Int. J. Nonlinear Anal. Appl. 10 (2019) No. 2, 131-140
ISSN: 2008-6822 (electronic)
https://dx.doi.org/10.22075/ijnaa.2019.4180

Application of 2D fractal dimension in content based video summarization

Hadi Yarmohammadi[a,*], Hossein Marvi[b], Hamid Hassanpour[a]

[a]*Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran*
[b]*Faculty of Electrical Engineering, Shahrood University of Technology, Shahrood, Iran*

(Communicated by M.B. Ghaemi)

Abstract

In this paper, a novel method for video content summarization has been proposed by calculating the fractal dimension of frames. Summarization of the video is the first step in automatic video analysis. In this paper, we use the support vector machine (SVM) and the decision - tree to identify the shot boundary and classify them. In order to compute the fractal dimension, the numerical method has been expressed. The results of this implementation were also reported on TRECVID 2006 data collection. The results show that the relative advantage of the method presented in this paper is compared to other articles.

Keywords: Video summarization, Shot boundary detection, Key frame extraction, Support Vector Machine.
2010 MSC: 68T45

1. Introduction

With the availability of digital video production tools such as smart phones and closed circuit cameras, Huge array of personal and commercial video archives have been created. quick accessing to the information and content of the video archives requires effective tools and methods. Therefore, the automated video analysis is one of the basic requirements of today's archives. video summarization is the first step in the automated video analysis, so in this paper a novel method is presented based on the fractal dimension of video frames to summarizing the video. the first and most important step

*Corresponding Author
Email addresses: hadi.yarmohammadi@shahroodut.ac.ir (Hadi Yarmohammadi), h.marvi@shahroodut.ac.ir (Hossein Marvi), h.hassanpour@shahroodut.ac.ir (Hamid Hassanpour)

Received: June 2017 *Revised:* September 2018

in the c video summarization is the shot boundary detection. if the detection of the shot boundary is properly done, the performance of the summarization system will also increase dramatically.

Shot is a "comprehensive entity" that can be used as a video base block that supports a large number of access operations in a high level video. Since the conceptual content of a video is highly dependent on the imaging process, the full segmentation of the video is as an initial step for most of the video processing tasks. According to whether the transmission between shots is a gradual or sudden , the shot boundaries are classified into two sudden transitions and a gradual transition categories . If the two shots are directly connected together, the location of this connection is called a sudden transition, on the ground between the end of the previous shot and the beginning of the next shot there is no gap; on the other hand, if the video frames are integrated based on some specific ways to make the connection more visually smoother, this connection is called a gradual transition. The shot boundary detection, which is also called time segmentation of the video, is the process where the transition between adjacent shots is identified. so far, many methods have been proposed for the shot boundary detection.

2. Related work

shot boundary detection is very important in automated video analysis, so several methods have been proposed for shot detection. In [1] a method has been introduced based on edge information and histogram. This system has also given an application for embedded systems. In [2] a method has been proposed found on local and global dissimilarity criterias. Efficiency of this system has been proved in some TV shows. In [3] a method also has been reported based on mutual information. In this method mutual information in neighbor frames of a video is calculated. Then mutual information signal which is corresponding to the entire mutual information of neighbor frames of a video is extracted. In this signal minimum values of mutual information indicates the neighbor frames that have least similarity and they can be chosen as boundary frames of two different shots. In [4] a technique has been published based on fuzzy sets. In order to achieve high accuracy in boundary shot detection , low-level selected feature is essential. But in a frame or a video there are many features like pixel values , different color channels , static features , color histogram and etc. So that , as selected features for showing a shot or a video is more suitable , calculation costs decreases and performance increases. To reach this goal , the procedure of efficient selection is based on fuzzy sets, has been reported. For shot boundary detection, 12 candidate features which are classified in 5 types usually are extracted for typical applications. Also in [5] a method has been proposed based on hidden Markof model. It is used three video segmentation features In this model which are: 1.histogram, 2.measure sound distances and 3.estimation of object movements are used between two neighbors frames.

Histogram measures distance between two neighbor frames based on distribution of brightness levels which in those points includes 64 classes placed on brightness degree. Sound distance which is measured each 20 milisec by using its first transform to sequence of SEPTAL vectors , is calculated.

Also color histogram as an efficient method which has a suitable performance has been published in [6]. But this algorithm faces a problem in selecting identical boundaries for shot detection in big pictures because of using vast set of different features and its needed to develop in order to boundary adaptation.

Histogram technique uses three 64-bit histogram(one histogram based on brilliance and two histograms based on color distribution) then these three histograms are concatenated to each other and N-dimensional vector is created. Which N corresponds to the sum of entire units in these three histograms. In this technique by using of cosine function we will be able to compare different frame

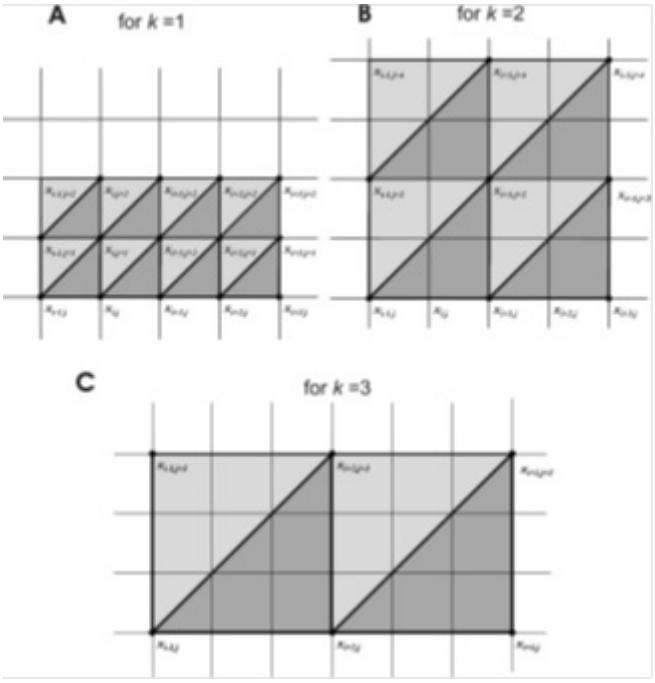

Figure 1: dividing main shape to the basic shape[10]

histograms. In [7] for achieving shot boundaries of a video a searching pattern called octagon square has been used. In first step features such as color, edge, movement and etc. are extracted for finding shots. Then by using of movement feature and octagon square searching pattern it can detect a shot. in [8,9] convolution neural network and deep learning model have been used in order to detect shots range.

3. CALCULATING THE FRACTAL DIMENSION

In this paper, fractal dimension is used to determine the similarity criterion between two frames. Fractal dimension shows the complexity of a frame. If there is no changes in shots, fractal dimension will not vary that much for neighbor frames. Despite this, for the frames which are between the shot boundaries, fractal dimension will differ obviously. In order to calculate the fractal dimension we have used the proposed method in [10]. As it is shown in figure. 1 , each frame is divided into basic shapes and sum of each shapes area are counted. In next step, size of basic shapes are doubled and sum of areas are calculated again. This procedure will be continued by doubling the size of each shape until it be as big as the size of the frame. By multiplying the sum of these areas by a constant factor, A(k) , the fractal dimension will be as follow:

$$D = \frac{\ln A\,(k)}{\ln \frac{1}{k^2}} + 1 \tag{3.1}$$

Where D represents fractal dimension and A(k) is the sum of areas of the basic shapes in kth level.

One of the problems in the proposed algorithm in [10] is ignoring the size of the basic shapes. In this method, the gray value difference between two neighbor pixels are considered as the distance between these pixels. Because of this, In most of the neighbor pixels, the size of the line will be zero. Consequently, the area of the basic shapes will be turned to zero. To avoid this problem In this paper, it is attempted to use a constant value when the area of the basic shape is zero. The

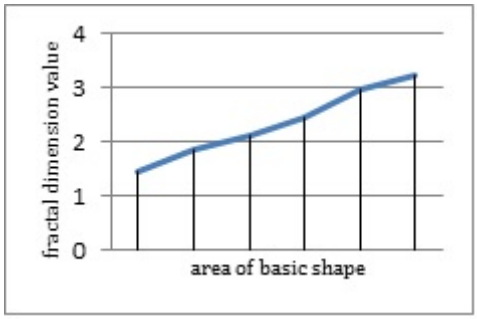

Figure 2: fractal dimension value for different area of basic shape

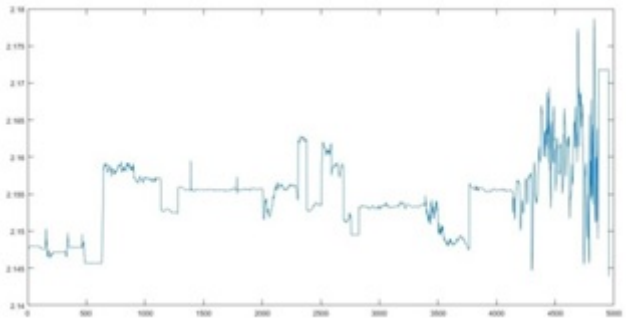

Figure 3: signal of fractal dimension

result of this work is shown in figure.2. According to figure.2 it is determined that the area of the basic shapes shouldnt be less than 0.7 and more than 1.5. Because in this condition, the calculated fractal dimension will not be acceptable. In this paper, If Euclidean distance between two neighbor pixels be zero, area of that basic shape will be considered one as default.

After calculating the fractal dimension of each frame and collecting these numbers, we get a signal which represents the whole fractal dimensions of the video. In Figure.3 one these signals are represented.

4. WAVELET TRANSFORM

By using the wavelet transform, we can get a good perspective of the signal and also see the difference between values and base line. For this purpose, the wavelet transform is applied on the signal in Figure.3 and it is shown in Figure.4.

5. Windowing

After the applying of Violet transform, we perform an overlapping window on the signal from it. We also consider the size of each window 50. The purpose of this windowing is to convert the signal into smaller segments, Which can make a more accurate comments on the shots. After the windowing, the signal positioned under each window is given as input to the support vector machine.

6. Support vector machine and decision tree

The training data for the support vector machine is the sub-signals obtained from the signals witch extraction of fractal dimension from the frames. For the detection of shot or absence of a

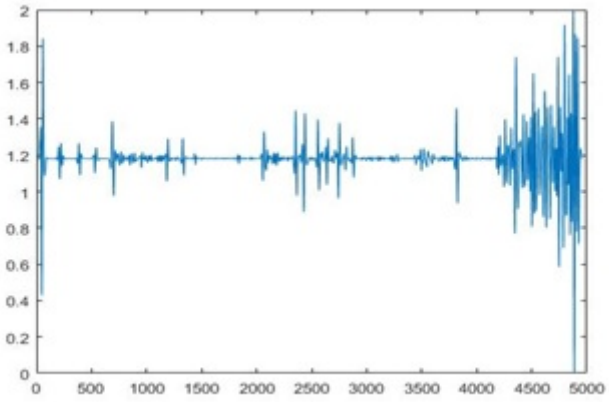

Figure 4: wavelet transform of fractal dimension signal

signal which located in a window, we refer it to the support vector machine as the test input. The support vector machine using the training data it has already observed will be able to determine whether the sub-signal is a shot or not After being shot confirmed by the support vector machine, we use the decision tree to specify the type of shot. The proposed system in this paper is able to detect two types of shot transitions gradual transition and sudden transition In order to better understand the proposed system, the flowchart of this system is presented in Figure.5.

7. KEY FRAME EXTRACTION

In this paper, a static video summarization system has been proposed. In this system ,in order to summarize a video file, after shot detection, key frames are extracted over the shots. Extraction means finding a frame which is independent of other shot frames. Input of this procedure is a shot. In this step, fractal dimension is used for finding the key frames like the pervious section. For this purpose, we extract the fractal dimension of each frame in shot .Result of this step is a vector with N length which N is the length of shot. Then after sorting this vector in a Descending way, we extract first m values as key shot frames. m depends on the length of the shot and it can differ for each shot. In this paper we have used summarization value of 1/30 which means 1 of 30 frames is chosen as representative. The reason of using these values is in TRECVID2006 database all videos have frame value of 30 per second. In fact with use of summarization values we extracted a representative frame over each second. Experiments show that if there arent some fast movements in a video, this value will have good performance in extracting the key frame. To compare the results of the summarization we used fidelity criterion.

8. DATASET

First data used in this research is TRECVID2006, it includes more than 10 hours video wich contains various shots, second data created by authors to evaluate proposed method, this set includes more than 2 hours of video in which videos gathers from sport and financial news and movies. Table 1 shows the video parameters and number of shots occurred in each of them.

Table.1 shows the video set parameters and types of shot changes in these set. TrecVid can be accessed by nist.gov website. Sample frames of Dataset shown in figure.6.

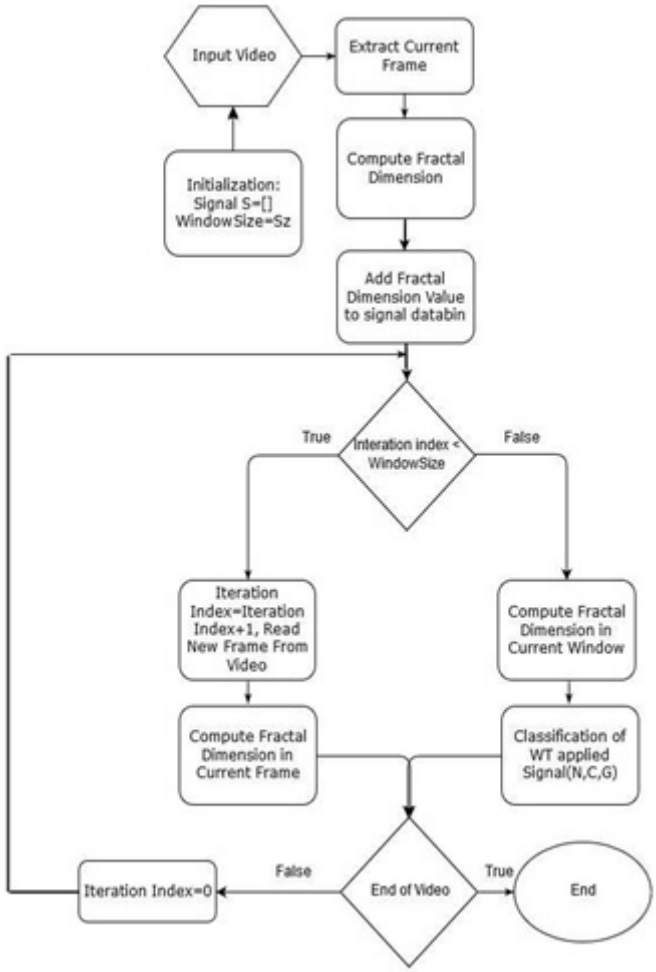

Figure 5: flowchart of proposed method for shot boundary detection

Table 1: dataset

Video Set	Number of frames	Number of shots	Shot types
TrecVid 2006	More than 500000	More than 1000	Abrupt-Gradual(Zoom, Fade-in,Fade-Out)
Our set	430000	508	Abrupt-Gradual(Zoom, Fade in)
Total	More than 900000	More than 1500	-

Figure 6: sample frames from dataset, top row shows the sample frame from our dataset from internet archives, second row shows frame from TrecVid2006

9. Evaluation result

The performance of shot boundary detection and classification can be evaluated using precision ,recall. to evaluate the methods also we used Correct Detection rate.

$$Precision = \frac{correctdetectedchanges}{correctdetectedchanges + wrongclassifiedchanges} \tag{9.1}$$

$$Recall = \frac{correctdetectedchanges}{correctdetectedchanges + undetectedclassifiedchanges} \tag{9.2}$$

$$CorrectDetection = \frac{correctdetectedchanges}{Allshotchanges} \tag{9.3}$$

Table.2 shows the analysis results for proposed method and compare it with the other state of art methods. We show the ROC curve for proposed method which shows the detection capabilities. Our classes for each part of signal can be non, abrupt, gradual(zoom), gradual(fade).

Figure.7 shows sample results on video set, changes in frames also seen in the figure. Precision-Recall Curve shows how is misclassification rate to undetected shots. Figure.8 shows the results using SVM classification method.

To show the accuracy of proposed method we should clarify these method is better than recent methods, table 2 shows the comparison between methods using various metrics. In table 2 results analyzed based on correct detection rate, precision, recall and ability to detect shot sub-types.

Result shows that the proposed method has better precision, recall and detection rate, also few methods able to classify shot changes sub-types. Computational time of proposed method is real-time and dont need major overhead in video analysis.

According to table.2 our method compared with other shot boundary detection and classification, methods detection rate calculate on our datasets by our evaluation, classification ability based on main methods. Figure.9 shows Recall-precision curve . figure.10 shows that our method achieve good result when key frame number is less than 15.

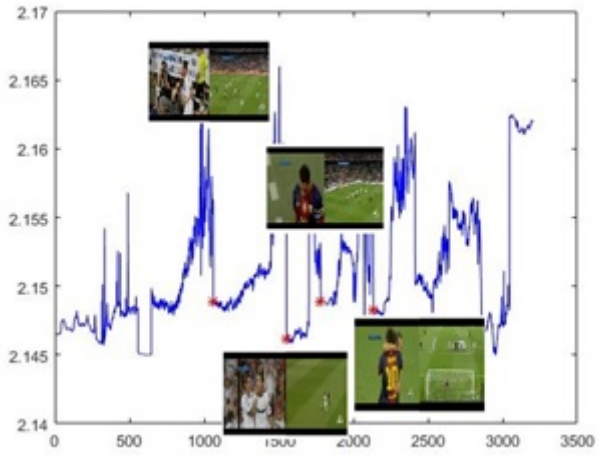

Figure 7: fractal dimension signal for video and detected shots and their frames

Figure 8: sample of extraction shot

Table 2: comparison reuslt

Method	Precision	Recall	Detection Rate	Sub-type classification
Tong et al [11]	98.60	86.90	95	N
Chasanis et al[12]	97.50	98.87	96	Y
Amiri et al[13]	94	96.80	92	N
Bi et al[14]	98.93	98.80	98	N
Proposed	99.02	97.22	99	Y

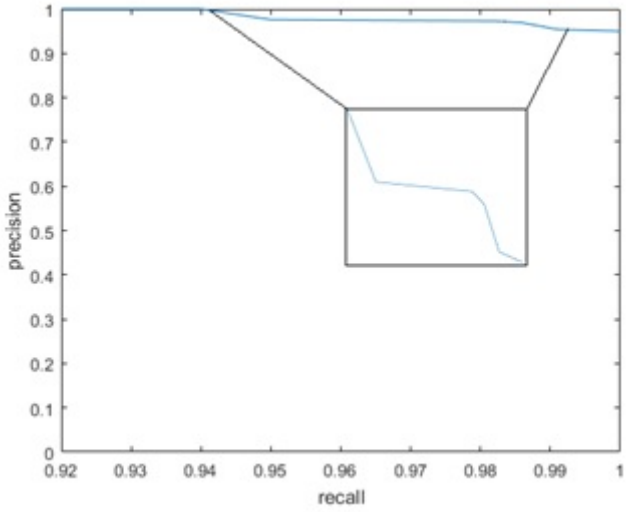

Figure 9: Recall-Precision Curve for shot detection and classification

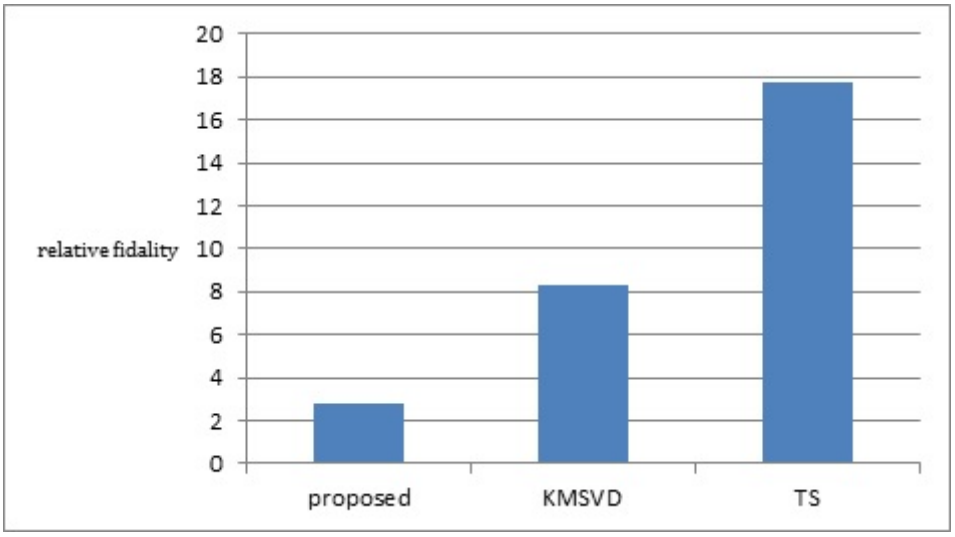

Figure 10: key frame extraction comparision

10. CONCLUSION

In this paper, a method for content based video summarization was presented. This method was based on the computation 2d fractal dimension . in order to ease the work after calculating the fractal dimension of the video, wavelet was used. the support vector machine was used as a tool for shot detection finally, the decision tree determined the type of shots. Key frame extracted based on 2D fractal dimension too. the results show the effectiveness of this system.

References

[1] S. Priyanka, J. Majumdar, S. Kumar, Video Shot Detection on Embedded System, International journal of advanced research in computer and communication engineering, vol. 4, no. 8, 2015.

[2] E. Amini, S. Jafarali Jassbi, A quick algorithm to search and detect video shot changes, vol. 115, no. 3, International Journal of Computer Applications, 2015.

[3] H. Yarmohammadi, M. Rahmati, Sh. Khadivi, Content based video retrieval using information theory, In proc. IEEE Iran Conf. ,Machine vision and Image Processing, pp. 214-218, 2013.

[4] B. Han, X. Gao, H. Ji, A shot boundary detection method for news video based on rough-fuzzy sets, International Journal of Information technology, vol.11, no. 7, 2005.

[5] S. Boreczky and D. Lynn, A hidden markov model framework for video segmentation using audio an image features, In Proceedings of IEEE International conference on Acoustics, Speech and Signal Processing, May 12-15, 1998.

[6] U. Gargi, R. Kasturi and S.H. Strayer, Performance characterization of video-shot-change detection methods, IEEE Transaction on Circuits and Systems, vol. 10, issue.1 2000.

[7] J. Kavitha, S. Sowmyayani, P. Arockia Jansi Rani, Shot Boundary Detection Using Octagon Square Search Pattern, International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:10, No:7, 2016.

[8] Michael Gygli, Ridiculously Fast Shot Boundary Detection with Fully Convolutional Neural Networks, International Conference on Content-Based Multimedia Indexing (CBMI 2018).

[9] Shitao Tang, Litong Feng, Zhangkui Kuang, Yimin Chen, Wei Zhang, Fast video shot transition localization with deep structured modesl, Asian Conference on Computer Vision (ACCV 2018).

[10] Sladjana Spasic, On 2D generalization of Higuchi's fractal dimension, Chaos, Solitons and Fractals, vol. 69, pp. 179187, 2014.

[11] W. Tong, L. Song, X. Yang, H, Qu, R. Xie, CNN-based shot boundary detection and video summarization, Proceeding on IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, 2015.

[12] V. Chasanis, A. Likas, N. Galatsanos, A Support Vector Machine Approach for Video Shot Detection, In Book: New Directions in Intelligent Interactive Multimedia, pp. 45-54, 2008

[13] A. Amiri, M. Fathy, Video Shot Boundary Detection Using generalized eigenvalue decomposition and gaussian transition detection, Journal of Computing and Informatics, vol. 30, pp. 595-619, 2011.

[14] J. Bi, X. Liu, B. Lang, A Novel shot boundary detection based on information theory using SVM, International Congress on Image and Signal Processing (CISP), pp. 512-516, 2011.