World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:6, No:9, 2012

Shot Detection using Modified Dugad Model

Lenka Krulikovská, Jaroslav Polec

***Abstract*—**In this paper we present a modification to existed model of threshold for shot cut detection, which is able to adapt itself to the sequence statistics and operate in real time, because it use for calculation only previously evaluated frames. The efficiency of proposed modified adaptive threshold scheme was verified through extensive test experiment with several similarity metrics and achieved results were compared to the results reached by the original model. According to results proposed threshold scheme reached higher accuracy than existed original model.

***Keywords*—**Abrupt cut, shot cut detection, adaptive threshold.

I. INTRODUCTION

WITH the rapid rise of interest in analysis of audio-visual material, there is a corresponding growth in the need for methods to reliably detect shot boundaries within the video sequence. There are several approaches to the problem [1]-[3].

In many of the methods, the detection decision is based on a hard threshold of some dissimilarity measure, whose value is determined by experimentation. The optimal value depends on the requirements of the application and will be a trade-off between the number of false positives detected and the number of undetected true positives.

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 [4]-[6]. Also, one can use an adaptive threshold setting, by using statistics of the dissimilarity measure within a sliding window [7]-[9].

There are a number of different types of transitions or boundaries between shots [10]. 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.

In general, abrupt transitions are much more common than gradual transitions, accounting for over 99% of all transitions found in video [11]. 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.

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)

The majority of shot cut detection use frame by frame approach, it means pair of successive frames is compared and their similarity is evaluated by selected metric. Subsequently the decision if there is a shot cut is made based on threshold.

The threshold selection has significant influence on the final accuracy of video shot cut detection algorithm. There exist fixed and adaptive thresholds.

Fixed threshold is easy to implement, but achieved lower efficiency due to high amount of missed cuts and false detections. The way how to improve the accuracy of shot boundary detection algorithm is the use of threshold that adapts itself to the sequence statistics.

Other approach, which can be applied on the task of shot cut detection, is employing the similarity evaluation of selected regions of interest [12], [13].

Our goal was to propose an adaptive threshold that will operate in real time for later use in adaptive GOP structure determination with aim to optimize video encoding efficiency.

We have chosen Dugad model [7] and propose its modification for real time adaptive threshold and therefore real time video shot cut detection. We run extensive experiments to verify proposed adaptive threshold and achieved results were compared to results obtained by original Dugad threshold.

This paper is structured as follows: in the second section Dugad model is described. The proposed modification is presented in the third section. Fourth section provides information about used video shot cut detection model and employed measures for evaluation the similarity of compared frames. Fifth section includes experimental results and all achievements are summarized in the conclusion.

II. DUGAD MODEL

Dugad et al [7] proposed a local threshold method whereby the frame difference of successive M frames is examined. The length of sliding window M has to be odd number, as the middle frame is examined and the threshold value is determined from left or right half of sliding window, in other words the means and standard deviations from either side of the middle sample in the window is calculated.

The middle sample represents a shot change if the conditions below are simultaneously satisfied:

1. The middle sample is the maximum in the window
2. The middle sample is greater than threshold m_T

The threshold m_T is calculated as follows:

$$m_T = \max(\mu_{left} + T_d\sqrt{\sigma_{left}}, \mu_{right} + T_d\sqrt{\sigma_{right}}) \tag{1}$$

where μ stands for the mean, σ for standard deviation and T_d is determined based on carried experiment for each evaluation metric.

The decision threshold m_T is recalculated for each new frame and a decision made. However, after a shot cut is detected, no new decisions are made until *M/2* frames have elapsed.

III. Proposed Modification of Dugad Model

Our aim was to propose adaptive threshold scheme for use in real time. Original Dugad model would cause some delays, because it examines frames in the left and right half of defined sliding window. Therefore we can examine frame only when we know the result of similarity evaluation for next *M/2* frames.

We propose a modification, which removes this limitation. We examine the last frame in the window, not the middle one. Thus there is no limitation on windows length, it can be both odd and even number.

The last sample in the sliding window represents a shot change if the conditions below are simultaneously satisfied:

1. The last sample is the maximum in the window
2. The last sample is greater than threshold m_T

The threshold m_T is calculated as follows:

$$m_T = \mu + T_d \sqrt{\sigma} \tag{2}$$

where μ stands for the mean, σ for standard deviation and T_d is determined based on carried experiment for each evaluation metric.

The decision threshold m_T is recalculated for each new frame and a decision made. However, after a shot cut is detected, no new decisions are made until *M/2* frames have elapsed.

IV. Video Shot Cut Detection Algorithm

We have used standard frame by frame approach, it means two successive frames are compared and their similarity, respectively dissimilarity is evaluated by chosen metric. The value of similarity measure is calculated separately for each color component Y, U and V and the final value is achieved as the average of components values.

Subsequently the decision is made based on achieved metric value and selected threshold. If the measure value is higher (or lower in case of metric with reverse logic - higher value achieved by measure means higher similarity of compared frames) than the value of threshold, the frames are classified as cut frames, in second case as non-cut frames.

We have employed following metrics as similarity measure: mean square error (MSE), Pearson correlation coefficient (PCC), mean sum of absolute differences (MSAD) and mutual information (MI). These metrics are described below.

The accuracy of video shot cut detection for all metrics and both original Dugad model and proposed modification was evaluated by standard evaluation metrics: precision (p), recall (r) and F1 score measure (F1) [10]. For all metrics, the range of values is from 0 to 1, where 1 (100%) means the highest possible accuracy of shot cut detection algorithm.

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 value of recall descends with ascending number of missed cuts.

The precision measure is defined as the ratio of correct experimental detections over the number of all experimental detections. In other words, the value of precision is lower with higher amount of false detection.

F1 score measure is a combined measure that results in high value if, and only if, both precision and recall result in high values. F1 score gives more global look at the accuracy of examined shot cut detection algorithm, because it takes into account both missed cuts and false detections.

A. MSE

MSE is the simplest and the most widely used full-reference quality metric. The MSE can be calculated for two images as follows [14]:

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

B. PCC

In statistics, the Pearson's correlation coefficient has been widely employed to measure the correlation (or strength of linear dependence) between two variables X and Y [15]. 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:

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

where Xm and Ym stand for mean pixel intensity of images X and Y.

As the range of values for PCC is from -1 to 1 we used absolute value of PCC as for shot cut detection is not important the direction of similarity. Shot cut detection using PCC has reverse logic, because higher value means higher similarity of compared frames.

C. 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 [12]:

$$MSAD = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left|X(i,j)-Y(i,j)\right|}{M.N} \tag{5}$$

D. MI

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 [10]:

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{6}$$

MI is also metric (as PCC), which has reverse logic for classification of examined frames to cut or non-cut.

V. Experimental Results

We have verified the efficiency of proposed modification of Dugad model through extensive test experiment. As dataset we have used video sequences from the database of project TRECVID. Dataset contains video sequences in CIF resolution (352 x 288 pixels) with 420 abrupt cuts.

As similarity measures we have employed MSE, PCC, MSAD and MI. For MSAD and MSE conditions described in section II and III are valid as these metrics reach the higher value for higher dissimilarity. But for PCC and MI we have to applied the reverse logic, because for similar frames they achieve the highest values.

Therefore for Dugad model the minimum of left and right half of window is chosen as threshold m_T and the examined middle frame has to be minimum in sliding window and it value has to be below calculated threshold. Analogically for proposed modification the examined last frame has to be minimum in monitored sliding window and its value has to be lower than threshold value.

Obtained results were classified by Dugad model and proposed modification for real time processing. The length of sliding window and value of T_d were set according to optimal values found by experiment. Efficiency for both adaptive threshold schemes was evaluated by standard evaluation criterions: precision, recall and F1. Achieved results were compared to evaluate accuracy of proposed modification.

Table I shows the results of video shot cut detection algorithm where Dugad adaptive threshold were used for classification of frames. Table II contains the results for case, where the proposed modification was employed. In all tables first column (SM) represents used similarity metric for evaluation, C stands for correctly detected cut, M represents missed cuts, F is for false detections, p stands for precision, r represents recall and F1 is for F1 score measure.

According to results displayed in Table I, the lowest accuracy was obtained for shot cut detection using MSAD as similarity measure for evaluation. It is caused by low differences between cut and non-cuts values, what led to 33 missed cuts and 102 false detections. Therefore the accuracy evaluated by F1 score measure reached value around 85%.

Shot cut detection algorithm with PCC reached no missed cut, but 9 false detections, thus the value of F1 measure is nearly 99%. Algorithm that uses MI resulted in only 2 false detections, but it missed 6 cuts. Its value for F1 score is about 98%. The best result at all was achieved by method, which employs MSE as similarity measure. Final result for this method is no missed cut and 6 false detections. The value reached by F1 score measure is more than 99%.

TABLE I
Shot Cut Detection Accuracy With Dugad Threshold

SM	C	M	F	p	r	F1
MSE	420	0	6	0.98592	1	0.99291
PCC	420	0	9	0.97902	1	0.9894
MSAD	387	33	102	0.79141	0.92143	0.85149
MI	413	7	8	0.981	0.98333	0.98216

Table II displays the accuracy results for proposed modified Dugad model. The lowest accuracy in terms of F1 score nearly 94% was obtained by MSAD. This algorithm missed 24 cuts and found 29 false detections. For MI the reached F1 score is nearly 97%, because there is 5 false missed cuts and 22 false cuts. Correlation metric PCC achieved the accuracy 99.5% with 1 missed cut and 3 false detections. The best result was obtained by MSE, where algorithm found only 1 false detection and missed 1 cut.

TABLE II
Cut Detection Accuracy With Modified Dugad Threshold

SM	C	M	F	p	r	F1
MSE	419	1	1	0.99762	0.99762	0.99762
PCC	419	1	3	0.99289	0.99762	0.99525
MSAD	396	24	29	0.93176	0.94286	0.93728
MI	415	5	22	0.94966	0.9881	0.96849

According to comparison of results achieved by original adaptive threshold scheme and the proposed modification, we can state, based on results in Table I and Table II, the proposed modification reached in general higher accuracy. The only case, where it showed worse results is shot cut detection algorithm employing MI. Original Dugad model had 7 missed cuts and 8 false detections, the proposed modification 5 missed cuts and 22 false detections. In terms of F1 score it is 98% for original and 96.8% for modified adaptive threshold scheme. For the rest of simulated metrics better results were obtained by proposed threshold. The most significant improvement (more than 8%) is for MSAD.

If we exclude MI measure, the results obviously showed that proposed modification of existed original adaptive threshold scheme is able to improve efficiency of video shot cut detection algorithm and achieves high accuracy.

Fig. 1 shows the graph of shot cut detection, where original Dugad model detected false cut, but with proposed modified adaptive threshold this false detection wasn't found. The place where curve of Dugad threshold cut the curve of shot cut detection is the position of false detection (point 3752 on x axis.)

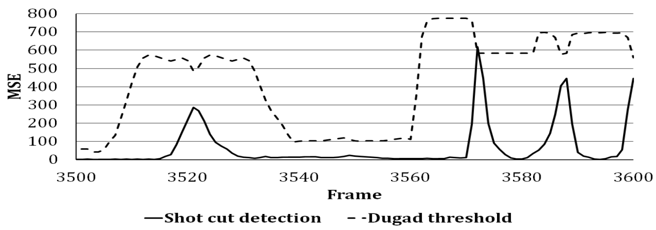

Fig. 1 Shot cut detection with found false detection Dugad model, which was suppressed by proposed modification

The frames classified as cut according to graph in Fig. 1 are shown in Fig. 2. We can see it is not change of shots, there is only larger movement of object between two frames.

Fig. 2 An example of false cut detected by Dugad model, which was suppressed by proposed modification

VI. Conclusion

In this paper we have proposed a modification of existed adaptive threshold scheme for real time processing for later use in determining adaptive GOP structure during encoding process with aim to improve the video encoding efficiency.

Proposed modified threshold model doesn't cause any delays, because it uses only previously evaluated frames for adapting itself to the sequence statistics and subsequently for classification of the examined frames as cut or non-cut.

We have verified the efficiency of proposed algorithm through extensive test experiment with several similarity metrics used for comparing of examined frames. Achieved results were evaluated by standard criterions and compared to original Dugad model.

According to obtained results, the proposed modification achieved really high accuracy and is able to improve efficiency of original threshold.

Acknowledgment

Research described in the paper was financially supported by the Slovak Research Grant Agency (VEGA) under grant No. 1/0602/11.

Sound and Vision video is copyrighted. The Sound and Vision video used in this work is provided solely for research purposes through the TREC Video Information Retrieval Evaluation Project Collection.

References

[1] J. S. Boreczky and L. A. Rowe. "Comparison of video shot boundary detection techniques." *Journal of Electronic Imaging*, 5(2):122–128, 1996.

[2] H. Jiang, A. S. Helal, A. K. Elmagarmid, and A. Joshi. "Scene change detection techniques for video database systems," *ACM Multimedia Systems*, 6(3):186–195, 1998.

[3] Y. Yusoff,W. Christmas, and J. Kittler. A study on automatic shot change detection. In D. Hutchison and Ralf Schafer, editors, *Proc 3rd. Eur. Conf. on Multimedia Apps, Services and Techniques* (ECMAST), pages 177–189. Springer, May 1998.

[4] 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". In *Proc. IEEE Int. Conf. on Image Proc.*, volume 2, pages 884–887, Oct 1998

[5] 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.

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

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

[8] 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, Dec 1995.

[9] 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.

[10] 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.

[11] 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

[12] E. Šikudová, M. Gavrielides and I.Pitas, "Extracting semantic information from art images," *Computer Vision and Graphics : International Conference, ICCVG 2004*, pages 394-399, Warsaw, 22.-24.9.2004.

[13] M. Gavrielides, E. Šikudová, and I. Pitas: "Color-based descriptors for image fingerprinting," IEEE *Transactions on Multimedia*, Vol. 8, No. 4 (2006), s. 740-748.

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

[15] 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.