Shot Boundary Detection In Videos Sequences Using Motion Activities

Tabii Youness *youness.tabii@gmail.com*
ENSA
Abdelmalek Essaadi University
Tétouan, Morocco

Sadiq Abdelalim *sadiq.alim@gmail.com*
Computer sciences departement
Faculty of sciences
Ibn Tofail University, Kenitra, Morocco

Abstract

Video segmentation is fundamental to a number of applications related to video retrieval and analysis. To realize the content based video retrieval, the video information should be organized to elaborate the structure of the video. The segmentation video into shot is an important step to make. This paper presents a new method of shot boundaries detection based on motion activities in video sequence. The proposed algorithm is tested on the various video types and the experimental results show that our algorithm is effective and reliably detects shot boundaries

Keywords: Information Retrieval, Shot Detection, Video Segmentation, Motion Vector, 2D Variance.

1. INTRODUCTION

In recent years, thanks to the rapid growth in multimedia information and the advance in internet communication, multimedia information indexing and retrieval has become more and more important. Multimedia information contains audio and visual data in addition to text information. Although, many research efforts have been devoted to the video retrieval and video analysis based on audio or visual features. This analysis show that, when developing retrieval applications and video indexing, we first have to consider the issue of structuring the huge and rich amount of heterogeneous information related to video content. In addition, to retrieve information from the audio or visual content is a very challenging since it requires the extraction high-level semantic information from Low-level audio or visual data.

In order to achieve a reliable video description, the requirement is to structure the video into elementary shots. This structuration consists of detecting transition effects between homogenous segments (shots). This video partitioning step enable us to provide content-based browsing of the video and should facilitate higher-level tasks such as video editing, indexing and information retrieval. For ease of reference, we have to provide brief definition of the different kinds of shots boundaries. A cut is an abrupt transition between two shots that occurs between two adjacent frames. A fade is a gradual change in brightness, either starting or ending with a black frame. A dissolve is similar to a fade except that it occurs between two shots.

In [1], Bescos analyses several frame disparity functions, i.e. functions which measure frame dissimilarities. Deterministic (Summation of absolute differences), statistic parametric (likelihood ratio test) as well as statistic non-parametric disparity functions are considered. The authors choose two metrics which undertake the best divergence between the "cuts" and "noncuts" classes and computes a third features which uses a small sliding window of size 1. Furthermore, a simple supervised parallelepipedic classifier is applied. The following results are reported for s subset of MPEG-7 test set (2074 cuts): 99% recall and 95% precision.

Chua et all. [2] propose a unified approach to detect cuts and gradual transition by using a temporal multi-resolution approach. This method affected by applying a wavelet transform to frame dissimilarity measures. They use histogram differences as well as coarse representation of MPEG motion vector. First, they detect candidates from the set of local maxima and they apply an adaptive threshold technique. Finally, they use support vector machines via active learning to find an optimal hyperplane to separate cuts and non-cuts.

In [3], the authors present an algorithm shot changes the discreet cosines transform (DCT), they calculate the DCT of the luminance matrix by blocks of 8x8, then the two distances between the neighboring pixels (vertical and horizontal distance). The only threshold for the shot changes is that the average of vertical and horizontal distances is superior to 1/2. Authors in [4], enhance and optimize this last method using DCT multi-resolution. A shot boundary detection algorithm is based on the same assumption: visual discontinuity between consecutive shots, this discontinuity have been aimed to detect via various Visual descriptors such as color histogram [5].

Park et al [6]. Used an object recognition algorithm, namely SIFT [7], in order to exploit the similarities between frames and hence detect shot boundaries. Their assumption was that if certain amount of change belongs to the same shot. They compared consecutive frames for detecting abrupt shot changes (hard cut) and non-adjacent frames with a fixed distance apart for detecting gradual transition. However their method considerably suffered on the heavy computational cost of the SIFT algorithm. Moreover, by relaying on the difference between adjacent frames with a fixed threshold, their accuracy is relatively low under high motion and sudden illumination changes.

2. SHOT BOUNDARY DETECTION METHOD

In video sequences, the motion activity is one of the most features included in the visual information. It also used to describe the level of activity, action, or motion in that video sequence.

In this paper, we propose that the low or high motion show how much a video sequence is changing and can be measured and present a shot transition. Figure 1 show the stages of our method for shot boundary detection in video sequence. The first step consists the extraction of the intra-frames (I-frames) of the whole video, frame by frame. The transformation of frames extracted to luminance is the second step. After, the motion vector calculation using diamond search algorithm (the motion vectors are considered as random variables). Next, the calculation of 2D variance. Finally, using threshold, we detect shot boundaries and we generate the shots sequences.

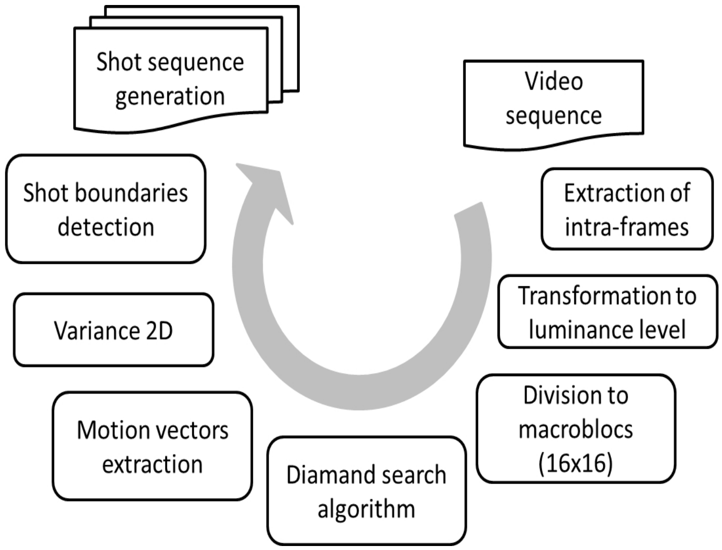

FIGURE 1: Stages of our shot boundary detection method using motion activity.

3. MOTION VECTOR

The motion activity can be used to describe the video content using the level of this motion. This description is used in various applications such as dynamic video summarization, video surveillance, fast browsing, content-based querying, shot classification or scene segmentation.
In our case we will use a different kind of videos like news, action films, cartoon and sports and we propose that the low or high motion activity shows how much the video is changing, this property allows us to detect transition between shot using motion vector (MV).

In order to extract motion vector, we use the Diamond Search (DS) algorithm. The DS algorithm is proved to be the best block matching algorithm of the last century [8]. Using of this motion vector to compute the variance 2D that allow as to locate the set of frames which are stable between them and candidate to be one shot (belongs to the same shot).

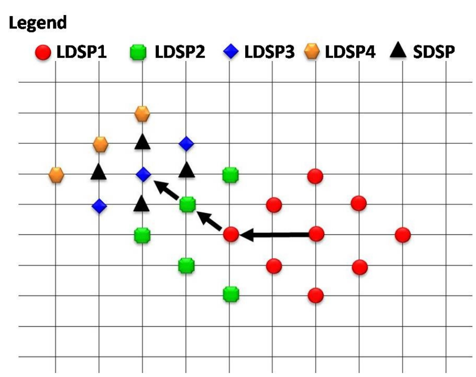

FIGURE 2: DS Algorithm.

The figure 2 shows the large diamond search pattern and the small diamond search pattern. It also shows an example path to motion vector (-4, -2) in five search steps four times of LDSP and one time of SDSP.

To compute motion vector, there are various cost functions, the most popular and less computationally expensive is *Mean Absolute Difference* ***(MAD)*** given by Equation (1). Another cost function is *Mean Squared Error* ***(MSE)*** given by Equation (2) which are used in this work.

$$MAD = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|C_{ij} - R_{ij}\right| \qquad (1)$$

$$MSE = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(C_{ij} - R_{ij}\right) \qquad (2)$$

Where ***N*** is the size of the macro block, $\boldsymbol{C_{ij}}$ and $\mathbf{R_{ij}}$ are the pixels being compared in current macro block and reference macro block, respectively. Next step is to calculate the variance 2D of the motion vectors.

4. 2D VARIANCE

In this section we will present how to compute the variance in two dimensions of motion vector. We define the 2D variance as given in Equation 3:

$$\sigma\_i_{t+1}^{t} = \sqrt{\frac{\sum_{l=1}^{M}\sum_{c=1}^{N}\left(MV_i(t+1)_{lc} - MV_i(t)_{lc}\right)^2}{M*N-1}} \qquad (3)$$

where ***M*** and ***N*** are the height and the width of the matrix ***MV*** respectively, and ***i*** refers to the samples number ***i***.

When we get the ***2D*** variance vector of each block of frame during the whole video sequence with sampling (just intra-frames), then we compute the mean variance (Equation 4):

$$\sigma\_i_{mean} = \frac{1}{2*S*p}\sum_{t=1}^{k}\sigma\_i_{t+1}^{t} \qquad (4)$$

In Equation (4), ***S*** represents the number of samples and ***k*** is the number of macro blocks in DS algorithm and ***p*** is the search parameter.

The last step in searching of the static set of frames that are candidate to be one coherent shot, we use the Equation (5) for this purpose.

$$\begin{cases} if \quad \sigma_{mean}^{i} < T_{thl} \quad then \quad NonTransition \\ \quad else \quad ShotTransition \end{cases} \qquad (5)$$

Where $\boldsymbol{T_{thd}}$ is the threshold. In section of experiments we will show the obtained results of the proposed method.

5. EXPERIMENTAL RESULTS

We put on the test a different kind of video to show the robustness of our method. Figure 3, exhibit the frames taken from our videos sequences database.

News

Action film

Sport

Cartoon

FIGURE 3: Frames from video sequence used in experiments.

In table 1 we present more details about the video database. The number of frames and the number of transition which are extracted manually to compare them we our algorithm results.

Videos	Length (minutes)	Number of frames	Number of transitions
News	5.43	8498	24
Action film	4.30	6842	17
Sport	7.50	11953	65
Cartoon	4.20	6654	36
Total	22.23	33947	125

TABLE 1: Video Database.

Figure 4 display the result obtained of algorithm applied on ***500 I-frames*** that is about ***6000 frames*** and ***4 minutes***. The results show peaks that present the shot changes in clip (Action film). Also as shown in this histogram of 2D variance mean (Equation 5), the threshold taken for detection shot change is $T_{thd} = 1/2$.

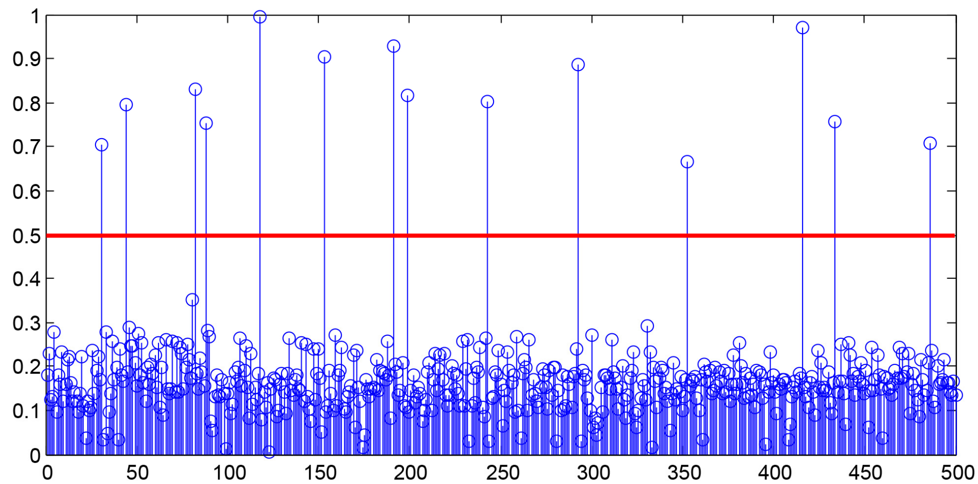

FIGURE 4: Result for 500l frames (≈ 4 minutes). Video : Action film.

The performance of a shot boundary detection system is measured in terms of precision π and recall ρ : the first is the fraction of frames identified as shot boundaries by the system that correspond to real shot boundaries; the second is the percentage of real shot boundaries that have been detected as such by the system. In mathematical terms, if R_d is the set of frames that the system claims to be shot boundaries and R_t is the set of frames that correspond to real shot boundaries, then the precision is defined as follows:

$$\pi = \frac{|R_d \bigcap R_t|}{|R_d|} \tag{6}$$

Where $|.|$ is the cardinality of the set.

The recall is defined as:

$$\rho = \frac{|R_d \bigcap R_t|}{|R_t|} \tag{7}$$

The table 2 shows the result we obtained for shot detection. The detection rate in News, Sport and cartoon is high, but for Action is relatively low. This is maybe due to the edition style and a motion activity which is differing from a video to another. However our algorithm works satisfactorily.

Video	Detected	Correct	False detection	Missed	Precision (%)	Recall (%)
News	23	21	2	3	91.30	87.5
Action film	18	16	2	1	88.89	94.11
Sport	61	56	5	9	91.80	86.15
Cartoon	35	32	3	4	91.42	88.89

TABLE 2: Experimental Results.

6. CONCLUSION

In this paper we presented a new method for the shot detection based on motion activity using variance 2D applied to different domain of video. The main contribution of the presented work is an algorithm for abrupt shots detection. The advantage of our algorithm is clearly seen in its simplicity and effectiveness in providing better results for the detection of the majority of shots. Besides, the analysis of video on the basis of motion segmentation is very promising.

7. REFERENCES

[1] J. Bescos. “Real time shot change detection over online mpeg-2 video.” IEEE Transactions on Circuits and Systems for Video Technology, vol. 1(4), pp.475-484, April. 2004.

[2] T.S Chua and H.M. Feng and C. Anantharamu. “An unified framework for shot boundary detection via active learning”. In Proc. ICASSP, 2003, pp. 845–848.

[3] H. Koumaras and G. Gardikis and G. Xilouris and E. Pallis and A.Kourtis. “Shot boundary detection without threshold parameters.” Journal of Electronic Imaging, vol. 15(2), April. 2006.

[4] Y. Tabii and R. O. H. Thami. “A new method for soccer shot detection with multi-resolution dct,” CORESA’07 Compression et REprsentation des Signaux Audiovisuels, France, 2007.

[5] U. Gargi and R. Kasturi and S.H. Strayer. “Performance characterization of video-shot-change detection methods.” Circuits and Systems for Video Technology, IEEE Transactions, vol. pp. 1-13, 2002.

[6] M. H. Park and R. H. Park and and S.W. Lee. “Shot boundary detection using scale invariant feature matching. ” In Proc. SPIE Visual Communications and Image Processing, 2006, pp. 569–577.

[7] D.G. Lowe. “Distinctive image features from scale-invariant keypoints.” International journal of computer vision, vol. 60, pp. 91-110. 2004.

[8] A Barjatya. “ Block matching algorithms for motion estimation,” DIP 6620 Spring 2004 Final Project Paper, 2009.