



Fernando, W. A. C., Canagarajah, C. N., & Bull, D. R. (1999). DFD based scene segmentation for H.263 video sequences. In Proceedings of ISCAS, Image and Video Processing, Multimedia and Communications. (Vol. 4, pp. 520 - 523). Institute of Electrical and Electronics Engineers (IEEE).  
10.1109/ISCAS.1999.780056

Link to published version (if available):  
[10.1109/ISCAS.1999.780056](https://doi.org/10.1109/ISCAS.1999.780056)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/about/ebr-terms.html>

### Take down policy

Explore Bristol Research is a digital archive and the intention is that deposited content should not be removed. However, if you believe that this version of the work breaches copyright law please contact [open-access@bristol.ac.uk](mailto:open-access@bristol.ac.uk) and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline of the nature of the complaint

On receipt of your message the Open Access Team will immediately investigate your claim, make an initial judgement of the validity of the claim and, where appropriate, withdraw the item in question from public view.

# DFD BASED SCENE SEGMENTATION FOR H.263 VIDEO SEQUENCES

W.A.C. Fernando, C.N. Canagarajah, D. R. Bull

Centre for Communications Research,  
Department of Electrical and Electronic Engineering, University of Bristol,  
Merchant Ventures Building, Woodland Road, Bristol BS8 1UB, United Kingdom.

Voice - +44-117-954-5198, Fax - +44-117-954-5206

Email - W.A.C.Fernando@bristol.ac.uk

## ABSTRACT

Content based indexing and retrieval of video is becoming increasingly important in many applications. Identifying scene changes and special effects in a video scene is an essential pre-requisite for indexing. In this paper, a sudden scene change detection algorithm for H.263 video sequences is proposed. This method uses the number of intra-coded macroblocks and Displaced Frame Difference (DFD) signal of the video signal. Experimental results show that the performance of this algorithm is independent of the encoder threshold. Furthermore, this algorithm is capable of detecting abrupt scene changes accurately even the video sequence contains special effects.

## 1. INTRODUCTION

Due to rapid advances in compression technology and imaging hardware, the expansion of low-cost storage media, and the explosion of the internet, the digital video "for every one" is now becoming a reality. The demand for digital video is also increasing in areas such as video conferencing, multimedia authoring systems, education and video-on-demand systems. Development of these various multimedia applications requires fast and efficient storage, browsing, indexing, and retrieval techniques. Since, video is typically stored efficiently in a compressed format, it is preferable to analyse the compressed representation directly. This avoids the costly overhead of decompressing and operating at the pixel level. Compressed domain parsing of video has been presented in earlier work where a video clip is divided into shots, sub-shots and scenes [1-7]. Yeo proposed an algorithm for sudden scene change detection by exploiting the DC component, which can be extracted from compressed video, e.g., MPEG. This method leads to partial decompression of video and hence it is very costly. Hierarchical scene change detection technique proposed by Shin [4] and Zhang [2], is another alternative technique for H.263 sequences. But, again this algorithm fails when adjacent frames are highly correlated (during

special effects like fade-in, fade-out, dissolving and wiping etc.). Furthermore this method depends on the number of intra-coded and inter-coded MBs. Therefore in this method, setting the thresholds are very critical and heavily depend on the sequences to be processed. The reason for this is, at the indexing level segmentation algorithm has no information about the threshold, which is used to insert I or P-frame at the encoder.

In this paper an algorithm is proposed based on DFD and number of intra-coded MBs for H.263 compressed sequence. Section 2 presents an overview for H.263 and section 3 describes the proposed DFD based algorithm. Section 4 presents some results with the proposed algorithm and section 5 the conclusion.

## 2. H.263 OVERVIEW

The H.263 video codec is the ITU-recommended standard [8] for very low bit-rate video compression. Currently, H.263 encoding and decoding is the predominant technology used in videoconferencing across analogue telephone lines, optimised at rates below 64 kbits/s. Primary goal of the H.263 encoder is to represent information as compactly as possible without compromising video fidelity. In the encoder, all frames in a video sequence are categorised as either I-frames or P-frames. I-frames, or intra-frames, are encoded without reference to any other frame in the sequence, in much the same manner that a still image would be encoded. In contrast, P-frames, or (predicted) inter-frames, depend on information from a previous frame for its encoding. Each frame in the sequence is further subdivided into "macroblocks," (MBs) which consists of a 16x16 pixel luminance block, plus its corresponding two 8x8 chrominance blocks. MBs in an I-frame are intra-MBs as they are all encoded independent of other MBs. In most cases, MBs in P-frames are inter-MBs, which depend on a previous MB for information. Inter-MB can be represented by two pieces of information: its motion vector, which indicates the location of the best-matching MB in the reference frame, and the prediction error (DFD), which is the difference between the prediction

and the target MB pixel values. If the motion estimation is poor, some MBs in P-frames can also be intra-MBs.

The next step in the encoding process is that of the Discrete Cosine Transform (DCT). The DCT is a two-dimensional linear transformation that maps values in the spatial domain into the frequency domain. For inter-MBs, it is the prediction error that is transformed. For intra-MBs, the pixel values themselves are directly transformed.

### 3. DFD BASED ALGORITHM

The ratio of the number of MBs without motion vectors to the number of MBs with motion vectors (here after the ratio) [2,4], is an economical method for compressed domain sudden scene change detection. According to this method, if an abrupt scene change occurs current and the previous P-frames are not correlated. Therefore the ratio is increased since, encoder needs to introduce more intra-coded MBs. However, there is a threshold ( $T_E$ , where,  $T_E$  = absolute sum of the difference between luminance values in a particular MB and its mean value - sum of absolute values of luminance DFD signal for the same MB) is present as the encoder to decide whether to introduce intra-coded or inter-coded MB. Therefore simply counting the number of intra-coded MBs or taking the ratio of the number of MBs without motion vectors to the number of MBs with motion vectors is not sufficient to identify an abrupt scene change. This is due to the number of intra-coded blocks and number of inter-coded blocks are functions of the encoder threshold. Therefore the compressed domain scene change detection algorithm should not only depend on the number of intra-coded blocks. Specially, in H.263 bit rate is highly limited and this leads the encoder to take all possible steps to predict current MBs from the previous frame. When it tries to do so, all inter-coded MBs must have a large DFD values and also the variance of these DFD values will be large. This is due to poor correlation between two frames when a scene change is occurred. Thus, in the proposed procedure, it has two stages. During the first stage it counts the number of intra-coded MBs. If it exceeds the predetermined percentage threshold ( $T_r$ , set it to a high value to avoid false detection), then the abrupt scene change is detected. If the above condition is not satisfied then, DFD information is exploited. In this stage, it calculates the mean ( $m_i$ ) and variance ( $\sigma_i$ ) of the DC term of the luminance DFD signals according to the equations (1) and (2). Then, mean square error ( $MSE_i$ ) is calculated with equation (3) and if this ratio is above

the pre-determined threshold ( $T_m$ ), then an abrupt scene change is declared.

$$m_i = \frac{\sum_{j=0}^{N-1} DFD_j(0,0)}{N} \quad (1)$$

$$\sigma_i^2 = \frac{\sum_{j=0}^{N-1} (DFD_j(0,0) - m_i)^2}{N} \quad (2)$$

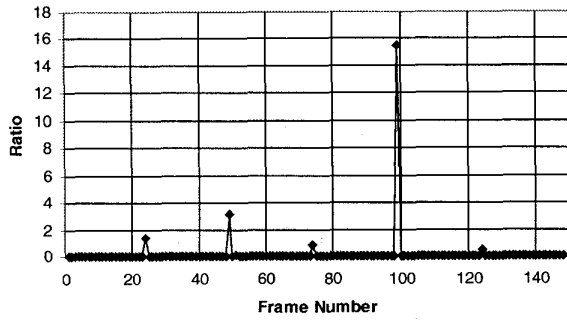
$$MSE_i = (m_i - m_{i-1})^2 + |\sigma_i^2 - \sigma_{i-1}^2| \quad (3)$$

where “N” is the total number of MBs and “i” is the frame number.

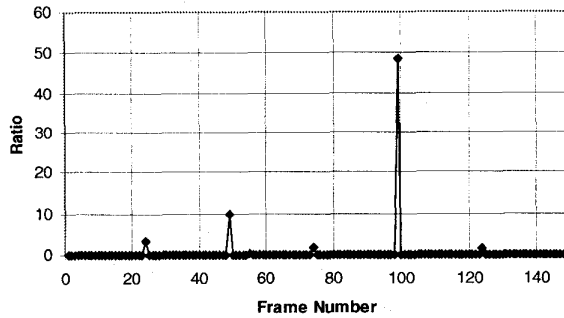
### 4. SIMULATION RESULTS

For the simulation, H.263 bit streams are used with 176x144 QCIF format. Here, five different abrupt scene changes are considered with the proposed algorithm. Results are compared against the ratio method [2,4]. According to the ratio method, scene change is declared if the ratio exceeds the pre-determined percentage threshold ( $T_r$ ). Results are also analysed with three different encoder threshold levels. Figure 1 – Figure 3 show the ratio of the number of MBs without motion vectors to the number of MBs with motion vectors for three different encoder thresholds of  $T_E$ . It is really impossible to set a threshold for  $T_r$  to detect all sudden scene changes. If  $T_r$  set to a low value, then probability of false detection is increased. Now, set  $T_r = 90\%$  (which is very large). This leads to missing a large number of sudden changes. This can be avoided using the DFD information as shown in Figure 4-6. They show the MSE of DFD signals against frame number. When,  $T_r < 90\%$ , algorithm consider the MSE value. If it is less than  $T_m (= 100)$ , there is no any sudden scene change occurred. If it is greater than  $T_m$  and previous frame MSE is not greater than  $T_m$  (This is necessary because, when a scene change is occurred MSE is large for the two consecutive frames), then scene change is declared. For instance, consider the fourth scene change at frame number 99. Figure 1 and Figure 2 show that the ratio is less than  $T_r$ . Therefore, it should go

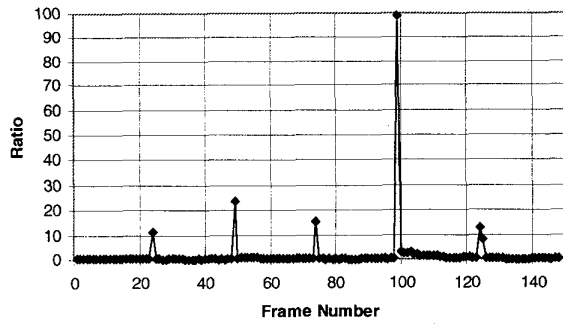
to the second stage and there, MSE is greater than  $T_m$  (Figure 4 and Figure 5). In Figure 3, ratio is greater than  $T_r$  and there is no need to proceed to the second stage. Therefore in all three cases, scene change is detected accurately. Similarly, other four changes are detected at frame 24, 49, 74 and 125. Table 1 shows the summarised results. It is clear that DFD method can detect all five abrupt scene changes irrespective of the encoder threshold where the ratio method fails.



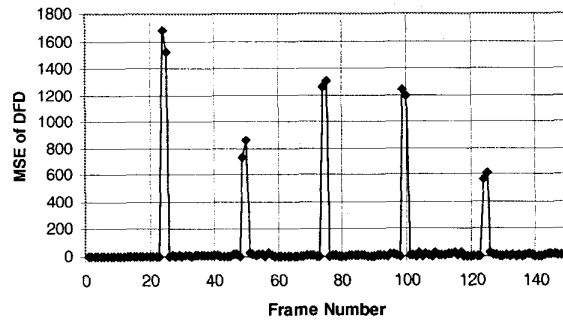
**Figure 1:** Ratio of number of intra and inter-coded MBs (%) with frame number ( $T_E = -2000$ )



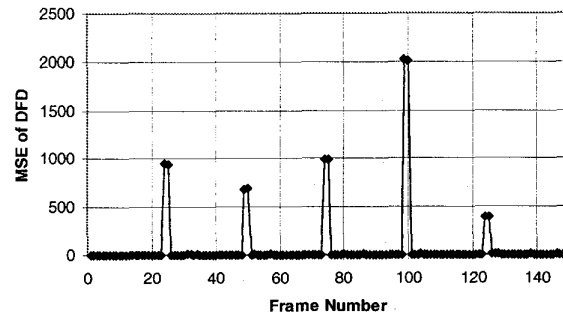
**Figure 2:** Ratio of number of intra and inter-coded MBs (%) with frame number ( $T_E = 0$ )



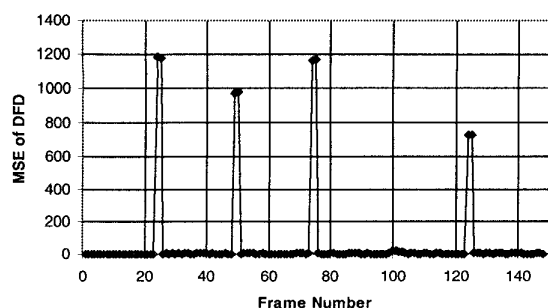
**Figure 3:** Ratio of number of intra and inter-coded MBs (%) with frame number ( $T_E = +2000$ )



**Figure 4:** MSE of DFD signal with frame number ( $T_E = -2000$ )



**Figure 5:** MSE of DFD signal with frame number ( $T_E = 0$ )



**Figure 6:** MSE of DFD signal with frame number  
( $T_E = +2000$ )

Encoder Threshold	Detected		Undetected		False Detected	
	Ratio	Prop.	Ratio	Prop.	Ratio	Prop.
-2000	3	5	2	0	0	0
0	2	5	3	0	0	0
2000	5	5	0	0	6	0

**Table 1 :** Performance comparison of the DFD based algorithm ( $T_m = 100$ ,  $T_r = 90\%$ ) against ratio method ( $T_r = 2\%$ ) for the sequence considered. (where, Ratio - Ratio Method and Prop. - Proposed algorithm.)

Encoder Threshold	Detected		Undetected		False Detected	
	Ratio	Prop.	Ratio	Prop.	Ratio	Prop.
-2000	14	30	16	0	2	0
0	12	30	18	0	5	0
2000	25	30	5	0	45	0

**Table 2 –** Performance comparison of the DFD based algorithm ( $T_m = 100$ ,  $T_r = 90\%$ ) against ratio method ( $T_r = 2\%$ ).

Proposed algorithm is also tested with video special effects such as fading, tilting, panning and observed that the algorithm is capable of detecting all sudden scene changes unlike ratio method produced many false detections. Table 2 shows the summarised results for a sequence, which is having special effects and several sudden changes. This scene is having 30 sudden changes. Results show that the proposed algorithm can detect all scene changes accurately and ratio method fails to declare the exact sudden scene changes.

## 5. CONCLUSIONS

This paper presents a real time algorithm, which can detect abrupt scene changes in H.263 compressed video. This algorithm is proposed based on DFD signal and the number of intra-coded MBs. Results are compared

against well-known ratio method. Experimental results show that the performance of this algorithm is independent of the encoder threshold unlike the ratio method. Furthermore, this algorithm is capable of detecting abrupt scene changes accurately even when the video sequence contains special effects. Results indicate that the proposed algorithm based on DFD can be used for H.263 compressed video with a high reliability rate. Further work is required to extend this algorithm for MPEG-2 compressed video.

## ACKNOWLEDGEMENTS

Authors would like to express their gratitude and sincere appreciation to the university of Bristol and CVCP for providing financial support for this work.

## REFERENCES

- (1) Kasturi and Jain, R., "Dynamic Vision," Computer Vision: Principles, R. Jain, IEEE Computer Society Press, Washington, pp. 469-480, 1991.
- (2) Zhang, H.J., Kankanhalli, A., and Smoliar, S.W., "Automatic Partitioning of Full-Motion Video," ACM/Spinger Multimedia Systems, Vol.1, No.1, pp. 10-28, 1993.
- (3) Nagasaka, A., and Tanaka, Y., "Automatic Video Indexing and Full-Video Search for Object Appearances," Visual Database Systems II, Eds.E.Kunth, and L.M. Wegner, Elsevier Science Publishers B.V., IFIP, pp. 113-127,1992.
- (4) Shin, T., "Hierarchical scene change detection in an MPEG-2 compressed video sequence", Paper Number 1099, TPA 13-6, ISCAS '98.
- (5) Nagel, N.H., "Formulation of an Object Concept by Analysis of Systematic Time Variation in the Optically Perceptible Environment," Computer Graphics and Image Processing, Vol-7, pp. 149-194, 1978.
- (6) Lee, S.Y., and Kao, H.M., "Video Indexing - An Object Based on Moving Object and Track," Proc. IS&T/SPIE, Vol. 1908, pp. 25-36, 1993.
- (7) Yeo, B.L., "Rapid Scene Analysis on Compressed Video", IEEE Tans. on ccts. and sys. for video tech., Vol5, No 6, December, 1995.
- (8) ITU-T, Draft H.263, May 1996.