An Automatic Cut Detection Algorithm Using Median Filter And Neural Network ITC-CSCC'2000

Seung-Chul Jun and Sung-Han Park
Department of Computer Science & Engineering
HanYang University, Ansan, KyungKi-Do
Tel: +82-345-400-4109, Fax: +82-345-406-4673
E-mail: {scjun, shpark}@cse.hanyang.ac.kr

Abstract: In this paper, an efficient method to find cut in the MPEG stream data is proposed. For this purpose, histogram difference and pixel difference is considered as a noise signal. The signal is then filtered out by a median filter to make the frame difference larger. The frame difference obtained in this way is classified into cut frame and non-cut frame by the 2-means clustering without using any threshold value. To improve the classification ratio, a back-propagation neural network is constructed, where outputs of 2-means clustering are used as the inputs of the network. The simulation results demonstrate the performance of the proposed methods.

1. Introduction

An automatic video parsing is needed to search and index a video data. The video parsing method consists of video segmentation and video indexing. In the video segmentation, video stream is segmented into shots as the basic unit of physical meaning and scenes as the basic unit of logical meaning. A shot is the smallest unit of video data which is recorded by continuous camera moving and do not be edited. A shot has various transitions.(cuts, fade, dissolve, wipe, etc.). To find cut frame corresponding to shot boundary having those various transitions is the core technology for constructing efficient content based video indexing and searching. The cut frame is the frame which has the largest total differences between the all pixels of the current frame and those of the previous frame.
The video data to be parsed, in general, is the MPEG compressed data. There are two parsing methods depending on how the MPEG data is used. The first method using the raw video data after decoding all of the MPEG data[1]-[3] takes long time for decoding the MPEG data. The other one uses the partially decoded MPEG data[4]-[8]. This method has no overhead to decode all of the MPEG bitstream.

The most instinctive way of finding cut is to calculate the pixel difference measure(PDM) between consecutive two frames. This method can implement very easily but has many problems. It is very sensitive to the camera movement. The false cut can also be detected because when flash light occurs. The use of Histogram difference measurements (HDM) is the general method of finding cut in the raw video data. This method is simple and efficient but ignore the spatial information of the frames.

Several methods have been developed to solve these problems. Deng proposed the video segmentation method using the weighted sum of several measurements rather than only one measurement[1]. It finds cut frames using the weighted sum of the color, the texture and the motion measurement of every frames. But it can not simply decide the weight values of these three measurements when those are merged. In addition, a concrete method to know the threshold value which is used to detect cut frames is not provided. Emile proposed a method that searches the video data using the object's moving vectors rather than the difference measurements of every frames[2]. He gets the object's moving trajectory between several frames from the motion vectors in the video data. Then the video is searched using the difference of trajectory, which is computed by comparing the trajectory of query and one of database. This method is suitable for when searches one object base, but it is very hard to find any one object's trajectory when many objects exist in the video data. Naphade convert the problem of the cut detection into the process of clustering the two feature vectors, the HDM and PDM of each frame, into two groups, cut and not-cut[8]. For detecting cut using the measurements retrieved by HDM and PDM methods, the threshold value[1] or k-means clustering algorithm [8] can be applied. However, these methods detect many not-cuts as cuts(false positive) because the measurements of frame differences do not distinct cuts and not-cuts. So they have to do the post-processing step of eliminating illegal cuts.

To solve these problems, in this paper, we propose an efficient parsing process which extracts the needed data with the partially decoded MPEG bitstream. To make the clear measurements between cut and not-cut frames, the measured data is treated as a signal and applied to a median filter. The comparison of the original signal and the filtered signal provides a clear difference measure. The filter outputs are clustered by 2-means clustering and then used as the input data set of a back-propagation neural network. Section 2 explains the method to calculate measurements used in this paper. The proposed algorithm is presented in Section 3. Experimental results are presented in Section 4, followed by the conclusion in Section 5.

2. The Difference Measures of Frames

For representing the variation between frames, the HDM and PDM are used in this paper. These two measurements complement each other's problem. The spatial information which is not exists in HDM is exist in

PDM. The sensitivity of camera moving in PDM is complemented by HDM. These measurements are obtained between I-pictures of the MPEG bit-stream. To further save the decoding time of the overall MPEG data, the MPEG data is partially decoded by only considering the Y component of the DC image of I-picture rather than the whole image. The HDM can be calculated by the equation (1).

$$HDM(f_t, f_{t+1}) = \frac{1}{M \times N} \sum_{j=1}^{256} |H_t(j) - H_{t+1}(j)| \qquad (1)$$

where f_t is a video frame in time t and f_{t+1} is a video frame in time $t+1$. $H_t(j)$ is the histogram value of j's bin in the video frame of time t. The HDM may have a large value in the case of rapidly changing color distribution. The second measurement, PDM, is given by the equation (2).

$$PDM(f_t, f_{t+1}) = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} d_{i,j}(f_t, f_{t+1}) \qquad (2)$$

Where $d_{i,j}(f_t, f_{t+1})$ is the intensity difference of a pixel at position (i,j) between frames at time t and $t+1$. The difference function d is presented as the equation (3).

$$d_{i,j}(f_t, f_{t+1}) = \begin{cases} 1 & if \ |I_{i,j}(f_t) - I_{i,j}(f_{t+1})| > 0 \\ 0 & otherwise \end{cases} \qquad (3)$$

The pixel differences measured by the equation(3) measures the whole variation of frame rather than the abrupt variation of intensity of the partial frame. The processing time of these two measurements can be reduced by using only the Y component of the DC image of I-picture of MPEG bit-stream.

3. Proposed Algorithm

The cut detection process consists of two parts. The first part is to acquires the difference measurements presenting the features of each frame. The second part is to detect cut using the difference measurements. For this purpose, we first perform a preprocessing step for transforming the simple frame difference measurements into the more distinct measurements for the easy discrimination between cut and not-cut. The next process is to detect cut using the back propagation neural network in which the measurements determined in the first part are used as the input patterns of the network.

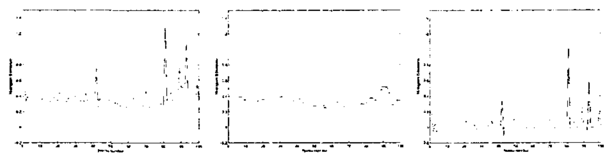
3.1 Preprocessing

We consider the difference measures retrieved from each frames is considered as a white noise signal. Figure 1(a) presents the HDM of 100 frames as a signal. In this signal, the value of cut can be treated as a noise of original signal. This noise has the character of white noise since occurring independently with other signals. To detect the white noises, a median filter can be used. Figure 1(b) presents the signal obtained by applying the median filter to the original signal. In this paper, to make the distinction between cuts and not-cuts, we subtract the Median filtered signal from the original signal. Figure 1(c) shows the subtracted signal. As shown in figure 1(c), the measurement of cut and not-cut frames is nearly 0 value and larger than 0, respectively. The measurements corresponding to the cut are removed as a noise by the median filter and then become a large value by the difference between the original signal and the filtered signal. However, the measured value of the not-cut data is not affected and the difference between the original signal and the filtered signal is very small. Therefore, the detection of cuts becomes easier and more accurate since the cut data has a large value than zero in the difference signal. The same procedures for the PDM are shown in figures 1(d) to (f)

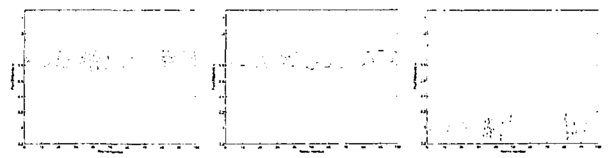
The equation (4) shows the process of acquiring the measurements proposed in this paper.

$$MFD(t) = DM(t) - mf(DM(t)) \qquad (4)$$

where $DM()$ is the difference measure function such as PDM or HDM and $mf()$ is the median filter function.



(a) Histogram Difference Signal (b) Median fitered signal of (a) (c) The difference between (a) and (b)



(d) Pixel Difference Signal (e) Median fitered signal of (d) (f) The difference between (d) and (e)

Fig. 1. Median Filtered Frame Differences

3.2 Cut Detection Using Neural Network

The method that decides cut using k-means clustering with the measurements of each frame shows the different result depending on distributions of measurements of the video data[8]. Therefore, it cannot detect cut exactly. In a critical moment it occurs that cut frames is considered as not-cut frames. To solve this problem, this paper proposes a method that detect cut using back-propagation neural network.

The network has two input nodes, six hidden nodes and two output nodes. The network is trained with 2600 sample data. The data of input nodes are two difference measurement retrieved at Section 3.1. If the data of input node is a cut, the output result of data is [1, 0]. If the data of input node is a not-cut, the output is [0, 1]. The input data is normalized at the previous stage. Since two measurements of not-cut frame is near 0, the proposed neural network can efficiently detect cut frame. The initial training set used in the neural network is acquired by k-means clustering algorithm. The more precise class data is acquired by user selection.

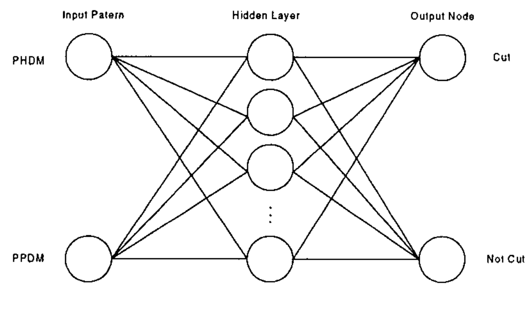
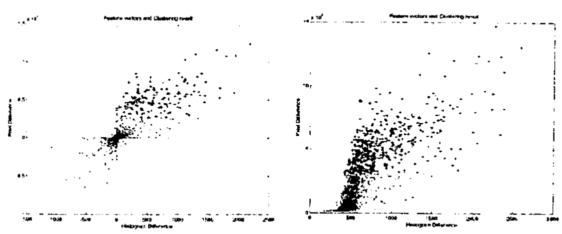


Fig.2. Backpropagation Network

4. Simulation Results

The TV news video data is used for the simulation. Figure 3 represents the difference between the proposed method and the previous method. Figure 3(a) and (b) are a two-dimensional representation of the results.of the cut detection by using the proposed method and k-means clustering of HDM and PDM values. In the figure, the dot('.') signs of the figure present the not-cut frames and the plus('+') signs present the detected cut frames. As shown in figure 3(a), the measurements of not-cut frame is located at nearly origin coordinate(0,0).

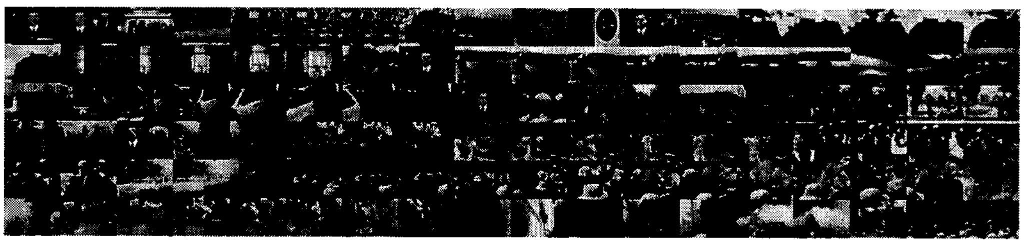


(a) The proposed method (b) The conventional method

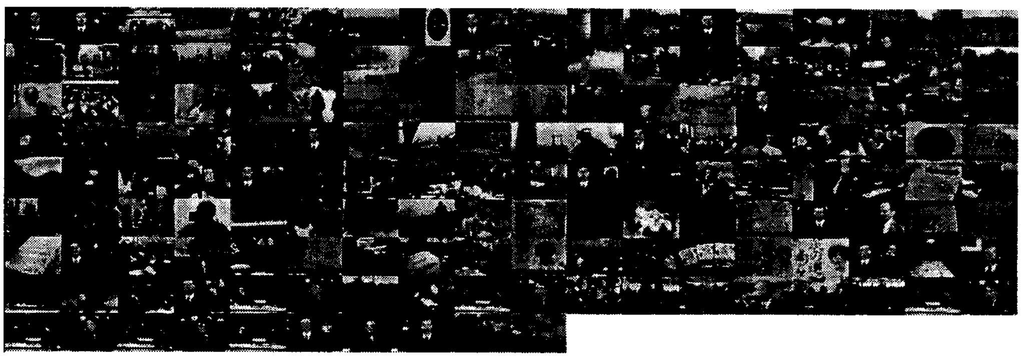
Fig. 3. The Results of Cut Detection

Figure 4 presents the DC images of Y component of detected cut frames. As shown in figure 4(a), many similar frames are detected. These false positive result make it difficult to retrieve the precise cut frame without post-processing. Figure 4(b) is the result of detecting cut using the proposed method. Figure 4(a) demonstrates that many false positives shown in figure 4(b) are eliminated. The cut detection using neural network trained by the proposed measurements has the 99% precision. Figure 4(b) presents the cut detection results of applying the simple k-means clustering and using the neural network.

The number of news video used is two news video. The news video is coded by MPEG-1. The first video is 20 minutes(30765 frames) and the second is 25 minutes(32730 frames). The GOP of the used MPEG data consists of 15 frames. All frames are 352x240 in size. The first video clip is training data. There are 168 cut frames and 1884 not-cut frames. Table 1 presents the experimental result of these two video clip. In the table 1, we can see that the method using back-propagation network with proposed measurements shows better result than k-means clustering algorithm in terms of the false positive and number of scenes detected.



(a) Cut images obtained by the proposed method



(b) Cut images obtained by the k-means clustering

Fig. 4. The Images of Detected Cut Frames

Table 1. The Comparison of simulation results for two videos

Video		Naphade Method	Proposed Method
News 1	# of frames	30765	
	# of scenes	164	
	# of scenes detected	468	168
	# of false positive	304	4
	# of miss-detection	0	0
News 2	# of frames	32730	
	# of scenes	215	
	# of scenes detected	496	219
	# of false positive	218	4
	# of miss-detection	0	0

5. Conclusion

This paper proposes an automatic cut and shot boundary detection algorithm in MPEG video stream without threshold value. For this purpose. we first preprocess the HDM and PDM using the median filter. Then, the back-propagation neural network is employed to exactly classify the cut and not-cut frames. The simulation results show that the most of false positive are eliminated. To solve simple k-means clustering algorithm for classifying the cut and not-cut with, a back propagation neural network is employed. The experimental result shows 99% precision of cut detection in TV news video. For the future research work, the semantic merging of the shots with this results will be necessary

References

[1] Y. Deng and B.B.Manjunath, "Content-based Search of Video Using Color, Texture and Motion," *1997 IEEE International Conference on Image Processing,* volume I, pp 534-537.

[2] E. Sahouria ans A. Zakhor , "Motion Indexing of Video," *1997 IEEE International Conference on Image Processing, 1997,* volume I, pp 526-529.

[3] K. J. Han and Ahmed H.Tewfik , "Eigen-Image Based Video Segmentation and Indexing,"*1997 IEEE International Conference on Image Processing, October 26—29,* volume II, pp 538-541.

[4] B. L. Yeo and B. Liu, "Rapid scene change detection on compressed video,"*IEEE Transaction on Circuit and Systems for Video Technology,* Vol 5, Num 6, pp 533-544, Dec, 1995

[5] Y. Nakajima , "Universal scene change detection on MPEG-coded data domain," *VisualCommunications and Image Processing '99, 12--14 February 1997 San Jose, California,* volume 3024.

[6] T. Kaneko and O. Hori , "Cut Detection Technique from MPEG Compressed Video Using Likelihood Ratio Test,"*14th International Conference on Pattern Recognition, August 16--20, 1998 Brisbane, Australia,* volume II, pp 1476-1480

[7] M. Lee and B.W. Hwang and S. Sull and S.W. Lee, "Automic video Paring Using Shot Boundary Detection and Camera Operation Analysis," *14th International Conference on Pattern Recognition, August 16--20, 1998 Brisbane, Australia,* volume II, pp 1481-1483

[8] M.R. Naphade and R.Mehrotra and A.M. Ferman and J. Warnick and T. S. Huang and A. M. Tekalp, "A High-Performance Shot Boundary Detection Algorithm Using Multiple Ques," *1998 IEEE International Conference on Image Processing, 1998,* volume I, pp 884-887

[9] M. Sugano and Y. Nakajima and H. Yanagihara and A Yoneyama , "A Fast Scene Detection on MPEG Docing Parameter Domain," *1998 IEEE International Conference on Image Processing,* volume I, pp 888-892