Gradual Shot Change Detection in Soccer Videos via Fractals

Efnan Sora Gunal [1], Selcuk Canbek [2], and Nihat Adar [2]

[1] Dept. of Electrical and Electronics Engineering, Eskisehir Osmangazi University, Eskisehir, Turkiye
esora@ogu.edu.tr
[2] Dept. of Computer Engineering, Eskisehir Osmangazi University, Eskisehir, Turkiye
selcuk@ogu.edu.tr, nadar@ogu.edu.tr

Abstract

Accurate shot change detection plays vital role to organize video contents into meaningful parts for video scene analysis. Shot changes may occur with either abrupt or gradual shot transition. In abrupt shot transition, the change of video content occurs over a single frame. In gradual shot transition, however, the content change takes place gradually through a short sequence of frames. The gradual transition is also divided into several subgroups such as dissolve, pan, zoom and wipe. In the literature, there exist numerous methods to detect shot changes. While these methods are able to detect abrupt changes properly, they usually fail to detect gradual changes. This paper proposes a new method for detecting gradual shot changes successfully. The proposed method employs fractal dimension information of gray scale video frames. In the experimental work, efficacy of fractal based gradual shot change detection method is tested on soccer videos including different types of gradual shot changes and compared with well known pixel and histogram based shot change detection methods. Results of the experimental study indicate that the proposed method yields higher accuracy with respect to other methods in most cases.

1. Introduction

In recent days, digital video combining audio and visual information are available everywhere. Rapid advances on video equipments such as digital cameras, camcorders and storage media increase usage fields of the video including internet conferencing, multimedia authoring systems and e-education. This situation increases the demand for not only huge video databases but also organizing and retrieving the desired information through these databases as well. Since audio-visual data includes semantic information, retrieving procedure is quite complicated [1-5].

The demand for organizing video databases causes scene change detection to become one of the popular research areas on video processing. While the scene change detection process is based on semantic content of the video, shot boundary detection is the fundamental preprocessing step in semantic video processing [6]. Scenes are formed by semantically related individual shots that are uninterrupted segment of a video frame sequence with static or continuous camera motion [7]. Accurate shot change detection therefore plays a crucial role to organize video contents into meaningful parts for scene analysis [8]. Shot changes may occur with either abrupt or gradual shot transition. In abrupt shot transition, the change of video content occurs over a single frame. In gradual shot transition, however, the content change takes place gradually through a short sequence of frames. The gradual transition is also divided into several subgroups such as dissolving, wiping, noise and camera movements (pan, tilt, zoom) [9].

Performance of the shot change detection methods directly depends on the features that are used to represent the video content. These features are extracted by means of reducing large dimensionality of video frames [6]. Existing shot change detection techniques utilize differences on features over consecutive video frames. Pixel difference and histogram difference are the most widely used and fundamental approaches among all [10-12].

Sum of Absolute Differences (SAD) is the fundamental pixel difference method. In this approach, the difference between two frames is obtained by calculating the value that represents overall change in pixel intensities of images [8, 10]. The sum of absolute pixel-wise intensity differences between two frames is used as a frame difference as shown in Eq. (1) where f_1 and f_2 are intensity values of consecutive frames, X and Y are the height and weight, x and y represent pixel coordinates of the frames, respectively. The main disadvantage of pixel difference methods is that they are very sensitive to noise and camera motion [8].

$$SAD(f_1, f_2) = \frac{1}{X \times Y} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} \left| f_1(x,y) - f_2(x,y) \right| \quad (1)$$

Histogram differences is another widely used approach in shot change detection [8-14]. The most popular histogram method, Bin-to-Bin (B2B), is computed as in Eq. (2) where h_1, h_2 are histograms of consecutive frames, and N is number of pixels in a single frame.

$$D_{B2B}(h_1, h_2) = \frac{1}{2N} \sum_i \left| h_1[i] - h_2[i] \right| \quad (2)$$

While the above mentioned methods are able to detect abrupt changes successfully, they may fail to detect gradual changes. This paper proposes a new method for detecting gradual shot changes successfully. The proposed method employs fractal dimension (FD) information of gray scale video frames. FD, which was introduced by Mandelbrot [15], is an important tool to extract roughness and self-similarity features from images. Most of the objects in natural life have complex and irregular structures that can not be described by using ideal mathematical shapes such as cubes, cones and cylinders defined in Euclidean geometry. It is however easier to describe those objects by FD. Using FD in image feature extraction and segmentation is induced by the observation that FD has a strong correlation with human judgment of surface roughness and is relatively insensitive to image deformation as well [16]. FD has been previously applied to many areas in image processing and pattern recognition. Textural segmentation, classification, shape analysis are applicable research areas of FD in satellite, medical and natural images [17].

FD difference (FDD) values for successive video frames constitute the features that are used for gradual shot change detection process. Using FD rather than pixel or histogram information mentioned above would help more to define visual complexity and textural information of video frames. The proposed method is tested on soccer videos including dissolve, zoom, wipe and pan-type gradual shot changes inside. Experimental results reveal that FD based method offers comparable and even better detection performance with respect to other methods.

Rest of the paper is organized as follows: In Section 2, computation of FD is briefly described and the proposed shot change detection method using FD is given. Section 3 provides the experimental work and results. Finally, in Section 4, conclusions of the paper are given.

2. Feature extraction based on FD

All FD computation methods follow the same principles of Hausdorff – Besicovitch(D) Dimension [18]. D dimension of a bounded set A in $\mathbb{R}^n$ is a real number used to characterize the geometrical complexity of A . Here, the set A is called a fractal set if its D dimension is strictly greater than its topological dimension [19]. D dimension of a subset X of Euclidean space can be defined by counting numbers of open balls used to cover X .

The box counting dimension (D_B) of set A is defined as

$$D_B = \lim_{r \to 0} \frac{\log(N_r)}{\log(1/r)} \tag{3}$$

where N_r is the number of the boxes of size r needed to cover A [15].

According to [20], gray level values are assumed to be a 3D surface to calculate the FD of an image. FD can be calculated using well known Differential Box Counting Dimension (DBCD) method that is also called as “Blanket” method as well. To calculate the N_r , the image of size M×M pixels is scaled down to a size s × s (Fig. 1) where M/2≥ s> 1 and s is an integer. r is scaling ratio and equals to (s/ M). The image is considered as 3-D space with (x, y, z) axes. While (x, y) is denoting 2-D position, the z axis denotes gray level. After (x, y) space is partitioned into grids of size s × s, each grid contains column of boxes of size s × s × h. If the total number of gray levels is G then [G / h] = [M / s]. Let the minimum and maximum gray level of the image in the $(i, j)_{th}$ grid fall in box number k and l, respectively.

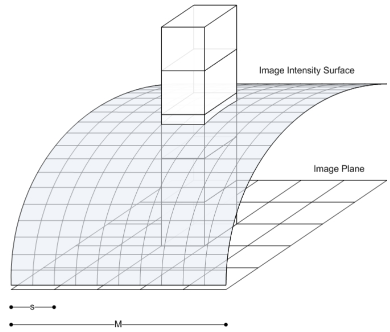

Fig. 1. Image intensity surface

For each $(i, j)_{th}$ grid, $n_r(i, j)$ (Eq. (4)) is the contribution of N_r.

$$n_r(i, j) = l - k + 1 \tag{4}$$

Taking contributions from all grids, we have

$$N_r = \sum_{i,j} n_r(i, j) \tag{5}$$

N_r is counted for different values of r, i.e., different values of s. Then, using Eq. (5), FD is estimated using least square linear fit of log (N_r) against log(1/r).

In our proposed method, each frame of the color video is assumed to be an image and converted to gray scale. As shown in Eq. (6), after computing FD of each frame using DBC method that is explained above, FDD are calculated by subtracting FD values of consecutive frames, f and $f+1$.

$$FDD = \left| FD_{f+1} - FD_f \right| \tag{6}$$

After computation of FDD values, shot changes can be located using a threshold approach. In our study, a dynamic threshold method (DTM) is used where threshold is dynamically changed during shot change detection process [21]. The threshold in DTM is computed by considering the presence of a shot change and the variation of frame contents. DTM consists of two threshold stages: fixed and dynamic. DTM is used as the thresholding tool in all shot change detection methods (FDD, SAD, B2B) for each gradual change type. Examples to various gradual shot change types (dissolve, pan, zoom and wipe) through consecutive video frames are provided in Fig. 2-5.

Fig. 2. Dissolve-type gradual transition

Fig. 3. Pan-type gradual transition

Fig. 4. Zoom-type gradual transition

Fig. 5. Wipe-type gradual transition

3. Experimental Work

Since soccer videos contain lots of object moves and shot changes, they are employed as the test video dataset in the experimental work. Before extracting the FD based features from video frames, each type of gradual shot changes are individually collected through compressed soccer video dataset and then combined into a single video file using video editing software. Hence, different types of gradual shot changes can be evaluated individually. A total of four different types of gradual change videos which consist of 10837 frames are used in testing. Size of each frame is 240 x 320 pixels. Before testing, actual locations of gradual shot locations through the test videos are determined by a comprehensive manual analysis. Types of the gradual shot changes encountered in the test videos are dissolve, zoom, wipe and pan.

Detection performance of FDD method is measured and compared with the histogram and pixel based detection methods using Recall and Precision parameters which are defined in Eq. (7) and Eq. (8). In these equations, N_m, N_C and N_F represent missed, correct and false detections, respectively.

$$RECALL = \frac{N_C}{N_C + N_m} \tag{7}$$

$$PRECISION = \frac{N_C}{N_C + N_F} \tag{8}$$

In case of our study, recall is the ratio of number of correctly detected shot changes obtained by a particular method to total number of actual shot changes. Similarly, precision is the ratio of number of correctly detected shot changes to total number of correctly and incorrectly detected shot changes. Efficacy of a detection method would increase as these rates get closer to one.

Gathering same type of gradual changes into a single video enables us to calculate precision rates for every single type. Otherwise, it is not possible to calculate precision due to uncertainty of the type of false alarm cases.

In addition to recall and precision, another commonly used metric, F1, combining precision and recall information is also used for evaluation. F1 measure is calculated as in Eq. (9). Value of F1 is high only when recall and precision are both high [6].

$$F1 = \frac{2 \times PRECISION \times RECALL}{PRECISION + RECALL} \tag{9}$$

Table1-5 shows the recall, precision and F1-measure rates for detection of gradual shot changes including dissolve, pan, zoom and wipe-type. In case of dissolve transition (Table 1), SAD and FDD detection methods have similar recall performances. On the other hand, FDD yields the best recall and precision values for pan-type shot changes as shown in Table 2. Recall rates of both zoom and wipe-type changes in Table 3 and 4 indicate that FDD method offer promising performance with respect to classical pixel and histogram based shot change detection methods. Moreover, FDD has the best F1 rate for all types of gradual transitions except dissolve. In addition to analysis for individual types of gradual shot change detection performance, Table 5 provides an overall analysis among all four types. Here, FDD provides best performance in terms of recall and F1 measure rates. In all tables, highest rates are indicated as bold for clarification.

Table 1. Results of dissolve-type shot change detection

Methods	Nc	Nf	Nm	Dissolve	RECALL	PRECISION	F1
FDD	2	5	8	10	0,200	0,286	0,235
SAD	2	4	8	10	0.200	0.333	0.250
B2B	8	7	2	10	0.800	0.533	0.640

Table 2. Results of pan-type shot change detection

Methods	Nc	Nf	Nm	Pan	RECALL	PRECISION	F1
FDD	9	5	10	19	**0.474**	**0.643**	**0.545**
SAD	2	2	17	19	0.105	0.500	0.174
B2B	7	5	12	19	0.368	0.583	0.452

Table 3. Results of zoom-type shot change detection

Methods	Nc	Nf	Nm	Zoom	RECALL	PRECISION	F1
FDD	13	13	14	27	**0.481**	**0.500**	**0.491**
SAD	3	6	24	27	0.111	0.333	0.167
B2B	0	3	27	27	0.000	0.000	0

Table 4. Results of wipe-type shot change detection

Methods	Nc	Nf	Nm	Wipe	RECALL	PRECISION	F1
FDD	22	30	8	30	**0.733**	0.423	**0.537**
SAD	4	1	26	30	0.133	**0.800**	0.229
B2B	20	41	10	30	0.667	0.328	0.440

Table 5. Results of overall gradual shot change detection

Methods	Nc	Nf	Nm	Overall	RECALL	PRECISION	F1
FDD	44	49	42	86	**0.512**	**0.473**	**0.492**
SAD	11	13	75	86	0.128	0.458	0.200
B2B	35	56	51	86	0.407	0.385	0.395

4. Conclusions

Accurate shot change detection is of great importance for organizing video contents into meaningful parts for video scene analysis. Most of the shot change detection algorithms are able to detect abrupt changes properly; however, they are not very successful to detect gradual changes where the video content change takes place gradually through a short sequence of frames. Hence, this paper addresses the detection of gradual shot changes with a novel method employing fractal dimension information. The proposed method is tested on soccer videos including different types of gradual shot changes and compared with well known pixel and histogram based shot change detection methods. Results of the experimental study indicate that the proposed method yields higher accuracy with respect to the other methods in most cases.

5. References

[1] W.A.C Fernando, C.N. Canagrajah, D.R. Bull, “A Unified approach to scene change detection in uncompressed and compresses video”, IEEE Transactions on Consumer Electronics, vol. 46, no. 3, pp. 769-779, Aug 2000.

[2] W.A.C Fernando, C.N. Canagrajah, D.R. Bull, “Scene change detection algorithms for content based video indexing and retrieval”, Electronics & Communication Engineering Journal, pp. 117-125, June 2001.

[3] M. Saric, H. Dujmic, D. Baricevic, “Shot Boundary Detection in Soccer Video using Twin-Comparison Algorithm and Dominant Region”, Journal of Information and Organizational Sciences. Vol. 32, no. 1, pp. 67-73, 2008.

[4] B. Yeo, B. Liu “Rapid Scene Analysis on Compresses Video”, IEEE Transactions on circuits and Systems for video Technology, vol. 5, no. 6, Dec.1995.

[5] X. Wang, Z. Weng “Scene Abrupt Change Detection”, in Canadian Conference on Electrical and Computer Engineering, vol.2, pp. 880-883, 2000.

[6] C. Cotsaces, M. Gavrielides and I.Pitas, "A Survey of Recent Work in Video Shot Boundary Detections", in Proc. of *2005 Workshop on Audio-Visual Content and Information Visualization in Digital Libraries (AVIVDiLib '05)*, Cortona, Italy, 4-6 June, 2005.

[7] O. Marques and B. Furht “Content-based Image and Video Retrieval”, Kluwer Academic Publishers, 2002.

[8] J. Korpi-Anttilla “Automatic color enhancement and scene change detection of digital video”, Licentiate thesis, Helsinki University of Technology, 2002.

[9] H. Lu, Y.P. Tan, “An effective post-refinement method for shot boundary detection”, IEEE Trans. on Circuits and Systems for Video Technology, vol. 15, no.11, pp. 1407-1421, 2005.

[10] J.S. Boreczky, L.A. Rowe, “A comparison of video shot boundary detection techniques”, J. Electronic Imaging, vol.5, no. 2, pp. 122-128,1996.

[11] A. Ekin, “Sports video processing for description, summarization and search”, Ph.D Thesis, Department of Electrical and Computer Engineering, University of Rochester, Rochester NY, 2003.

[12] H. J. Zhang, A. Kankanhalli, S.W. Smoliar, “Automatic partitioning of full-motion video”, Multimedia Systems, vol. 1, no.1, pp. 10-28, 1993.

[13] A. Nagasaka, Y. Tanaka, “Automatic video indexing and full-video search for object appearances”, Visual Database Systems, II, Elsevier Science Publishers, pp.113-127, 1992.

[14] Y. Tonomura, “Video handling based on structured information for hypermedia systems”, Proc. of International Conference on Multimedia Information Systems, 1991, pp. 333-344.

[15] B. Mandelbrot. Fractals: Form. Change. an Dimension. San Francisco, Freeman, 1977.

[16] Y. Liu, Y. Li, Image feature extraction and segmentation using fractal dimension, International Conference on Information, Communication and Signal Processing, Vol. 2, 1997, pp. 975-979.

[17] W.Sun, G. Xu, P. Gong, S. Liang, “Fractal analysis of remotely sensed images: A review of methods and applications”, International Journal of Remote Sensing, vol. 27, no. 22, pp. 4963-4990, 2006.

[18] M. F. Barnsley. Fractals Everywhere. Academic Press. Boston, USA, 1988.

[19] B. B Chaudhuri and N. Sarkar, “Texture segmentation using fractal dimension”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 17, no.1, pp.72-77, 1995.

[20] N. Sarkar, B.B.Chaudri, “An efficient differential box-counting Approach to compute fractal dimension of image”, IEEE Trans. on Systems, Man, and Cybernetics, vol. 24, no. 1, pp. 115-120, 1994.

[21] S. Youm, W. Kim, “Dynamic Threshold method for scene change detection”, Proc. of International Conference on Multimedia and Expo., 2003, pp. 337-340.