluated by standard used measures of test accuracy and compared with existing approach. Based on the results, we claim the proposed method is more effective and Pearson correlation coefficient obtained the best results among chosen similarity metrics.

***Keywords*—**Abrupt cut, mutual information, shot cut detection, Pearson correlation coefficient.

I. Introduction

PROGRES in the multimedia compression technology and computer performance has led to the widespread availability of digital video. There is a corresponding growth in the need for methods to reliably detect shot boundaries within the video sequence. The detection of shot boundaries provides a base for nearly all video abstraction and high-level video segmentation approaches. Therefore, solving the problem of shot-boundary detection is one of the major prerequisites for revealing higher level video content structure. Moreover, other research areas can profit considerably from successful automation of shot-boundary detection processes as well.

There are a number of different types of transitions or boundaries between shots [1]. A cut is an abrupt shot change that occurs in a single frame. A fade is a slow change in brightness usually resulting in or starting with a solid black frame. A dissolve occurs when the images of the first shot get dimmer and the images of the second shot get brighter, with frames within the transition showing one image superimposed on the other. A wipe occurs when pixels from the second shot replace those of the first shot in a regular pattern such as in a line from the left edge of the frames. Of course, many other types of gradual transition are possible.

Lenka Krulikovská is with the Institute of Telecommunication, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava, Slovakia (email: krulikovska@ktl.elf.stuba.sk)

Jaroslav Polec is with the Institute of Telecommunication, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, Ilkovičova 3, 812 19 Bratislava, Slovakia (email: polec@ktl.elf.stuba.sk)

ACKNOWLEDGMENT: Research described in the paper was financially supported by the Slovak Research Grant Agency (VEGA) under grant No. 1/0602/11 and by the Slovak University of Technology under project "Detection of cuts in video sequences" within the Program for support of young researcher for year 2011.

Different approaches have been proposed to extract shots. The major techniques used for the shot boundary detection are pixel differences, statistical differences, histogram comparisons [2], edge differences, compression differences and motion vectors [3]-[5].

There are various possibilities for improving on the basic methods. The variety of basic methods opens up the possibility of combining several of them into a multiple expert framework, explored in [6]-[8]. Also, one can use an adaptive threshold setting, by using statistics of the dissimilarity measure within a sliding window [9]-[11].

In general, abrupt transitions are much more common than gradual transitions, accounting for over 99% of all transitions found in video [12]. Therefore, the correct detection of abrupt shot transitions is a very important task in video segmentation and this paper is only focused on the detection of an abrupt cut.

Our aim was to propose method of abrupt cut detection, which can be used directly in video encoder without need of additional buffers or delaying the video encoding process. The paper is structured as follows: in the second section the chosen similarity metrics and evaluation measures are described. A proposed method of shot cut detection and obtained results are presented in the third section. All results are summarized and discussed in conclusion.

II. Similarity Metrics And Evaluation Measures

During the process of shot cut detection the position of cut is determined based on the similarity of compared frames. Huge dissimilarity between frames indicates the presence of abrupt cut. We have chosen mutual information, mean sum of absolute differences (MSAD), mean square error (MSE) and Pearson correlation coefficient (PCC).

The accuracy of shot cut detection algorithm is usually determined by precision measure, recall measure and F1 score measure.

A. Mutual information

The mutual information measures the amount of information about random variable X conveyed to random variable Y.

The average mutual information between the two processes can be calculated as the sum of the two self entropies minus the entropy of the pair [1]:

$$I(X,Y) = H(X) + H(Y) - H(X,Y). \qquad (1)$$

B. MSAD

MSAD is a widely used, extremely simple algorithm for measuring the similarity between image blocks. MSAD for images X and Y with dimension MxN is expressed as [13]:

$$MSAD = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left|X(i,j)-Y(i,j)\right|}{M.N} \quad (2)$$

C. MSE

MSE is the simplest and the most widely used full-reference quality metric. The MSE can be calculated for two images as follows [13]:

$$MSE = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(X(i,j)-Y(i,j)\right)^2}{M.N} \quad (3)$$

D. PCC

In statistics, the Pearson's correlation coefficient typically denoted by r (sometimes also referred to as the Pearson product-moment correlation coefficient) has been widely employed to measure the correlation (or strength of linear dependence) between two variables X and Y [14]. The value for a Pearson correlation coefficient can fall between -1 and 1, where 0 means no correlation. Generally, correlations above 0.80 are considered as really high. It is expressed as:

$$r = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(X(i,j)-X^m)(Y(i,j)-Y^m)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(X(i,j)-X^m)^2}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(Y(i,j)-Y^m)^2}} \quad (4)$$

where X^m and Y^m stand for mean pixel intensity of images X and Y.

E. Evaluation measures

The recall measure, also known as the positive true function or sensitivity, corresponds to the ratio of correct experimental detections over the number of all true detections. The precision measure is defined as the ratio of correct experimental detections over the number of all experimental detections. F1 score measure is a combined measure that results in high value if, and only if, both precision and recall result in high values [1].

III. Proposed Method of Shot Cut Detection

The novelty of the presented method is in the evaluation of the positions of abrupt cuts. The most of existing methods calculate similarity of two consecutive frames by chosen metric and determine the position of cut based on obtained values. Our proposed method compares the actual frame with its motion compensated prediction.

The principle is very easy, selected frames are compared and the positions of cuts are determined based on huge dissimilarity. At first we have used mutual information and the obtained results were compared with a method with common logic of frames comparison proposed by Cernekova in [1]. Subsequently we have evaluated the relevance of other selected similarity metrics.

Presented results were obtained by test experiment performed on a video sequence (1989 frames) at CIF resolution (352 x 288 pixels) with 7 abrupt cuts sampled at rate of 30 frames per second. The test video sequence consists of eight standard test sequences. For prediction of frames we have employed motion estimation scheme used in H.264 video encoding standard. The total value of similarity metric is calculated as the average of values for components Y, U and V.

An example of abrupt cut from used video sequence is displayed on Fig. 1.

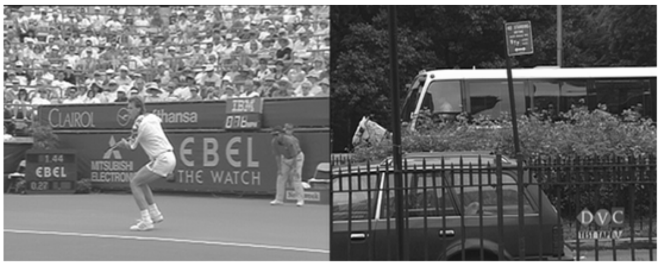


Fig. 1 An example of abrupt cut

A. Mutual Information

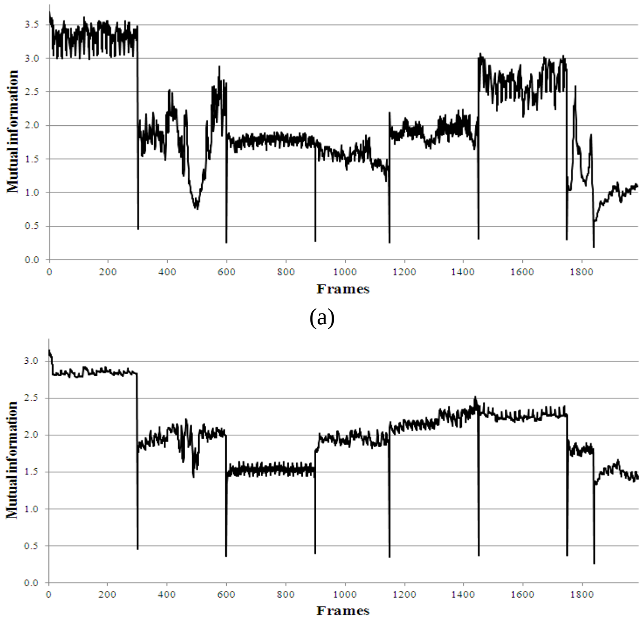
Fig. 2 (a) shows the result obtained by existing approach proposed by Cernekova [1] and Fig. 2 (b) shows the result of proposed algorithm with mutual information.

Cuts are expected in minimums, because the mutual information obtains the highest values, when very different variables, in our case frames, are compared. Based on results displayed on Fig 2, both methods were able to detect all cuts. The difference is in the range of obtained values and in the behavior of these methods for non cuts frames.

Method proposed by Cernekova is based on frame by frame comparison and it is more sensitive to object or camera motion within the shot. This can be observed for example within the second shot (frames 300-600) and for the last shot. Sensitivity to motion can lead to false detection with use of threshold due to reached small values for non cut frames. In opposite, the proposed algorithm suppressed the local minimums in non cut values, which can cause increasing of the accuracy of shot cut algorithm, thus it is more robust.

For better comparison and to show the proposed method is more robust, we have simulated automatic shot boundary detection with three fixed thresholds. The accuracy was evaluated by precision, recall and F1 score measures. The results for Cernekova's method are in Table I and for our proposed method in Table II.

In all tables C stands for correctly detected cuts, M for missed cuts, F for false detected cuts, P for precision measure, R for recall measure and F1 for F1 score measure.



(a)

(b)

Fig. 2 Plot of cut detection using mutual information for (a) method proposed by Cernekova [1] and (b) our proposed algorithm

The measures used for evaluation can obtain values from 0 to 1, where 1 represents the highest accuracy. The results show the proposed methods reached highest accuracy for two of simulated thresholds, existing methods only for one. For the threshold 1.5 the accuracy was decreased due to false detections, 359 for existing method and 152 for proposed. For threshold value 1 the existing method reached 116 false detections in contrast to no false detection by presented method. These results demostrate the method with novel logic of frames' comparison is more robust to huge object or camera motion within the shot.

TABLE I
RESULTS OF CERNEKOVA'S METHOD USING MUTUAL INFORMATION

Threshold value	C	M	F	P	R	F1
0.5	7	0	0	1	1	1
1	7	0	116	0.057	1	0.108
1.5	7	0	359	0.019	1	0.038

TABLE II
RESULTS OF PROPOSED METHOD USING MUTUAL INFORMATION

Threshold value	C	M	F	P	R	F1
0.5	7	0	0	1	1	1
1	7	0	0	1	1	1
1.5	7	0	152	0.044	1	0.084

B. MSAD

Fig. 3 displays the plot of shot cut detection for proposed method using MSAD as evaluation measure. The cuts are expected in the peaks, because MSAD has the highest values for frames with different content.

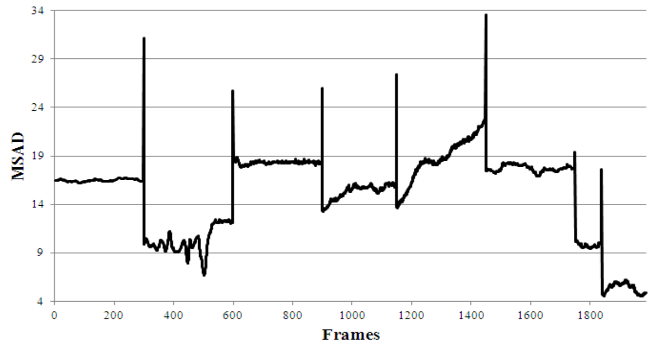


Fig. 3 Plot of cut detection using MSAD for proposed method

The behavior of MSAD for non cut frames shows this measure is not suitable for shot cut detection, because some cut values are lower than non cut values. This feature would cause false detections. To illustrate this fact, we have run simulation of shot cut detection with threshold values equal to the two lowest cut values. The results are shown in Table III.

TABLE III
RESULTS OF PROPOSED METHOD USING MSAD

Threshold value	C	M	F	P	R	F1
17.6	7	0	724	0.010	1	0.019
19.35	6	1	120	0.048	0.857	0.09

First threshold is set to the lowest MSAD value for cut (at position 1839 in Fig. 3). We do not use lower value with effort to avoid higher amount of false detections during automatic detection. For this threshold 724 cuts were false detected, what caused decreasing of precision measure and subsequently of F1 score.Second threshold is set to the value of the second lowest cut (at position 1749 in Fig. 3). With this threshold there were 120 false detections, but one real cut was missed.It is obvious MSAD is not appropriate metric for shot cut detection due to small difference among cut and non cut values.

C. MSE

Similarly to MSAD, MSE measure reached the highest values for different frames, thus the cuts are expected in the local maximums. The plot of shot cut detection using MSE is displayed on Fig. 4.We can observe that all seven cuts are situated to the peaks, the smallest difference in MSE value between cut and non cut frames is about 400, what provides sufficient distance.

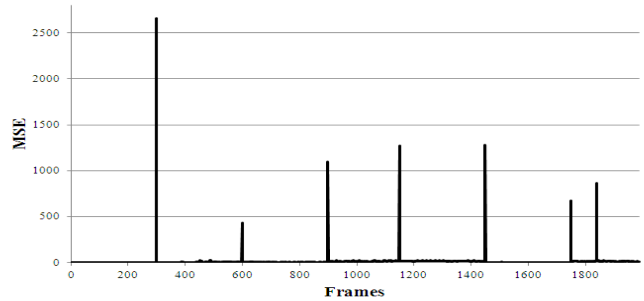


Fig. 4 Plot of cut detection using MSE for proposed method

The values for cuts are also in huge range, from 434 to 2661, this could cause problems (missed detections) if we use threshold. To demonstrate this situation, we have run automatic shot boundary detection using threshold in range from 100 to 1300 with step 200. Results can be found in Table IV.

TABLE IV
RESULTS OF PROPOSED METHOD USING MSE

Threshold value	C	M	F	P	R	F1
100	7	0	0	1	1	1
300	7	0	0	1	1	1
500	6	1	0	1	0.857	0.923
700	5	2	0	1	0.714	0.833
900	4	3	0	1	0.571	0.727
1100	3	4	0	1	0.429	0.6
1300	1	6	0	1	0.143	0.25

According to Table IV, first missed cut occurred for threshold set to 500. For threshold with value 1300, there were 6 missed detected cuts, what cause the decrease of algorithm accuracy (performed by F1 score measure) to 25%.

MSE is able to locate position of cuts, but due to huge differences in cut values, there is a risk of missed detections.

D.PCC

PCC can obtain values from -1 to 1. For purposes of shot cut detection we have used absolute value of PCC. The plot of detection for PCC is on Fig. 5. As PCC is a kind of correlation metric, the cuts are expected in local minimums.

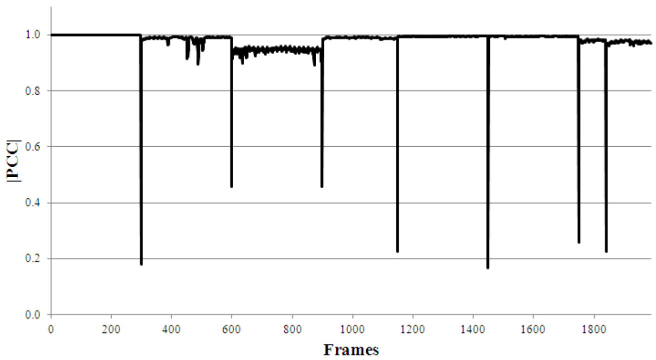


Fig. 5 Plot of cut detection using │PCC│ for proposed method

Shot cut detection using PCC resulted with sufficient distance between cut and non cut values. The another important advantage of using absolute value of PCC is that all values will be in range from 0 to 1, therefore it is easier to select appropriate threshold. The best choice is threshold value around 0.5.

E. Comparison of used metrics

With aim to provide the comparison of selected metrics in shot cut detection, we have normalized all values of each metric to the range from 0 to 1.

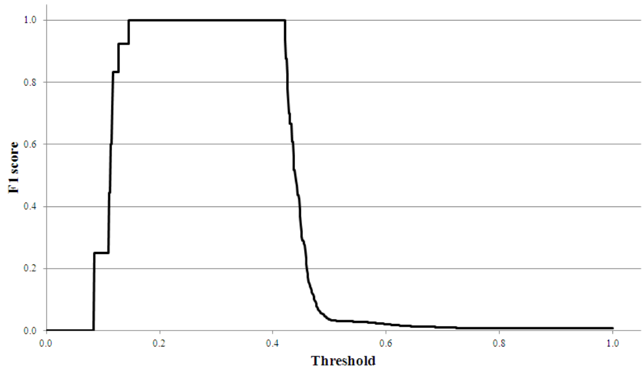


Fig. 6 The dependency of F1 score to threshold for shot cut detection using mutual information

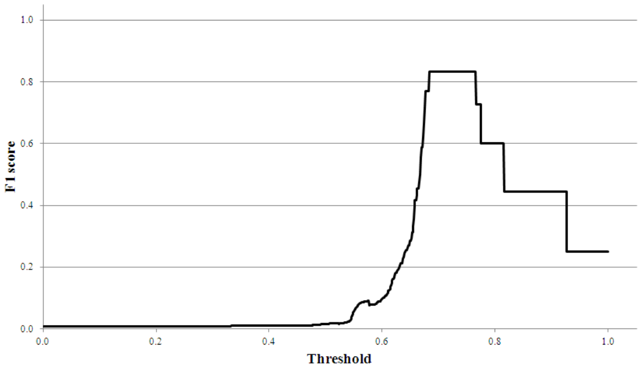


Fig. 7 The dependency of F1 score to threshold for shot cut detection using MSAD

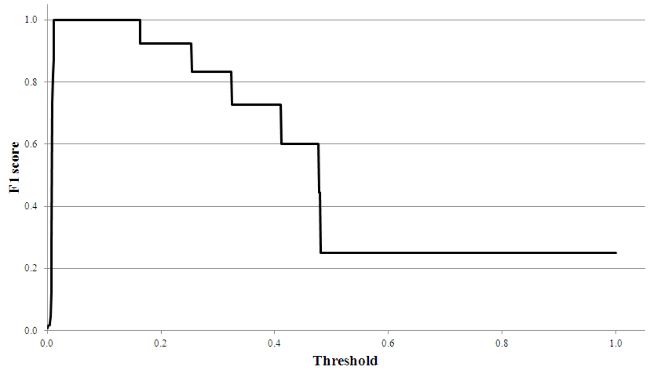


Fig. 8 The dependency of F1 score to threshold for shot cut detection using MSE

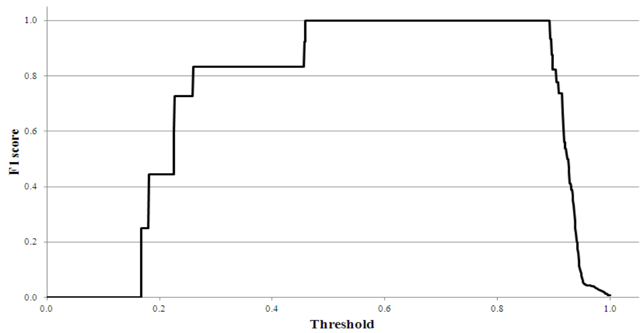


Fig. 9 The dependency of F1 score to threshold for shot cut detection using │PCC│

Next, we have run simulation of the dependency of precision, recall and F1 score to the selected threshold. Threshold obtains values from 0 to 1 with step 0.001.

The dependency for F1 score measure for individual metrics is displayed on Fig. 6 – Fig. 9. We have chosen F1 score, because this evaluation metric takes into account false detections and missed cuts together, there it provides more complex view of shot cut detection accuracy in comparison with precision and recall. All evaluation measure can obtain values from 0 to 1, where 1 stands for the highest accuracy of algorithm.

Proposed algorithm of shot cut detection using mutual information for about 27% of selected threshold range. Algorithm using MSAD never reached the highest possible accuracy due to false detections and missed cuts. Algorithm using MSE holds the highest accuracy for about 15% of threshold range, and then it is degraded due to missed cuts. Finally algorithms using the absolute value of PCC reached the highest accuracy for about 46% of simulated threshold range.

Based on obtained results and performed comparison, the algorithm using absolute value of PCC gives the best result in the terms of ability to keep the highest accuracy and simplicity of determining the threshold (we know the range of values we can expect).

IV. Conclusion

In this paper we presented a novel method for shot cut detection. The novelty of proposed method is in a new logic of frames' comparison. We have employed several similarity metrics and all results were evaluated and compared to existing approach.

Based on presented results we claim the absolute value of Pearson correlation coefficient provides the best results and another advantage of this metric is we know the range of values we can expected, what is important for decision about threshold value.

For future work we would like to employ proposed method directly to H.264 encoder for purposes of adaptive GOP structure.

References

[1] Z. Cernekova, I. Pitas, and C. Nikou, "Information theory-based shot cut/fade detection and video summarization", IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no.1, page(s): 82-91, January 2006

[2] A. Amiri and M. Fathy, "Video shot boundary detection using QR decomposition and Gaussian transition detection," EURASIP Journal on Advances in Signal Processing, Volume 2009, Article ID 509438.

[3] A. Hanjalic, "Shot-boundary detection: unraveled and resolved?," IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 2, pp. 90–105, 2002.

[4] J. S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques," Storage and Retrieval for Still Image and Video Databases IV, Proc. SPIE 2664, pp. 170-179, 1996.

[5] R. Lienhart, "Comparison of automatic shot boundary detection algorithms," Storage and Retrieval for Image and Video Databases VII, vol. 3656 of Proceedings of SPIE, pp. 290–301, San Jose, Ca, USA, 1999.

[6] M. R. Naphade, R. Mehrotra, A. M. Ferman, J. Warnick, T. S. Huang and A. M. Tekalp, "A high-performance shot boundary detection algorithm using multiple cues," Proc. IEEE Int. Conf. on Image Proc., volume 2, pages 884–887, 1998.

[7] C. Taskiran and E. J. Delp, "Video scene change detection using the generalized sequence trace," Proc. IEEE Int. Conf. on Image Proc., pages 2961–2964, 1998.

[8] Y. Yusoff, J. Kittler and W. Christmas, "Combining multiple experts for classifying shot changes in video sequences," Proc. 6th Int. Conf. on Multimedia Comp. and Systems (ICMCS), volume 2, pages 700–704, Florence, Italy, 1999.

[9] R. Dugad, K. Ratakonda and N. Ahuja, "Robust video shot change detection", IEEE Workshop on Multimedia Signal Processing, 1998.

[10] B. L. Yeo and B. Liu, "Rapid scene analysis on compressed video," IEEE Trans. On Circuits and Systems for Video Technology, 5(6):533–544, 1995.

[11] R. Zabih, J. Miller and K. Mai, "A feature-based algorithm for detecting and classifying production effects," ACM Multimedia Systems, 7(2):119–128, 1999.

[12] S. PASCHALAKIS and D. SIMMONS, (2008, April 24), "Detection of gradual transitions in video sequences" [Online]. Available: http://www.wipo.int/pctdb/en/wo.jsp?WO=2008046748&IA=EP2007060594&DISPLAY=STATUS.

[13] S. CUMAR (22.10.2001), "An introduction to image compression" [Online]. Available: http://www.debugmode.com/imagecmp/

[14] Y. K. Eugene, and R. G. Johnston, "The Ineffectiveness of the Correlation Coefficient for Image Comparisons," Technical Report LA-UR-96-2474, Los Alamos, 1996.