International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:4, No:8, 2010

Gradual Shot Boundary Detection and Classification Based on Fractal Analysis

Zeinab Zeinalpour-Tabrizi, Faeze Asdaghi, Mahmooh Fathy, Mohammad Reza Jahed-Motlagh

***Abstract*—** Shot boundary detection is a fundamental step for the organization of large video data. In this paper, we propose a new method for video gradual shots detection and classification, using advantages of fractal analysis and AIS-based classifier. Proposed features are "vertical intercept" and "fractal dimension" of each frame of videos which are computed using Fourier transform coefficients. We also used a classifier based on Clonal Selection Algorithm. We have carried out our solution and assessed it according to the TRECVID2006 benchmark dataset.

***Keywords*—** shot boundary detection, gradual shots, fractal analysis, artificial immune system, choose Clooney.

I. Introduction

Today, indexing and retrieval of digital videos is an active research area, and shot boundary detection is a fundamental step for the organization of large video data. Therefore, the issue of analyzing and automatically indexing the video content by retrieving highly representative information (e.g., shot boundaries) has been raised in the research community.

We need analyzing our video to achieve high accuracy in video processing. Most of time, shot boundary detection have been noticed between different type of video structure (frame, shot, scene and scenario), in content-based video processing [1, 2]. A video shot is defined as a sequence of frames captured by one camera in a single continuous action in time and space [3]. According to whether the transition between shots is abrupt or gradual, the shot boundaries can be categorized into two types: cut (CUT) and gradual transition (GT). The GT can be further classified into dissolve, wipe, fade out/in (FOI), etc., according to the characteristics of the different editing effects [4].

Z. Zeinalpour-Tabrizi is with the *Computer Engineering Department, Iran University of Science and Technology, Narmak, 16846-13114 Tehran, Iran* (phone: 0098-511-8455658; e-mail: zeinab.zeinalpour@comp.iust.ac.ir).

F. Asdaghi, Jr., is with the *Computer Engineering Department, Iran University of Science and Technology, Narmak, 16846-13114 Tehran, Iran* (e-mail: asdaghi@comp.iust.ac.ir).

M. Fathy is with *Computer Engineering Department, Iran University of Science and Technology, Narmak, 16846-13114 Tehran, Iran* (e-mail: mahfathy@iust.ac.ir).

M. R. Jahed-motlagh is with *Computer Engineering Department, Iran University of Science and Technology, Narmak, 16846-13114 Tehran, Iran* (e-mail: jahedmr@iust.ac.ir).

A large number of shot boundary detection methods have been proposed. Pair-wise comparison of the pixels (which is also called template matching), evaluates the differences in color or the light intensity between two similar pixels in two sequential frames. Although some irrelevant frame differences to the outside have been filtered, these approaches are still sensitive to the movements of the object and the camera. In [5, 6] this method has been used to detect shot boundaries.

A block based method has been proposed by [7], in which each frame is divided into 12 blocks which have no overlap and for each of them the best match in the neighboring according blocks in the previous image based on the light intensity was found. Also in [8], the segmenting of each frame to 4*4 areas and the comparison of colored histograms in according areas has been proposed. In [9], the idea of video sampling in special has been expanded into both the temporal and special area.

In [10], two features of the histograms difference and the pair-wise comparisons of pixels in the clustering method have been combined and the result was that when these filtered features are complementary, they end in the recognition of existing shots and higher accuracy. The first task in analyzing a video directly in the discrete area has been directed by [11], in which a method to recognize the abrupt shot based on discrete cosine transform's coefficients of fames, is proposed.

In [12], the method proposed was called DC- images and was created and compared. DC-images, are the declined images of the original images regarding the location: the (i,j) pixel of the DC image is equal to the average block (i,j) of the original image. In [13], shot boundaries are detected by the comparison of colored histograms of DC-images of sequential frames. These kinds of images are formed by DC terms of discrete cosine transform coefficients for a frame.

In [14], authors used the information theory to recognize the abrupt and dissolve shot boundaries. In [15], authors used genetics algorithm for video segmentation; the reverse value of frames' similarity which is calculated through colored histograms, is used to calculate the Fitness function value. In case of shot boundary detection, there are some surveys [16], [17] and [18].

In this paper, a novel method has been proposed to detect gradual shot boundaries using fractal features that its efficiency compared with the previous ones in nearly similar and remarkable. The proposed method of this article will be completely elaborated in the next parts. Section III illustrates

empirical evaluations of our solutions and implementations on video SBD tasks using the TRECVID test bed. At the end, the summary and the suggestion for the continuation of the task will be presented.

II. Proposed Method

The main goal of this study is gradual shot boundary detection via classification of video according to their fractal features. To achieve the goal, Fourier analysis is used to compute fractal features. The proposed phases are illustrated on fig. 1. In the following section, each phase will be described in detail.

A. Feature Extraction

There are various algorithms to extract fractal dimension from images such as Fourier, Kolmogorov, Korcak, Minkowski and Mean Square Error. These methods are different in computation, accuracy and border estimation. In [19, 21, 22] authors showed that there is high correlation between fractal values. Among the mentioned methods, Fourier analysis is more appropriate for our purpose because of following reasons:

- This method is almost not sensitive to noises. [20]
- A fast algorithm exists for Fourier Coefficient computation.
- In addition to linear regression angel, vertical intercept could be computed which is an effective feature.

Assume *V* is a grey video with p frames:

$$\mathbf{V} = \{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(p)}\} \quad (1)$$

Which $v^{(k)}$ is the K^{th} frame with M*N dimensions ($1 \le k \le p$).

$$\mathbf{v}^{(k)} = \{v_{i,j}^{(k)}\}_{M\times N}, \quad 1 \le i \le M, \quad 1 \le j \le N, \quad 1 \le k \le p \quad (2)$$

Here $v_{i,j}^{(k)}$ is the grey level value of pixel which is located in i^{th} row and j^{th} column of k^{th} frame of *V*. At this point, each $v^{(k)}$ frame is converted to frequency domain using 2-dimensional discrete Fourier transform. [23]

$$F^{(k)}(u,v)=\frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[v_{i,j}^{(k)} \times e^{\left[-j2\pi\left(\frac{iu}{M}+\frac{jv}{N}\right)\right]}\right] \quad (3)$$

For Fourier coefficients of each $F^{(k)}$ frames, the power spectral density (PSD) are computed:

$$S^{(k)}(u,v) = \left\|F^{(k)}(u,v)\right\|^2 \quad (4)$$

Then coordinate system is transformed to Polar coordinates to compute power spectrum density which is needed for fractal dimension computation:

$$f = \sqrt{u^2+v^2} \quad (5)$$

$$\theta = \tan^{-1}\left(\frac{v}{u}\right) \quad (6)$$

After variables replacement in to new system, $S^{(k)}(u,v)$ is converted to $S^{(k)}(f,\theta)$. Average values of $S^{(k)}(f,\theta)$ is computed in every radial range of *f* and for each θ:

$$S^{(k)}(f) = \sum_{\theta} S^{(k)}(f,\theta) \quad (7)$$

In [21] authors proved that the power spectrum shows a linear variation between logarithm of *S(f)* and logarithm of the frequency of surface:

$$S^{(k)}(f) = c\,.\,|f|^{-\beta} \quad (8)$$

In other words:

$$log\left(S^{(k)}(f)\right) = log(c) - \beta \times log|f| \quad (9)$$

Slope of linear regression line of these changes are related with fractal dimension of image as follow:

$$FD = \frac{(7-\beta)}{2} \quad (10)$$

Which FD shows fractal dimension of image.

Other important aspect is the tight relation of vertical intercept with log(c) which was shown in (9). In [20] this value is named vertical intercept and is used as a complementary fractal feature. In this paper both fractal dimension value and vertical intercept are utilized. In this paper we use both of these features, vertical intercept and fractal dimension of each frame, for gradual shot boundary detection.

B. Gradual shot detection and classification

For gradual shots detection, feature transitions over a time duration in the range of a few consecutive frames are used as a feature. Thus, we statistically review our training set; The minimum length of gradual shots (common gradual shots with more than 5 frames) was determined *minGSL*. Then for modeling gradual shots, sliding windows with double length of *minGSL* is constructed from vertical intercept feature and fractal dimension of consecutive frames. (Fig. 2)

i. AIS-based classifier

After construction of each frame's feature vector according to their fractal features, classification is done using artificial immune system. Artificial immune is a branch of soft computing algorithms which is modeled from vertebrate immune system and seems similar to genetic algorithm. Artificial immune system consists of diverse methods such as negative selection algorithm, colony selection, immune networks an etc. In this paper, we utilized colony selection algorithm. The proposed method consists of two sections: learning and classification.

In colony selection algorithm, for each class, strings with similar value of the class features' vector, is randomly created. Then these strings are muted and according to the training data which are belong to the other class, a detection radial is defined. These strings are named "Colon". If the colon is able to detect the data belonging to its class, it is kept else they are

omitted. Among the remaining colons, the colon which detects more data is remained and the others are removed. The data which the colon is detected is then discarded and the process is continued with the remaining data. This process continues repeatedly till a set of colons is created for each class which covers all the class's data.

After learning phase is completed, system is ready for determine input frame's class. When a frame is fed to the system for labeling, its feature vector is compared to the created sets of colonies' feature vector. If input frame is located in the detection radial of even one of the colons of that special class, it is considered to belong to that class. If the frame is detected with colons of more than a class, it would belong to the class with less distance to that its colons. In this way the class of input data is determined.

Since the classification method is performed on the window of features which belong to consecutive frames, for detection of shot border position, each window's middle frame is considered as exponent of its window and window label is applied to that frame. At last, windows exponents' frames are classified to four classes according to the their windows' label: «*dissolve*» ،«*fade in/out*» ،«*other*» *and* «*none-shot*».

III. Experimental results

In this section, experimental results on the TRECVID 2006 dataset are presented. The method was tested on the reference video test set TRECVID 2006 [24] containing newscasts having many commercials in-between. Both news and commercials are characterized by significant camera effects like zoom-ins/outs, pans, and significant object and camera motion inside one shot.

Video sequences have been digitized with a frame rate of 29.97fps at a resolution of 352 × 240. We used 60% of TRECVID 2006 videos for training and 40% for our test set.

According to fig. 3, feature changes along shot occurrence have the potential to be modeled. Figure 4 illustrates our modeling of feature changes along gradual occurrence due to training its variation to classifier along its training phase.

The performance of proposed method is measured by "Precision" and "recall" which are the most usual measures that used for shot detection. Precision of shot detection means that, numbers of detected borders are true and recall means what number of existing borders in the video is detected. These two factors are computed as follow:

$$Precision = \frac{number\ of\ true\ detected\ shots}{number\ of\ detected\ shots} \quad (11)$$

$$Recall = \frac{number\ of\ true\ detected\ shots}{number\ of\ shots} \quad (12)$$

In Table I, the recall and precision rates of first nine methods of TRECVID's report are illustrated and our method for gradual transitions, as well as the FrameRecall and FramePrecision for the gradual transitions, is located at the last row of table. The best two performance values obtained on each category are marked in bold.

Due to the information of Fig. 5, it is clear that the achieved results are almost better than those reported in the TRECVID 2006 competition [25]. The overall results of detection from TRECVID 2006 participants are illustrated in fig. 5.

IV. Conclusion

In this paper a novel method is proposed for video shots segmentation using advantages of fractal analysis for gradual shot boundary detection. This method consist of three sections: fractal feature extraction for each frame, produce a sliding window of sequential frames over double minimum length of gradual shots and finally, gradual shot detection via classification of these ranges using of colony selection algorithm.

The experimental results are evaluated on the TRECVID 2006 videos for shot border detection. Our proposed method improved recall of gradual shot boundary. The proposed method novelty is using fractal features and AIS based classifier for shot boundary detection. In future we continue our research on detection of abrupt shot using fractal based features.

References

[1] S. W. Smoliar and H.-J. Zhang, "Content-based video indexing and retrieval," *IEEE Multimedia*, vol. 1, no. 2, 1994, pp. 62–72.
[2] R. Lienhart, S. Pfeiffer, and W. Effelsberg, "Video abstracting," *Commun. ACM*, vol. 40, no. 12, 1997, pp. 55–62.
[3] X. U. Cabedo and S. K. Bhattacharjee, "Shot detection tools in digital video," in *Proc. Non-Linear Model based Image Analysis*, 1998, pp.121–126.
[4] V. Kobla, D. DeMenthon, and D. Doermann, "Special effect edit detection using videotrails: a comparison with existing techniques," in *Proc. SPIE Conf. Storage Retrieval Image Video Databases VII*, Jan. 1999, pp. 302–313.
[5] A. Nagasaka and Y. Tanaka, "Automatic Video Indexing and Full-Video Search for Object Appearances," Proceedings of the IFIP TC2/WG 2.6 Second Working Conference on Visual Database Systems II, North-Holland Publishing Co., 1992, pp. 113-127.
[6] H. Zhang, A. Kankanhalli, and S.W. Smoliar, "Automatic partitioning of full-motion video," Multimedia Systems, vol. 1, Jan. 1993, pp. 10-28.
[7] B. Shahraray, "Scene change detection and content-based sampling of video sequences," Apr. 1995.
[8] A. Nagasaka and Y. Tanaka, "Automatic Video Indexing and Full-Video Search for Object Appearances," Proceedings of the IFIP TC2/WG 2.6 Second Working Conference on Visual Database Systems II, North-Holland Publishing Co., 1992, pp. 113-127.
[9] W. Xiong, J. C.-M. Lee, "Efficient Scene Change Detection and Camera Motion Annotation for Video Classification", *Computer Vision and Image Understanding,* Vol.71, Issue 2,1998, pp.166-181.
[10] A.M. Ferman and A.M. Tekalp, "Efficient Filtering and Clustering Methods for Temporal Video Segmentation and Visual Summarization," Journal of Visual Communication and Image Representation, vol. 9, Dec. 1998, pp. 336-351.
[11] F. Arman, A. Hsu, and M. Chiu, "Image processing on compressed data for large video databases," Proceedings of the first ACM international conference on Multimedia, Anaheim, California, United States: ACM, 1993, pp. 267-272.
[12] B. Yeo and B. Liu, "Rapid scene analysis on compressed video," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 5, 1995, pp. 544, 533.
[13] K. Shen and E.J. Delp, "A fast algorithm for video parsing using MPEG compressed sequences," Proceedings of the 1995 International Conference on Image Processing (Vol.2)-Volume 2 - Volume 2, IEEE Computer Society, 1995, p. 2252.
[14] Z. Cernekova, I. Pitas, and C. Nikou, "Information theory-based shot cut/fade detection and video summarization," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 16, 2006, pp. 82-91.

[15] P. Chiu, A. Girgensohn, W. Polak, E. Rieffel, and L. Wilcox, "A genetic algorithm for video segmentation and summarization," Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on, 2000, pp. 1329-1332 vol.3.
[16] . Lienhart, "Reliable transition detection in videos: a survey and practitioner's guide," *Int. J. Image Graph.*, vol. 1, no. 3, 2001, pp. 469–486.
[17] M. Albanese, A. Chianese, V. Moscato, and L. Sansone, "A Formal Model for Video Shot Segmentation and its Application via Animate Vision," Multimedia Tools and Applications, vol. 24, Dec. 2004, pp. 253-272.
[18] Jinhui Yuan, Huiyi Wang, Lan Xiao, Wujie Zheng, Jianmin Li, Fuzong Lin, and Bo Zhang, "A Formal Study of Shot Boundary Detection," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 17, 2007, pp. 168-186.
[19] Sarkar N., B.B. Chaudhuri, "An efficient differential box-counting approach to compute fractal dimension of image", IEEE Transaction on System Man and Cybernet, vol.24, no.1, 1994, pp.115-120.
[20] Zhang J. , Regtien P.P.L. , Korsten M.J. , "Monitoring of dry sliding wear using fractal analysis", 10th TC-10 IMEKO Conference on Technical Diagnostics, 9-10 June, 2005, Budapest, Hungary
[21] Russ J. C.: Fractal Surfaces. New York [etc.], Plenum Press 1994.
[22] Rawers J., Tylczak J.: Fractal characterization of wear-erosion surfaces, Journal of Materials Engineering and Performance, vol. 8, no.6, 1999, pp. 669-676.
[23] R.C. Gonzalez and R.E. Woods, Digital Image Processing, Prentice Hall, 2002.
[24] Z. Cernekov´a, N. Nikolaidis, I. Pitas. AIIA shot boundary detection at TRECVID 2006, TREC Video Retrieval Evaluation[C], 2006
[25] NIST, Homepage of Trecvid Evaluation, http://www-nlpir.nist.gov/projects/trecvid/

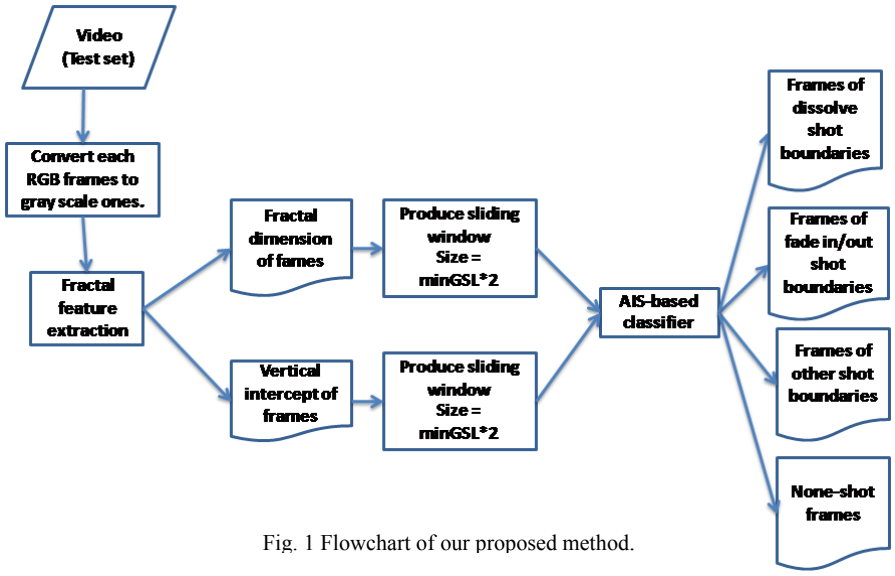

Fig. 1 Flowchart of our proposed method.

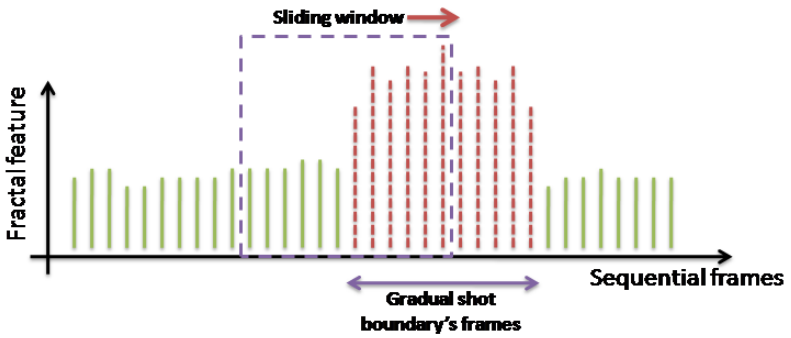

Fig. 2 An example of sliding window for *minGSL*=6.

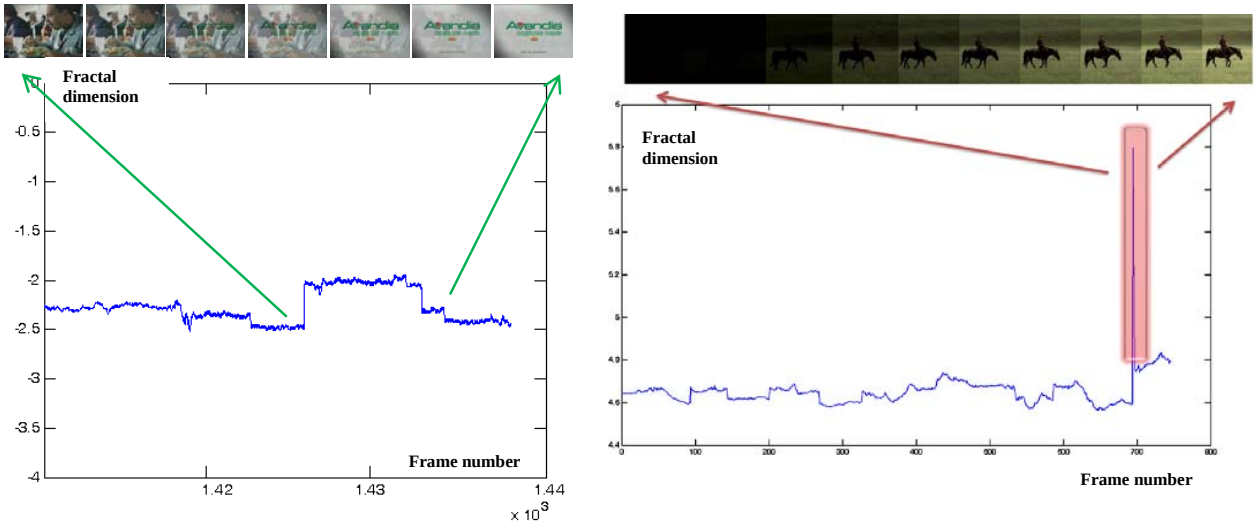

Fig. 3 Fractal dimension variation on a sample video frames.

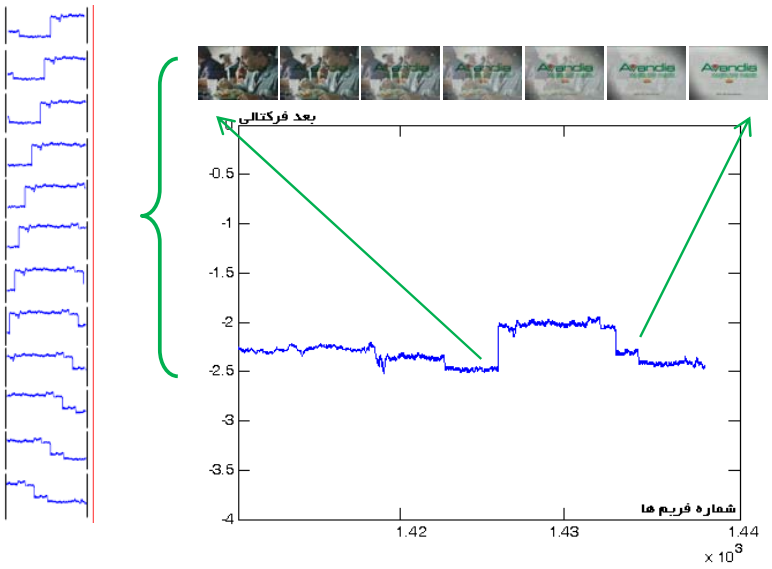

Fig. 4 Using sliding window for modeling gradual shot boundaries' feature variation.

TABLE I
THE RESULTS OF EXPERIMENTS ON TRECVID 2006 DATA SET.

Method	FRAME		GRADUAL	
	Recall	Precision		Precision
a	0/7243	0/7850	0/6446	0/6541
b	**0/8739**	**0/9261**	0/7416	**0/8355**
c	0/8275	0/7984	0/6030	0/8101
d	0/5639	0/7834	0/4031	0/5276
e	0/8527	0/5637	0/6420	0/6507
f	0/2540	0/7056	0/0159	0/7416
g	0/4269	0/7766	0/2126	0/3778
h	0/7716	0/8486	0/6013	0/8024
i	0/7726	0/7000	0/7565	0/5711
Proposed method	0/8271	0/8713	**0/7894**	0/8227

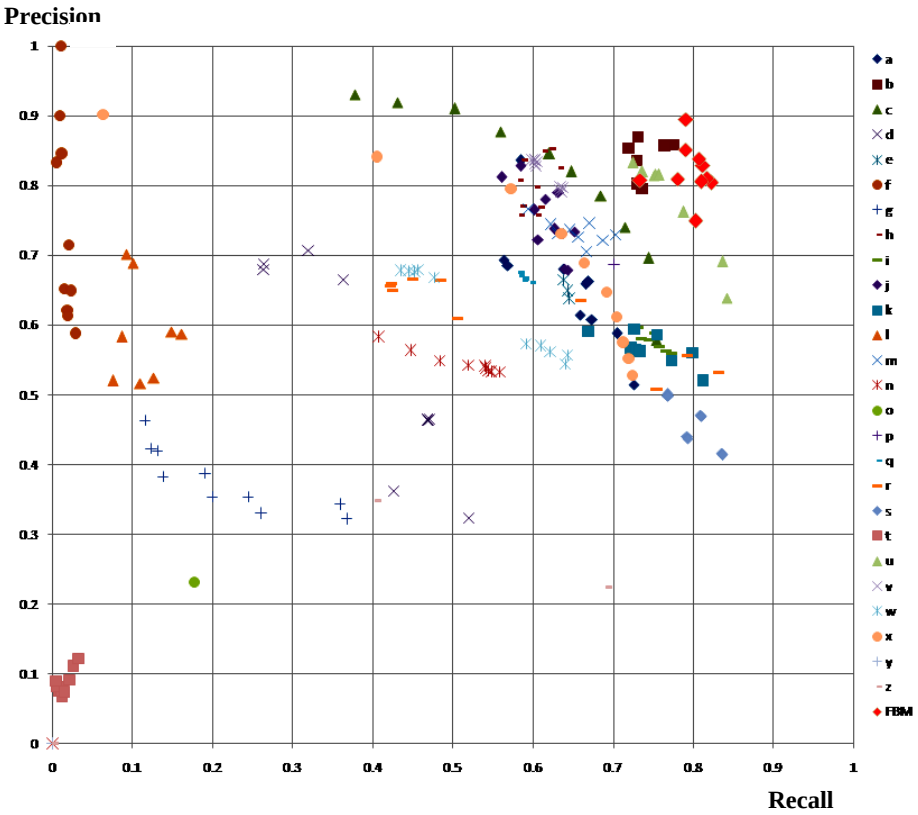

Fig. 5 Results on the gradual shot detection based on the data provided by the organizers in TRECVID-2006. Our approach is labeled FBM.