A Frame Quality Improvement before Video Shot Boundary Detection

Songpon Nakharacruangsak[1],
Maleerat Sodanil[2],
Department of Information Technology,
Faculty of Information Technology,
King Mongkut's University of Technology North Bangkok, Thailand
[1]songpon_n@hotmail.com
[2]msn@kmutnb.ac.th

and Supot Nitsuwat[3]
Department of Mathematic,
Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok, Thailand
[3]sns@kmutnb.ac.th

Abstract - **Video shot boundary detection is the first important step for the research in the following fields, automatic content analysis, content based-indexing and retrieval. In this paper, firstly the researcher presented, efficient method for video frame quality improvement to reduce noise, the flash-light problem and abrupt change illumination occurred within frame using logarithm transform with contourlet transform. Secondly, for shot boundary detection, we used gray-scale histogram differences with edge change ratio. Finally, we proposed the adaptive threshold algorithm based on a sliding window for shot transition detec-tion in order to get more accurate and precise detection. The experiment was then conducted on the standard video, and results are revealed that this pro-posed method can gain high precision and recall.**

Keywords - **Shot Boundary Detection, Adaptive Threshold, Logarithm Transform, Contourlet Transform**

I. INTRODUCTION

Currently, digital video is widely pro-duced and used due to advanced multimedia technology. Also, the rapid growth of Internet and computer technology results in a tremendous number of digital videos. Therefore, many researches were interested in the development of how to manage, search, retrieve and conduct the data mining of digital video as efficiently as possible for many years ago [1].

Video structure analysis is the first im-portant step of managing video before au-tomatic video content analysis. The analysis results are then utilized in indexing for digi-tal video retrieving later. In various video structure (frame, shot, scene, etc.), the first investigation is video shots. Therefore, a sequence of frame within video shot is suit-able for searching and retrieving video data [2].

The frames taken by the only one camera are arranged to be shots with the continuous action in the same area and time. An activi-ty between one shot to the next shot is the shot transitions which is the main problem because the location of video shot boundary has to be identified. Generally, there are two types of shot transition: 1) abrupt tran-sition: CUT, shot transitions in one frame and 2) gradual transition: GT. There are many kinds of the editing technique used in the GT such as dissolve, wipe and fade in /out that could appear between the last frames of the previous shots and the begin-ning frames of the next shots. Therefore, the detection is harder than that of the ab-rupt transition because it depends on these techniques of editing [3]. Hence, the point of the difference of two consecutive frames would show shot transitions that need deci-sion function to evaluate threshold. If the result of function is more than threshold, it means that the two frames are different shots. There are a lot of techniques to au-tomatically detect the shot boundary by calculating the difference between frames such as the difference of pixels [3, 4] the difference of statistic [5, 6] edge change ratio [7, 8] and comparison of histogram [1, 9]. However, if flash or abrupt change illumination active in a frame of shots, it may introduce the wrong difference of two consecutive frames that causes the wrong detection.

Therefore, in this paper, we propose the improvement of video frames quality before video shot boundary detection to solve the flash and abrupt change illumination within frames. Then we used the comparison of gray-scale histogram difference, edge change ratio and adaptive threshold algo-rithm in order to identify the position of shot transition within the video.

The rest of this paper is organized into 5 sections. Section 2 describes about materials and methods. Video Data and Evaluation are presented in section 3. The experimental results and analysis are presented in section 4, and section 5 provides conclusions and future work.

II. MATERIALS AND METHODS

We proposed an algorithm for shot boundary detection which composed of four stages, as shown in the Fig. 1. Firstly, frame quality preparation for reduction of the light or flash effect in video frame though logarithm transform and noise reduction by contourlet transform are proposed. Next, the gray-scale histogram difference method and edge change ratio are used to find the frame differences. Both methods are utilized to detect both the abrupt transition and gradual transition. Finally, the detection results of such methods are merged to specify position of shot boundary.

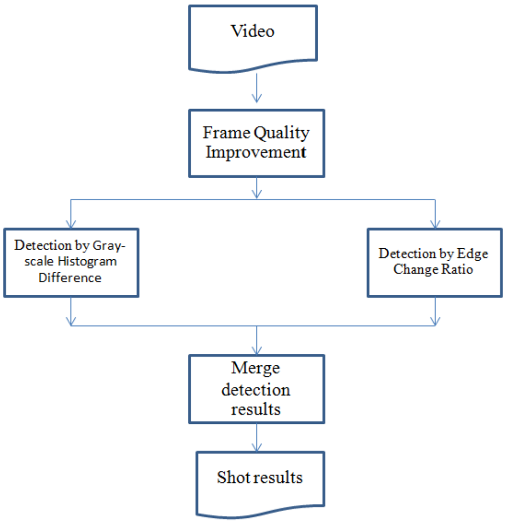


Fig. 1 Shot Boundary Detection Framework

A. Frame Quality Improvement

Zhang et. al [1] stated that the flash and illumination change in the two consecutive frames have strong effect on gray-scale histogram, i.e., given the high differences. It might make the detection of the abrupt transition in the wrong position. The sample of gray-scale histogram differences of video sequence with flashes is shown in Fig. 2.

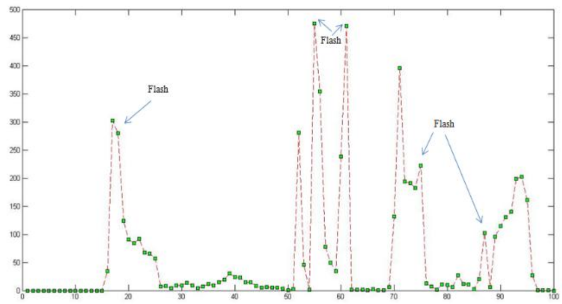


Fig. 2 The Sample of Gray-Scale Histogram Differ-Ences of Video Sequence with Flashes.

As illustrated in Fig. 2, two consecutive peaks appear in the position which a flash occurs. We applied logarithm transform to improve the frame resolution or contrast of frame in dark area or in low intensity area to have higher brightness. The calculation could be done by using;

$$g = c \times \log(1 + \text{double}(f)) \quad (1)$$

Where c is constant value and f is the required frames that need to be increased the brightness. Thus, if c is high, the frames would be more brightness, shown in Fig. 3.

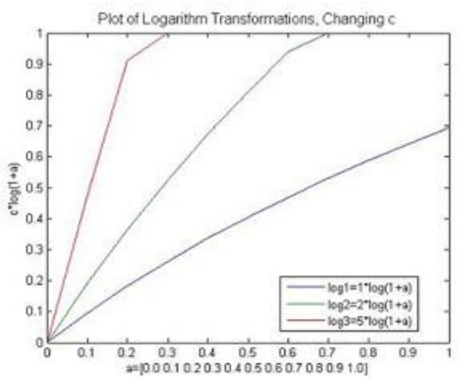


Fig. 3 Plot Logarithm Transform, when to Change the C Values

According to the proposed method above, under condition of the brightness of the light within a different frame, brightness adjustment may increase noise within the frame. Consequently, we used contourlet transform, proposed by Do and Vetterli [11], which combined Laplacian pyramid (LP) and Directional filter bank (DFB), to efficiently produce more details of sparse curve because of its direction and anisotropy. First of all, the result images after applying the logarithm transform are then divided into subbands by the Laplacian pyramid (LP). Detail of each image is analyzing by Directional filter bank (DFB), shown in Fig. 4. The 3-level LP and 8 directions contourlet coefficients of image frame are shown in Fig. 5. Examples of the reducing noise within frame by using the proposed method are shown in Fig. 6.

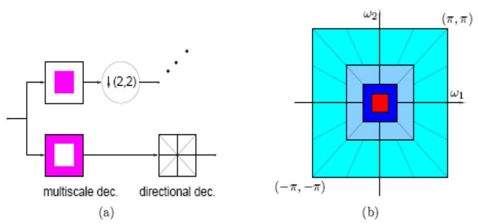
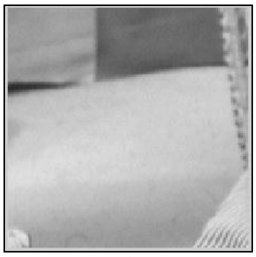
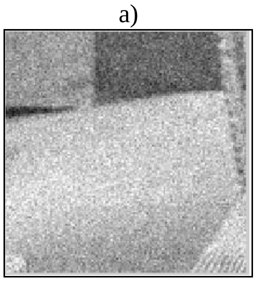


Fig. 4 Contourlet Transform Operation and Example of DFB Classified by Frequency

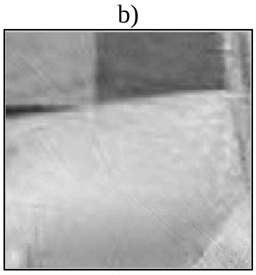
Fig. 5 Contourlet Transform of Babara by 3 LP and 8 Direction



a)



b)



c)

Fig. 6 a) Original Image
b) Original Image + noise (PSNR = 9.55 dB)
c) Reduced noise by Contourlet Tranform (PSNR = 13.08 dB)

We tested our proposed method with the video sequence with flashes. The results shown that, our method can suppress the flash problem and abrupt change illumination, shown in Fig. 7.

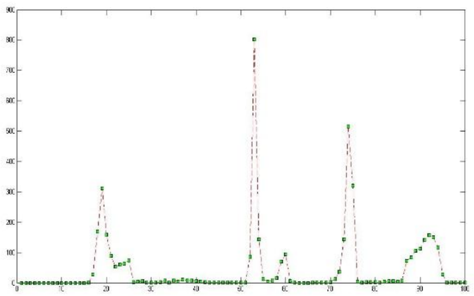


Fig. 7 The Gray-Scale Histogram Differences of Video Sequence after Suppress the Flash Problem by our Method.

B. Video Shot Boundary Features

Evaluation of efficiency in video shot boundary detection depends on function selection to be used to evaluate the two difference frames. We choose multiple features that combine the difference value of gray-scale histogram, calculated by using Euclidean distance and results from edge change ratio to identify the position of the changing shots. The features of the differences have been shown as follow:

1. **Gray-Scale Histogram Difference**

Research in image and video processing, most researchers usually take images in gray scale more than RGB color images because RGB color images composed of color, light and brightness values. It is difficult to compare them because of it is complicated processing. The transformation of RGB color images to gray scale images can be done by using the following equation;

$$Y=0.3 \times R+0.59 \times G+0.11 \times B \qquad (2)$$

Y is the gray-scale level at a specific pixel. R, G and B are the red, the green and the blue level at that pixel, respectively.

According to digital images, histogram gives number of pixels that have the same intensity value. In the histogram, the ab-scissa is range of intensity and the ordinate is the number of pixels. The calculation can be done using the equation;

$$H(i)=\frac{n_i}{N} \quad i=1,2,3,4,.......,k \qquad (3)$$

N is the number of all pixels in the im-age, n is the number of pixel of the i^{th} inten-sity, k is number of intensity value in the histogram. Therefore, histogram of image M is a vector $H(M) = (h_1,h_2,\ldots h_k)$.

Euclidean distance is the function used for evaluating the similarity of frames. It considers the distance of gray histogram between 2 consecutive frames. The calculation could be conducted by using the equation;

$$Dis(x,y)=\left[\sum_{i=0}^{n}|x_i-y_i|^2\right]^{\frac{1}{2}} \qquad (4)$$

Where x and y are the feature vector. n is the dimension of the vector.

To identify the position of shot transi-tion, we focus on the difference of gray-scale histogram between 2 frames. The cal-culation could be conducted by using the equation;

$$Dis(x,y)=\left[\sum_{i=0}^{n}|x_i-y_i|^2\right]^{\frac{1}{2}} \qquad (5)$$

Where Diff[i] is the difference between frame i and frame i-1, hi is gray-scale histogram, n is the dimension of frame i.

The difference of gray-scale histogram shows the frame positions of the shot transition. It is calculated by using equation 5, and is shown in Fig. 8.

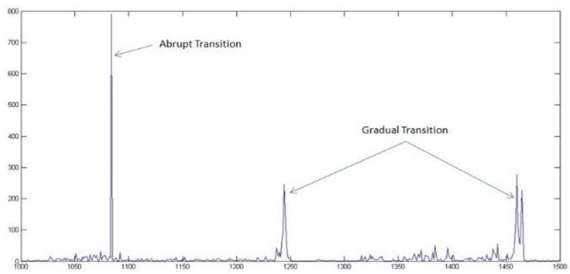


Fig. 8 The Example of the Difference of Gray-Scale Histogram for both the Abrupt Change and the Gradual Transition Change

The Fig. 8 shows the position of shot transition in video. At the abrupt transition, the higher difference of gray-scale histo-gram is given. It can be detected and recog-nized easier than the gradual transition.

2. **Edge Change Ratio**

Zabih et al. [10] presented edge change ratio: ECR. The main idea is that the continuous frames come along with the continuous structure. Therefore, the edge of object in the last frame before the hard cut usually cannot be found in the first frame of the next shot. The edge of object in the first frame after the hard cut usually cannot be found in the last shot before the hard cut. The ECR can be described in the following steps:

1) Detect edge of object in frame fn and fn+1, respectively.
2) Count the number of edge pixel δn and δn+1 in the fn and fn+1 frame
3) Define the entering and exiting edge pixels $En+1^{in}$ and En^{out}

To assume that two image frames, lm(n) and lm(n+1), the $En+1^{in}$ are fractions of edge pixel in the lm(n+1) frame which farther than fixed distance r away from the closed edge pixel in lm(n). Similarly, the En^{out} are fractions of edge pixel in the lm(n+1) frame which farther than fixed distance r away from the closed edge pixel in the lm(n+1) frame. Therefore, ECRn be-tween fn and fn+1 can calculated by:

$$ECR(n,n+1)=\left(\frac{En+1^{in}}{\delta n+1},\frac{En^{out}}{\delta n}\right) \quad (6)$$

If ECR is larger than a predefined thre-shold, the shot transition occurs at that frame. The process how to be performed on every frame in the video. The ECR sample is shown in the Fig. 9.

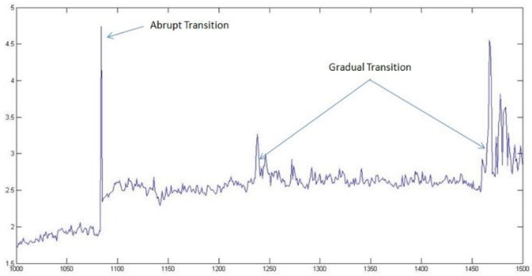


Fig. 9 The Difference of ECR

C. Shot Boundary Detection

In shot boundary detection, gray-scale histogram difference and edge change ratio need to be functioned for defining the cor-rect positioning of shot transition which are very important. Therefore, it is more appro-priate to set the threshold to suitable to that area. In this paper, adaptive threshold is used in principle of sliding window 2W+1, to compare the similarity of two consecu-tive frames. The calculation could be done by using to the following equation;

$$T_{(f_n)}=\Sigma\left(\frac{S\left(f_{n-1}+f_{n+1}\right)+c}{2w+1}\right) \quad (7)$$

In this paper, we defined w to 1, c is constant value and S(fn, fn+1) as the function of similarity of two consecutive frames such as fn and fn+1. Then we compares the results with T(fn). If the results are greater than T(fn), then the position of fn is shot transition. However, it is possible that S(fn,fn+1) is 0. Thus, it is necessary to specify c to 3-5 and apply to shot boundary detection with gray-scale histogram difference and edge change ratio respectively to detect abrupt transition and gradual transition, shown in Fig. 10 and 11.

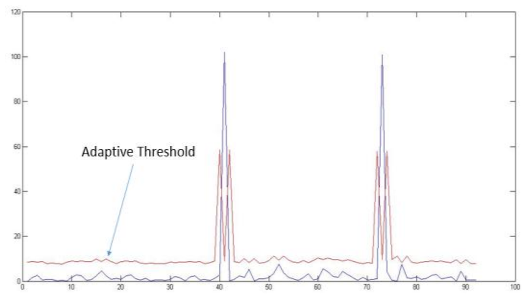


Fig. 10 Shot Boundary Detection by Using Gray-Scale Histogram with Adaptive Threshold

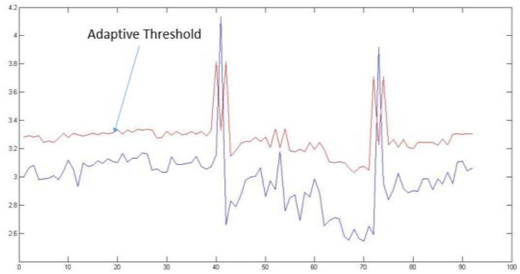


Fig. 11 Shot Boundary Detection by Using Edge Change Ratio with Adaptive Threshold

III. VIDEO DATA AND EALUATION

Video clips having inside operation such as object motion, camera panning and zooming were selected in order to evaluate the efficiency of the our proposed method. All data are downloaded from the Open Video Project (http://www.open-project.org). The evaluation of shot boun-dary detection is divided into abrupt transi-tions and gradual transitions their mea-surement methods are different. In addition, recall and precision are used as metrics which can be calculated as follows:

$$recall= \frac{N_c}{N_c+N_m} \text{ x } 100\% \quad (8)$$

$$precision= \frac{N_c}{N_c+N_f} \text{ x } 100\% \quad (9)$$

Where N_c is number of correct detection. Nm is number of missed and N_f is number of incorrect detection.

IV. RESULTS AND DSCUSSION

The results revealed that the proposed method about a frame quality improvement before shot boundary detection could sup-press the flash problem and abrupt change illumination. The experimental results shown, in table I and table II, that the ab-rupt transition and gradual transition can be obtained with high accuracy due to using the advantages of the detected object mo-tion and the number of pixels of edge in comparison with edge change ratio to be replaced disadvantages of gray-scale histo-gram difference which caused by compari-son between two consecutive frame with a similar color scheme. Those affected shot boundary detection to have more accu-racy and precision.

TABLE I
SHOT BOUNDARY DETECTION RESULTS

Video	Number of Frames	Number of Cut	Number of Gradual	Abrupt Detection			Gradual Detection		
				Nc	Nm	Nf	Nc	Nm	Nf
Anni003	4267	24	2	24	1	3	2	-	-
anni006	1379	0	2	-	-	-	2	-	-
anni007	1590	5	5	5	-	-	4	1	1
anni008	2775	2	12	2	-	-	12	3	1
Indi114	1813	5	2	5	-	-	2	-	-

TABLE II
PERFORMANCE EVALUATION

Abrupt Boundary Detection		Gradual Boundary Detection	
Recall	Precision	Recall	Precision
97.20%	92.30%	84.61%	91.66%

V. CONCLUSION

In this paper, a new method of a frame quality improvement before video shot boundary detection by using logarithm transform and reducing noise with coun-tourlet transform was proposed to suppress the flash problem and abrupt change illumi-nation within video. Then, frame differenc-es were calculated by gray-scale histogram difference with edge change ratio. We pro-posed adaptive threshold algorithm to detect abrupt transition and gradual transi-tion of the boundary. After performing some experimenting to evaluate the perfor-mance, we found that the precision value of the boundary position in there two types of shot boundary was in high-level. However, our proposed methods are not still absolute-ly successful because some shots are miss-ing and false positive occurs. As a result of this, in the future, we will improve this me-thod by combining multiple features such as motion feature within frame, and statistics so as to reduce missed shot detection and false positive.

VI. ACKNOWLEDGEMENT

This study was supported by Department of Information Technology, Faculty of Science and Technology, Southeast Bang-kok College. We are grateful to website www.open-video.com for supported video database. Thanks also to Associate Profes-sor Dr. Kanok Janchitrapongvej who pro-vided facilities for this work.

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)

[1] Zhang, H., Hu, R., and Song, L. (2011). “A Shot Boundary Detection Method Based on Color Feature”. International Confe-rence on Computer Science and Net-work Technology, IEEE 2011, pp. 2541-2544.

[2] Yuan, J., Wang, H., Xiao, L., Zheng, W., Li, J., Li, F., and Zhang, B. (2007). “A Formal Study of Shot Boundary Detection”. IEEE Transection on Circuits and Sys-tems for Video Technology, Vol. 17, No. 2, pp. 168-185.

[3] Borecsky, J.S. and Rowe, L.A. (1996). “Compari-son of video shot boundary detection techniques”. In Proceedings of SPIE, Vol. 2670, pp. 170-179.

[4] Hirzalla, N. (1997). “Media processing and re-trieval model for multimedia docu-ments”. PhD thesis, Ottawa University.

[5] Kasturi, R. and Jain, R. (1991). “Dynamic vision, in Computer Vision: Principles (Kasturi, R., Jain, R., and eds.)”. IEEE Computer Society Press, pp. 469-480.

[6] Kikukawa, T. and Kawafuchi, S. (1992). “Devel-opment of an automatic summary edit-ing system for the audiovisual re-sources”. IEEE Trans. on Electronics and Information, J75-A, pp. 204-212.

[7] Nasreen, A.S.G. (2013). “Key Frame Extrac-tion Using Edge Change Ratio for Shot Segmentation”. International Journal of Advanced Research in Computer and Communication Engineering Vol. 2 Issue 11, pp. 4421-4423.

[8] Mann, J. and Kaur, N. (2015). “Key Frame Extrac-tion from a Video using Edge Change Ratio”. International Journal of Ad-vanced Research in Computer Science and Software Engineering, Vol. 5, Issue 5, pp. 1228-1233.

[9] Wu, Z. and Xu, P. (2013). “Shot Boundary Detec-tion in Video Retrieval”. Electronics Information and Emergency Commu-nication (ICEIEC), 2013 IEEE 4th International Conference on 15-17 November 2013, pp. 86-89.

[10] Zabih, R., Miller, J., and Mai, K. (1995). “A feature-based algorithm for detecting and classifying scene breaks”. Proc. ACM Multimedia ‘95, San Francisco, CA, pp. 189-200.

[11] Do, M.N. and Vetterli, M. (2005). “The Con-tourlet Transform: an efficient direc-tional multiresolution image represen-tation”. IEEE Transaction on Image Processing 14(12), pp. 2091-2106.