Video Shot Boundary Detection based on Multifractal Analisys

Goran J. Zajić, Irini S. Reljin, *Senior Member, IEEE*, and Branimir D. Reljin, *Senior Member, IEEE*

***Abstract* — Extracting video shots is an essential pre-processing step to almost all video analysis, indexing, and other content-based operations. This process is equivalent to detecting the shot boundaries in a video. In this paper we presents video Shot Boundary Detection (SBD) based on Multifractal Analysis (MA). Low-level features (color and texture features) are extracted from each frame in video sequence. Features are concatenated in feature vectors (FVs) and stored in feature matrix. Matrix rows correspond to FVs of frames from video sequence, while columns are time series of particular FV component. Multifractal analysis is applied to FV component time series, and shot boundaries are detected as high singularities of time series above pre defined treshold. Proposed SBD method is tested on real video sequence with 64 shots, with manually labeled shot boundaries. Detection accuracy depends on number FV components used. For only one FV component detection accuracy lies in the range 76-92% (depending on selected threshold), while by combining two FV components all shots are detected completely (accuracy of 100%).**

***Keywords* — Shot boundary detection, multifractal analysis, threshold, feature vectors**

I. Introduction

RECENT rapid development of digital video technology and the low market price of video recording equipment have initiated an explosive growth of video archives. Video recordings, made by any user of video camera usually are not classified and labeled in video archives in an adequate way, while the size of video archives is growing very fast. Manual search is very difficult and time consuming. Searching process in large video collections can be significantly improved by using an automated system based on video content. Boundary shot detection in a video sequence is the initial step in the process of analysis and search of video sequences.

Usually, shot boundaries are characterized by a significant difference between successive frames in a video sequence. Two basic types of shot boundaries are usual: frames can be characterized by sharp changes or by gradual (smooth) changes (such as fade-out/fade-in, dissolve, wipe, and similar effects). The first case is much easier for detection, while the second one presents a challenge for researchers. Detecting of gradual changes is a major problem in shot detection process. The type of changes between frames is not the sole factor which affects the accuracy of detection. Other factors producing possible false detection can be, for instance, unpredictable changes in lighting (flashlight), jitter (panning, zooming, tilting) and sharp changes in the scene (explosion, fire, bleeding, etc.). Due to these factors the BSD algorithm should be more sophisticated, consisting of several procedures enabling more accurate shot detection.

A number of algorithms for video shots detection have been proposed in recent time. The easiest approach is the comparison between adjacent frames by the value of intensity in gray scale [1]. A modified approach [2] is counting pixels whose intensity or color are modified significantly. These methods are useful only for simple video sequences because they are very sensitive to lighting changes and moving of cameras and objects. An improved method based on the comparison of similarities between blocks is presented in [3]. This approach avoids false categorization of small moving objects and/or camera movement as shot changes. The next step in reducing the detection of moving cameras and objects is the comparison of grayscale or color histograms as global statistical information about an image [4]. In BSD algorithms for detection of gradual shot changes some other low-level features are used, for instance the information related to edges [5] and the vector of movement [6].

In addition to these algorithms, series of algorithms for shots detection based on machine learning are proposed: application of unsupervised clustering [7], Hidden Markov model (HMM) [8], neural networks [9] and the application of Support Vector Machine (SVM) [10]. There are also some other approaches based on the clusterization of frames in the same video shot having similar spatial-temporal information. [11], [12]

In this paper the detection of shot changes in a video sequence using multifractal analysis is proposed. Each frame in a video sequence is represented by its feature vector, which consists of components describing color and texture features. Feature vectors are stored in a feature matrix whose rows correspond to frames and the columns correspond to the components of feature vector. In this way, each column is the time series of corresponding feature vector component. Multifractal analysis was applied to these time series, in order to detect the position of shot changes in the time domain.

Goran Zajić is now with the ICT College of Vocational Studies Belgrade, Serbia (e-mail: goran.zajic@ict.edu.rs).

Irini Reljin is now with the Faculty of Electrical Engineering, University of Belgrade, Serbia (e-mail: irinitms@gmail.com).

Branimir Reljin is now with the Faculty of Electrical Engineering, University of Belgrade, Serbia (e-mail: reljinb@etf.rs).

The paper is organized as follows. After the Introduction, Section II gives a brief description of multifractal analysis and Section III describes pre-processing of video sequence. Section IV shows the methodology that was used in the detection of shot changes. The results of applying multifractal analysis to detecting shot changes are presented in Section V. Concluding remarks and directions for future research are given in Section VI.

II. Multifractal Analysis

Multifractal analysis (MA) presents the way of describing irregular objects and phenomena, which were conceived and developed by Benoit B. Mandelbrot, and then applied to problem solving in many fields of science [13]. Multifractal formalism is based on the fact that the highly nonuniform distributions, arising from the nonuniformity of the system, often have many scalable features including self-similarity [14]. Apart from studying the so-called long-term dependence (long range dependency), dynamics of some physical phenomena and the structure and nonuniform distribution of probability, the MA can be used for characterization of fractal characteristics of the results of measurements.

Multifractal analysis studies the local and global irregularities of variables or functions in a geometrical or statistical way [15]. Multifractal formalism describes the statistical properties of these singular results of measurements in the form of their generalized dimensions (local property) and their singularity spectrum (global) [14]. There are several ways to determine the multifractal parameters and one of the most common is called box-counting method. This method is based on the following procedure. Cover the time series of measurements by the squares with their side dimension l, and count those squares which contain some particular measurement value $\mu_i(l)$ and can be interpreted as the probability that the value of measuring μ is in the i-th square. It is shown that the measured value corresponds to the so-called power law

$$\mu_i(l) \approx l^{\alpha_i}, \qquad (1)$$

where exponent α_i describes the fractality of the structure, and is known as a rough Holder's exponent. Using different sizes of squares different values of the exponent α will be obtained, but they will converge on the common values of α, for a given measure μ, in the boundary process, when $l \to 0$. Parameter α depends on the position in the structure and describes its local regularity. Certainly, in the entire structure there is a number of squares (points in the boundary process) with the same value of the parameter α. Therefore, we can observe the distribution of this variable in the structure, which is known as Multifractal (MF) spectrum $f(\alpha)$ and provides a description of the global regularity. MF spectrum can be determined by box-counting method. If we count all the squares $N(\alpha)$ where Pi is the probability with a singularity intensity between α i $\alpha+d\alpha$, then $f(\alpha)$ can be adopted as a fractal dimension of the structure characterized by a value α, which is described as the following equation (2)

$$N(\alpha) \approx l^{f(\alpha)}. \qquad (2)$$

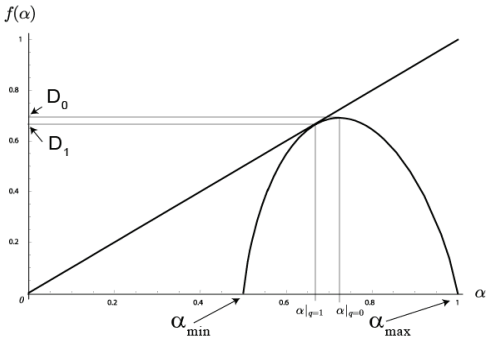


Fig. 1. Example of Multifractal spectrum.

This formalism leads to the definition of multifractal measures in the form of the Hausdorff dimension distribution of exponent α [14]. Typically, the spectrum of $f(\alpha)$ has a form of parables, as in Fig. 1.

Sudden transitions of signal values in time or spatial series are singularities of the signal. By calculating the intensity of singularity (exponents α) and the MF spectrum, we get the MF equivalent of the signal in the spatial or temporal domain, from which by applying MF inverse analysis [16] - [17] it is possible to find those parts in the starting structure, which are characterized by certain values of α and / or $f(\alpha)$.

III. Video Sequence Pre-Processing

For the purpose of the experiment, the introductory sequence of the film "Good Year" (filmed in 2006, director Ridley Scott) is used. Video sequence, which lasted 4 minutes, is converted from DIVX format to uncompressed AVI format, which is used for frame extraction. The total number of extracted frames was M = 4512 size 588x246 pixels. Some examples of extracted frames are shown in Fig. 2.

Fig. 2. Examples of extracted frames from video sequence (a) frame 440 (b) frame 881, (c) frame 952, (d) frame 1004.

From each frame low-level features (color and texture) are extracted and concatenated in the form of feature vector. A feature vector consists of the following features [18]: HSV Color histogram, Color moments, Color layout descriptor, Structural color descriptor, Color correlogram, Gabor transformation features, Radial coocurrence matrix features, Edge histogram and Wavelet texture feature. The

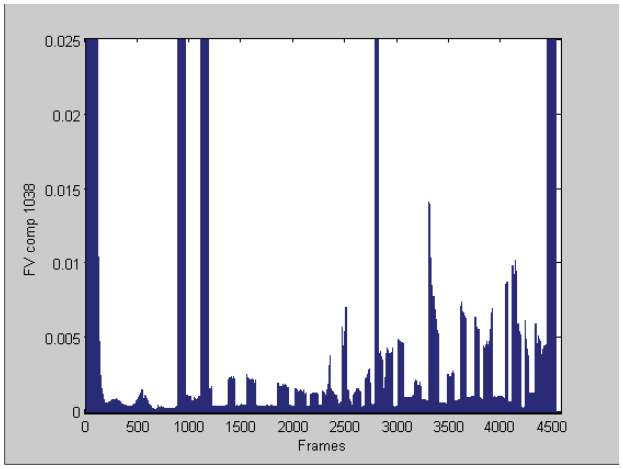
total number of FV coordinates is $N = 1369$. The selected sequence is characterized by feature matrix MxN = 4512x1369. Features matrix columns were normalized with a maximum value within a column.

The position of shot boundaries is defined as the position of the first frame of the next shot. Video sequence is visually inspected and frames, which represent the transitions of shot, are manually marked. The total number of marked frames was 64.

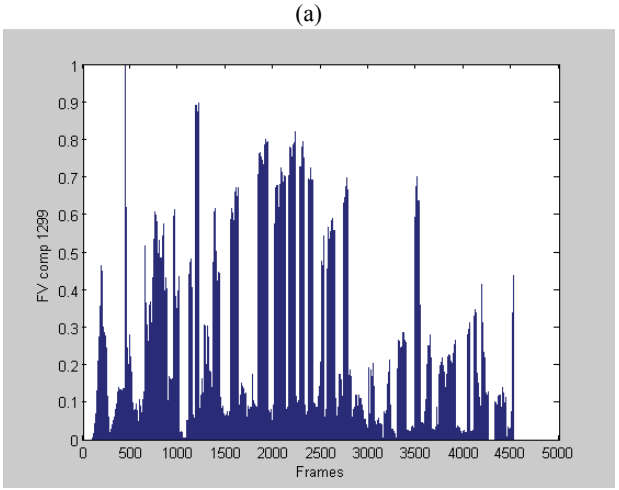
IV. Methodology of Shot Boundary Detection

Frames from the video sequence are described by feature vectors which consist of low-level feature for texture and color. The features are previously extracted from each frame. In a feature matrix FVs are subsequently stored and each column of matrix is a time series for a particular component of FV. Therefore, a feature matrix can be observed as a set of time series and MF analysis is applied to each time series. Any shot boundary is characterized by rapid transitions of texture or color between two frames, which are followed by rapid transitions of FV components values on observed frames position in a feature matrix. MF analysis of time series describes the singularity of series at the local level, and every transition in series is represented as the value of singularity in MF spectrum. High values of singularity correspond to sudden transitions of values in time series, and shot boundaries are represented in MF spectrum as high-value peaks in a local area of spectrum. Finding the positions of high-value peaks in MF singularity spectrum, which was applied in this experiment, sudden transitions in time series can be detected.

A large number of FV coordinates was used in this experiment, but MF analysis was applied to a reduced number of FV components which were selected using a pre-processing reduction criterion. Pre-processing of feature matrix was carried out, and all transitions on previously marked positions of real shots transitions in feature matrix were checked. A reduction criterion for the selection of a set of FV components consists of two conditions. The first condition requires a relative value of transitions of feature component greater than 55% on marked positions compared to the previous position. Difference in values between two coordinates in a time series on marked position is always positive, and calculation starts from a greater value between a marked and previous position. The second condition requires that the first condition is satisfied on at least 40 positions out of 64 in time series that were marked. Based on such defined criterion, the values of MF singularities on the required positions were very prominent in the selected columns in feature matrix. Among the initial set of 1369 time series, only 41 series satisfied the reduction criterion. This set consists of one coordinate of the HSV histogram, three coordinates of the moments of color, two color coordinates from correlogram feature, two coordinates of the features of radial Coocurrence matrix, three coordinates of the edge histogram and 29 coordinates from the Wavelet features. Fig. 3 shows the time series in column 1038 (Coocurrence feature) and 1299 (Wavelet feature) in the feature matrix, which are selected after reduction process.



(a)



(b)

Fig. 3. Time series of FV components (a) 1038 i (b) 1299.

As we said earlier, MF analysis is applied to a selected set of time series, and α exponent is calculated for each series which describes the local irregularity time series. $f(\alpha)$ is not calculated because we were just interested in local transitions. The simplest way of performing the detection of shot boundaries is the detection of high singularity in MF spectrum using thresholding technique [19]. For the given value of threshold, all values below the threshold in MF spectrum were replaced by zero. Due to different variations of values in the selected set of time series, a different threshold value for each time series is used in detection process. The remaining high-value peaks of exponent α, after thresholding is applied, represent the position of shot transitions with a possible error of one frame forward or backward. This error appears due to the nature of value transitions of the FV components and calculation algorithm for the exponent α.

V. Results of Experiment

Exponent α is calculated using the normalized values of selected FV components. Values of the exponent α for columns 1038 and 1299 of feature matrix, which belong to

Coocurrence feature set and wavelet texture feature set, are respectively shown in Figs. 4 and 5. In Figs. 4(a) and 5(a) high-value peaks of the exponent α for each frame can be seen, while the values of the exponent α that remain after applying threshold techniques can be seen in Figs. 4(b) and 5(b). It is obvious that the application of the threshold significantly reduces the number of peaks and only the most prominent peaks remain. However, some of the useful information is removed after thresholding. Some shot transitions, which were detected but with non-prominent peaks in MF spectrum, can be eliminated by thresholding. This disadvantage can be overcome by combining multiple FV components in detection process. Failure detection on any position of real shot boundary in one of the components can be substituted with correct shot boundary detection in other FV component. The results which are presented in this paper are the combination of shot boundary detection results for two FV components from a feature matrix (component on position 1038 in FV and component on position 1299). These two components have the highest number of correct shot boundary detections on marked positions.

TABLE 1. EFFICIENCY OF SHOT BONDARYS DETECTION FOR FV COMPONENT 1038.

FV comp.	*Number of correct detections**	*Number of incorrect detections**	*Threshold*
1038	50/64	4	0.3
1038	49/64	0	0.5
1038	43/60	0	0.7

TABLE 2. EFFICIENCY OF SHOT BONDARYS DETECTION FOR FV COMPONENT 1299.

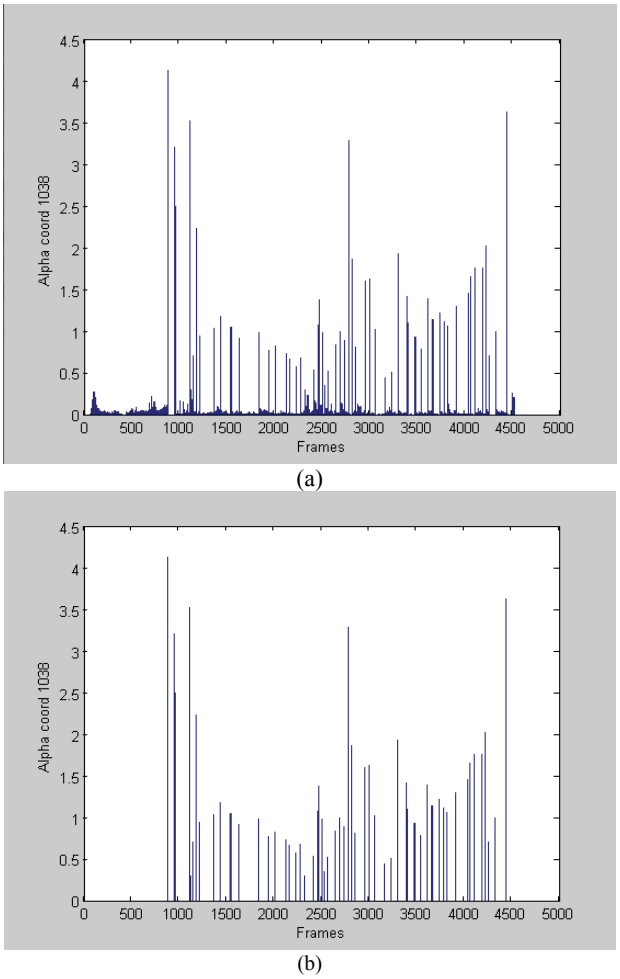
FV comp.	*Number of correct detections*	*Number of incorrect detections*	*Treshold*
1299	59/64	9	0.6
1299	46/64	3	0.8
1299	34/60	0	1.0

* First value represents results of detection for one FV component, and second value represents results of detection for comination of two FV components.

The fusion of the detection results for both FV components is problematic, because of the existing incorrectly detected shot boundaries. It is not easy to remove incorrectly detected transitions from the fused detection results. The examples of correct shot boundaries detection are shown in Fig. 6 and the examples of incorrect detection of shot boundaries are shown in Fig. 7. Figs. 7(a) and 7(b) represent two subsequent frames in a video sequence on a non-marked position, but the detection algorithm describes this transition as sudden. The fact is that the hand of an individual on the left side of the frame was moved very quickly between two frames and that transition was qualified in detection process as sudden. The next example of incorrect shot boundary detection is described in Figs. 7(c)-(e) where rotation of camera is the reason why fast changes appear in the frame. In this example the fast moved object is a tree on the right side of the frame.

A truncation threshold value is set manually. As mentioned above, threshold levels vary from column to column and were determined visually for the experiment. The values of exponent α for the same marked position in different FV components are significantly different. On some positions of shot boundary, some FV components have a very low value of the exponent α and these transitions cannot be detected with a high value of threshold.

The detection results of shot boundary for 1038 and 1299 components of FV are given in Tables 1 and 2. The efficiency of shot boundary detection is presented in Tables 1 and 2 for different values of the thresholds. The first column of tables represents the number of correctly detected shot boundaries. The first value shows the result of detection for only one FV component and the second value shows the result of detection for a combination of two FV components.



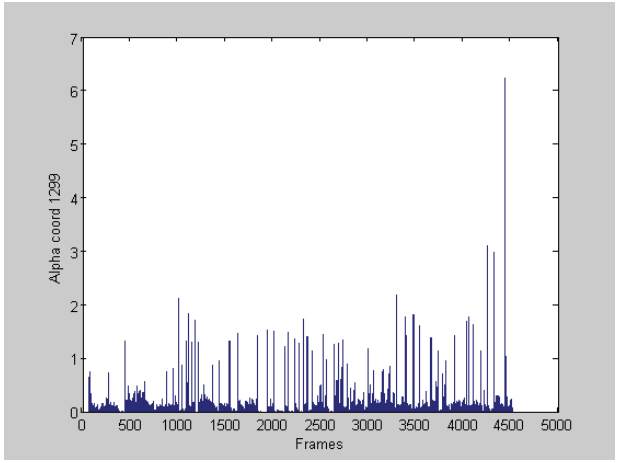
(a)

(b)

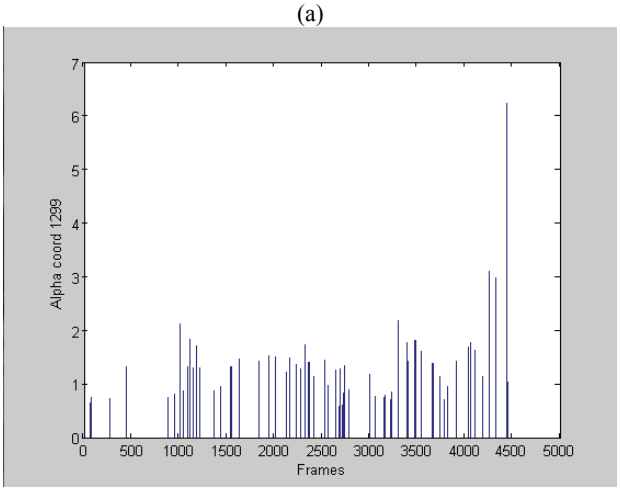
Fig. 4. Values of exponent α for FV component 1038, (a) before tresholding, (b) after tresholding.

The presented results indicate that the reduction of threshold value increases the number of correctly detected shot boundaries in both FV components, but also increases the number of incorrectly detected ones. In contrast, increase in the threshold value reduces the number of incorrectly detected shot boundaries, but also reduces the number of correctly detected ones. If the threshold value is

too high, there is a danger that some of the real shot boundaries cannot be detected in general. In Tables 1 and 2 when incorrectly detected shot boundaries for the threshold 0.7 (1038 FV component) and 1.0 (1299 FV component) do not exist, it can be seen that the number of correctly detected shot boundaries is much lower in each FV component. The total number of correctly detected shot boundaries for a combination of FV components is lower for 4. Four real shot boundaries were not detected. For the first two values of the thresholds in Tables 1 and 2, all shot boundaries were detected in one of the two FV components (1038 and 1299).

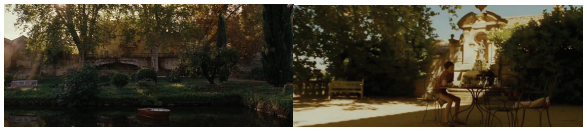


(a)



(b)

Fig. 5. Values of exponent α for FV component 1299, (a) before tresholding, (b) after tresholding.



(a) (b)

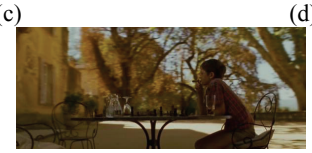
(c) (d)

Fig. 6. Example of correct shot bondary detection (a) frame 439 (b) frame 440, (c) frame 1539, (d) frame 1540.

Incorrectly detected shot boundaries appear due to the use of different effects in video editing process and sudden changes in texture and lighting within the video shot. These changes in video are singularities that are manifested as high-value peaks of the exponent α, in contrast to the low values of the exponent α for other frames within the shot.

(a) (b)

(c) (d)



(e)

Fig. 7. Example of incorrect shot bondary detection (a) frame 3206,(b) frame 3207, moving object (c) frame 646, (d) frame 647, (e) frame 648, rotation of camera.

VI. Conclusion

In this paper the application of multifractal analysis in video boundary shot detection is described. MA is applied to FV components which are extracted from video frames. Shot boundaries are represented with prominent high-value peaks of singularity in alpha domain of MA, and the positions of these peaks are detected using threshold method.

The results presented in this paper show that multifractal analysis can be used very successfully in this area. Real shot boundaries are detected in a video sequence but with the presence of incorrectly detected ones. Incorrectly detected shot boundaries originate from fast object movements between two subsequent frames in a video. This problem can be solved by further improvement of the detection algorithm and implementation of some kind of filter, which will remove incorrectly detected shot boundaries.

Further challenges in the development of presented technique for shot boundaries detection will be finding an optimal combination of FV components from features matrix for detection, automatic tuning of truncation threshold and identification and removal of incorrectly detected shot boundaries.

REFERENCES

[1] Kikukawa T, Karafuto S. “Development of an automatic summary editing system for the audio-visual resources”, *Trans Electr Inf*, 1992, 2(2): pp. 204—212.

[2] Meng J H, Juan Y J, “Chang S. Scene transitions detection in a MPEG compressed video sequence”, *Inter. Sym. Electronic Imaging*, 1995. 14–25.

[3] Hanjalic A. “Shot-boundary detection: Unraveled and resolved?”, *IEEE Trans Circuits Syst Video Tech*, 2002, 12(2): 90–105.

[4] Cernekova Z, Kotropoulos C, Pitas I. “Video shot segmentation using singular value decomposition”, *Proc. ICME*, 2003. 301–302.

[5] Nam J, Tewfik A H, “Detection of gradual transitions in video sequences using B-spline interpolation”, *IEEE Trans Multimedia*, 2005, 7(4): 667–679.

[6] Koprinska I, Carrato S, “Detecting and classifying video shot boundaries in MPEG compressed sequences”, *Europe Conf. Signal Processing*, 1998. 1729–1732.

[7] Zhuang Y T, Rui Y, Huang T S, et al. “Adaptive key frame extraction using unsupervised clustering”, *ICIP, 1998*. 866– 870

[8] Boreczky J S, Wilcox L D, “ A hidden Markov model framework for video segmentation using audio and image features”, *ICASSP*, 1998. 3741–374.

[9] Lienhart R, “Reliable dissolve detection” , *SRMD*, 2001. 219–230.

[10] Yuan J H, Zhang B, Lin F Z, “Graph partition model for robust temporal data segmentation”, *Pacific-Asia Conf. Knowledge Discovery and Data Mining,* 2005. 758–763.

[11] Cernekova Z, Pitas I, Nikou C.,” Information theory-based shot cut fade detection and video summarization,” *IEEE Trans Circuits Syst Video Tech*, 2006, 16(1), pp 82–91

[12] Boccignone G, Chianese A, Moscato V, et al. “Foveated shot detection for video segmentation,” *IEEE Trans Circuits Syst Video Tech*, 2005, 15(3), pp. 365–377

[13] Mandelbrot B B, “The fractal geometry of nature”, *New York, W. H.Freeman and company*, 1977.

[14] Chhabra B A, Meneveau C, Jensen V R, Sreenivasan R K, “Direct determination of the f(a) singularity spectrum and its application to fully developed turbulence", *Physical Review A,* Vol. 40, No. 9, Nov. 1, 1989.

[15] Evertsz G J C, Peitgen O H, Voss F R, “Fractal Geometry and Analysis”*, World Scientific Pub Co Inc*, 1996.

[16] Reljin I, Reljin B, “Fractal geometry and multifractals in analyzing and processing medical data and images”, *Archive of Oncology*, Vol. 10, No. 4, pp. 283-293, 2002

[17] Stojic T, Reljin I, Reljin B, “Adaptation of multifractal analysis to segmentation of microcalcifications in digital mammograms”, *Physica A: Statistical Mechanics and its Applications*, 2006, Vol. 367, Pages 494-508,July, Elsevier Publisher.

[18] Zajić G, Čabarkapa S, Kojić N, Radosavljević V, Reljin B, “Poređenje dve metode klasifikacija slika u CBIR sistemu sa modifikovanim vektorom obeležja” *ETRAN*, Vrnjačka Banja, Jun 15-19, 2009.

[19] Miene, A., Hermes, T., Ioannidis, G. T., and Herzog O., “Automatic Shot Boundary detection using adaptive thresholds,” *in Proceedings of the TRECVID 2003 Workshop*, pp. 275-278, Gaitherburg, Maryland, USA, 2003..