

Video-frame Quality Improvement before Shot boundary Detection by using Logarithm, Wavelet and Contourlet Transform

Songpon Nakharacruangsak*

Faculty of Science and Technology, Southeast Bangkok College, 10260, Thailand

ARTICLE INFO

Article history:

Received: 30 April, 2017

Accepted: 06 June, 2017

Online: 20 June, 2017

Keywords:

Shot Boundary Detection

Adaptive Threshold

Logarithm Transform

Wavelet Transform

Contourlet Transform

ABSTRACT

Video shot boundary detection is important step for the research in the content analysis and retrieval fields. In this paper, firstly we presented an efficient method for Video-frame quality improvement to suppress flash occurred within video frame using logarithm, wavelet and contourlet transform. In addition, wavelet and contourlet transform also performed denoising. Secondly, for shot boundary detection, we used gray-scale histogram differences with edge change ratio. Furthermore, the adaptive threshold algorithm for shot transition detection was proposed. The experiment results showed that using logarithm with contourlet transform gain the precision and recall higher than using logarithm with wavelet transform.

1. Introduction

Currently, digital video is widely produced 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 important step of managing video before automatic video content analysis. The analysis results are then utilized in indexing for digital 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 suitable 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 activity 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 transition: 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 beginning frames of the next shots. Therefore, the detection is harder than that of the abrupt 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 decision 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 automatically 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 algorithm in order to identify the position of shot transition within the video.

The rest of this paper is organized into 4 sections. Section 2 describes about method and material. The experimental results and analysis are presented in section 3, and section 4 provides conclusions and future work.

* Songpon Nakharacruangsak, Southeast Bangkok College, Thailand

Email: songpon_n@hotmail.com

www.astesj.com

<https://dx.doi.org/10.25046/aj0203108>

2. Method and Material

We proposed an algorithm for shot boundary detection which composed of four stages, as shown in the Figure 1. Firstly, frame quality preparation for reduction of the light or flash effect in video frame through 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.

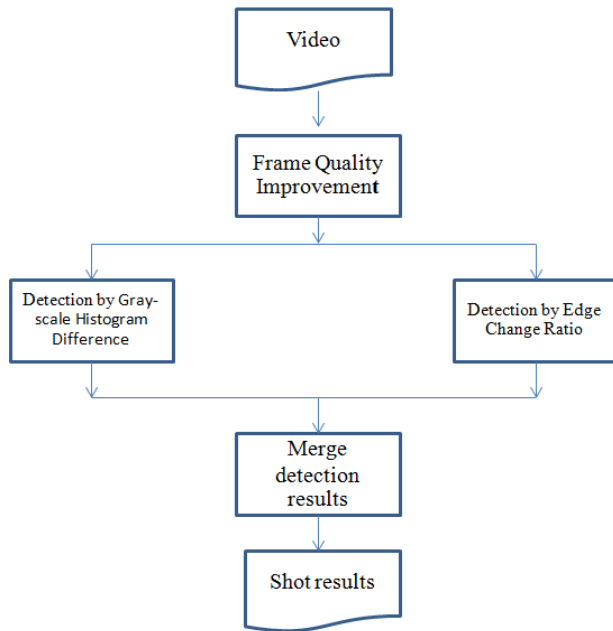


Figure 1: Shot Boundary Detection Framework

2.1. 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 Figure 2.

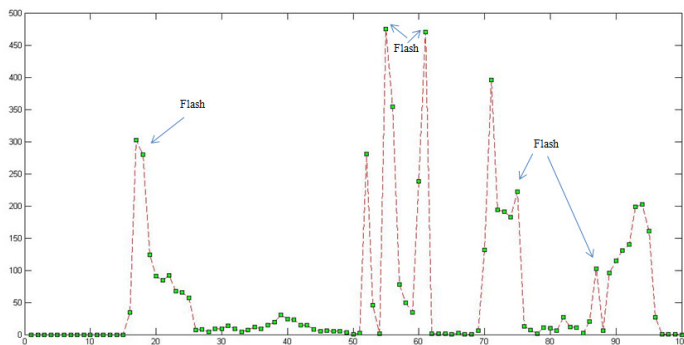


Figure 2: The sample of gray-scale histogram differences of video sequence with flashes

As illustrated in Figure 2, two consecutive peaks appear in the position which a flash occurs.

Logarithm Transform

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 Figure 3.

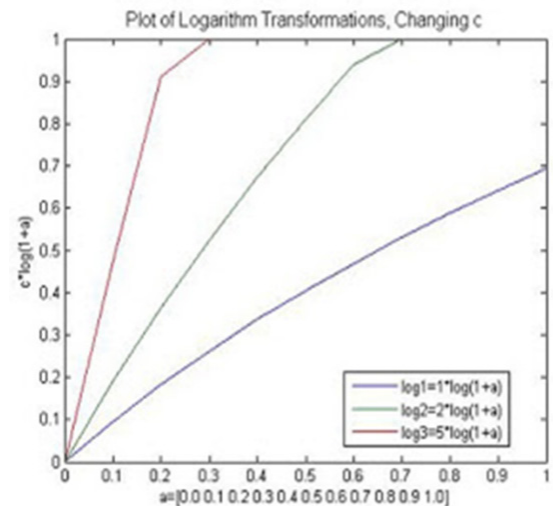


Figure 3: Shot Boundary Detection Framework

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 will reduce the noise inside the frame with wavelet transform and contourlet transform and compare the results of the two methods.

Wavelet Transform

We used wavelet transform to reduce noise, proposed by Li et al. [10] Firstly, this method decomposes the noisy image in order to get different sub-band image. Secondly, the low-frequency wavelet coefficients remain unchanged. After taking into account the relationship of horizontal, vertical and diagonal high-frequency wavelet coefficients and comparing them with Donoho threshold, we make them enlarge and narrow relatively. Thirdly, we use soft-threshold denoising method to achieve image denoising. Finally, the denoising image was achieved by performed inverse wavelet transform. The result shown in Figure 6 c)

Then, we used this method for video-frame quality improvement. The result shown in Figure 4.

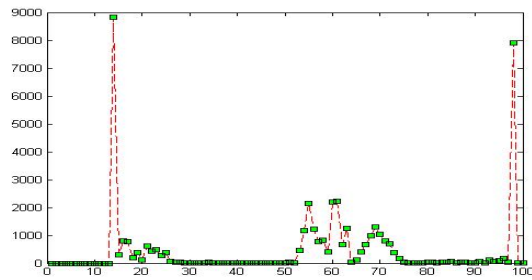


Figure 4: The gray-scale histogram differences of video sequence after suppress the flash problem by wavelet transform.

Contourlet Transform

Contourlet transform, proposed by Do and Vetterli [11], 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 Figure 5.

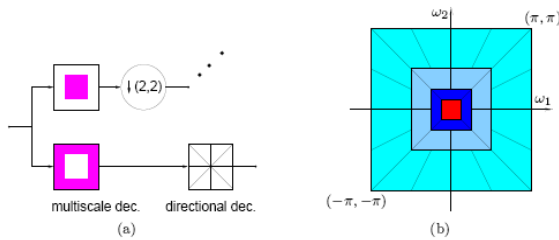


Figure 5: Contourlet Transform Operation and example of DFB classified by frequency

Then, we tested contourlet transform to denoising images. The results shown in Figure 5 d)

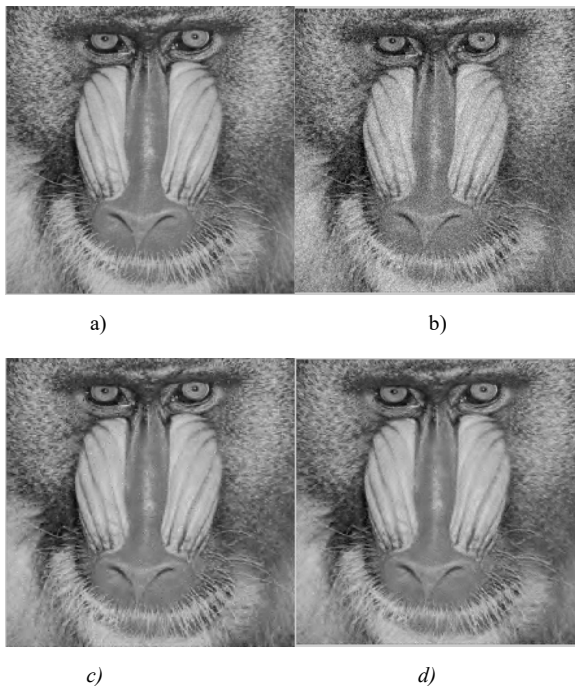


Figure 5: a) Original Images, b) Noisy Image (SNR = 9.52 dB) c) Denoise using Wevelet Transform (SNR = 12.38 dB), d) Denoise using Countourlet Transform (SNR = 13.45 dB)

Then, we tested contourlet transform method with the video sequence with flashes. The results shown that, the flash problem and abrupt change illumination can be suppressed, shown in Figure 6.

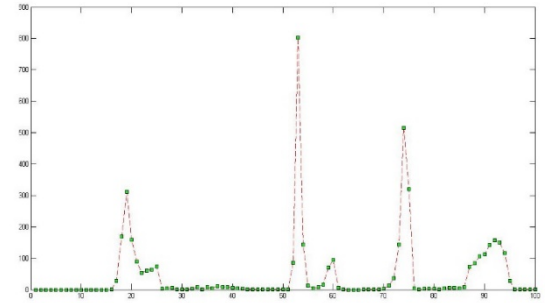


Figure 6: The gray-scale histogram differences of video sequence after suppress the flash problem by our method

2.2. 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 shot transition. The features of the differences have been shown in the following section.

Gray-scale Histogram Difference

For 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;

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

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 abscissa is range of intensity and the ordinate is the number of pixels. The calculation can be done using the equation;

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

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

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;

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

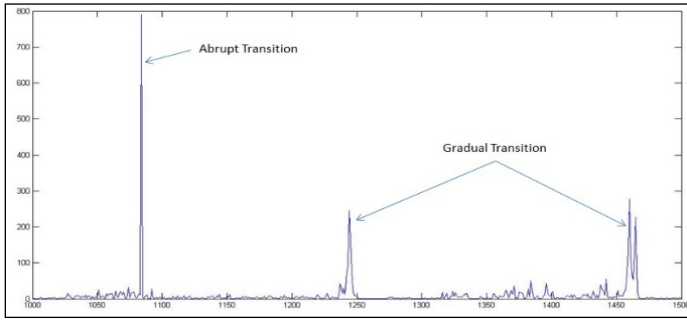
Where x and y are the feature vector. n is the dimension of the vector.

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

$$\text{Diff}[i] = \left[\sum_{j=0}^n |h_{i,j} - h_{i-1,j}|^2 \right]^{\frac{1}{2}} \quad (5)$$

where $\text{Diff}[i]$ is the difference between frame i and frame $i-1$, h_i 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 Figure 7.



Edge Change Ratio

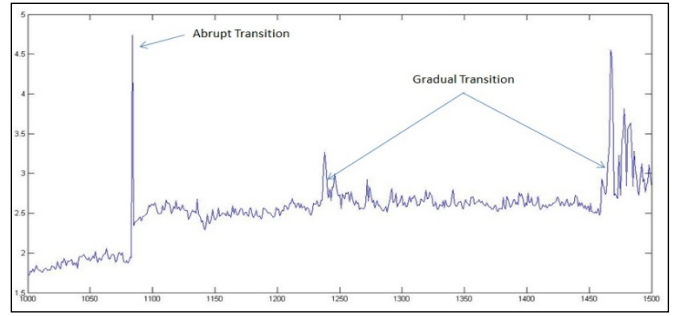
Zabih et al. [12] presented edge change ratio: ECR. The main idea is that the continuous frames come along with the continuous structure. Therefore, the edge of object in the last frame before the hard cut usually cannot be found in the first frame of the next shot. The edge of object in the first frame 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 f_n and f_{n+1} , respectively.
2. Count the number of edge pixel δ_n and δ_{n+1} in the f_n and f_{n+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, ECR_n between f_n and f_{n+1} can be 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 threshold, 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 Figure 8.



2.3. Shot Boundary Detection

In shot boundary detection, gray-scale histogram difference and edge change ratio need to be functioned for defining the correct positioning of shot transition which are very important. Therefore, it is more appropriate 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 consecutive frames. The calculation could be done by using the following equation;

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

In this paper, we defined w to 1, c is constant value and $S(f_{n-1}, f_{n+1})$ as the function of similarity of two consecutive frames such as f_n and f_{n+1} . Then we compares the results with $T(f_n)$. If the results are greater than $T(f_n)$, then the position of f_n is shot transition. However, it is possible that $S(f_{n-1}, f_{n+1})$ 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 Figure 9 and Figure 10.

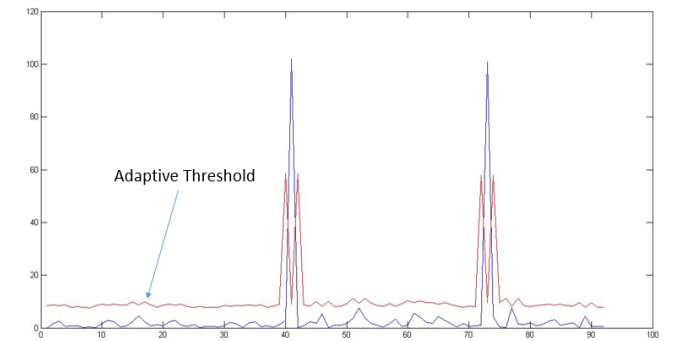


Figure 9: The difference of ECR

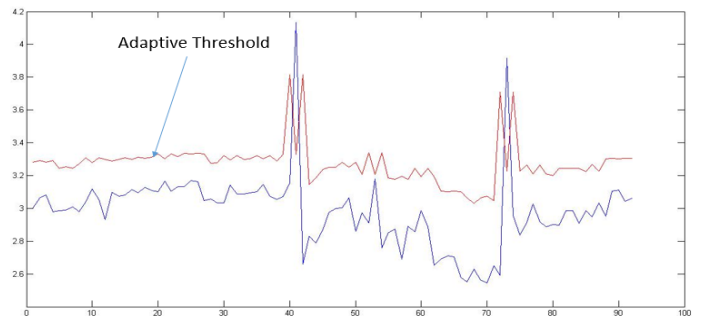


Figure 10: The difference of ECR

After, we obtained the results of shot boundary detection from Gray-Scale Histogram Difference and Edge Chang Ratio shown in Table 1.

Table 1: Results of Shot Boundary detection from two methods

Frame#	Shot1	Shot2	Shot3	Shot4	Shot5	Shot6	Shot7
Histogram Difference	998	1167	1292	1359	2081	2184	
ECR	86	998	1167	1292	2081	2184	2312

We will merged the results of two methods and selected the duplicate number of frames to define position of shot transition. The results are shown in Table 2.

Table 2: The Position of Shot Transition

Shot No.	Shot 1	Shot 2	Shot 3	Shot 4	Shot 5
Frame Number	998	1167	1292	2081	2184

2.4. Video Data and Evaluation

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-video.org>). The evaluation of shot boundary detection is divided into abrupt transitions and gradual transitions their measurement methods are different. In addition, recall and precision are used as metrics which can be calculated as follows:

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

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

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

3. Results and Discussion

The results revealed that the proposed method about a frame quality improvement before shot boundary detection could suppress the flash problem and abrupt change illumination. The experimental results of shot boundary detection by denoising use wavelet transform and contourlet transform shown in Table 3 and are shown performance evaluation of each method in Table 4, that the abrupt transition and gradual transition can be obtained with high accuracy due to using the advantages of the detected object motion and the number of pixels of edge in comparison with edge change ratio to be replaced disadvantages of gray-scale histogram difference which caused by comparison between two consecutive frame with a similar color scheme. Those affected shot boundary detection to have more accuracy and precision.

Table 3: Shot Boundary detection Results

Video	Number of Frames	Number of Cut	Number of Gradual	Using Wavelet Transform			Using Contourlet Transform		
				N_c	N_m	N_f	N_c	N_m	N_f
Vipsce	93	2	-	1	1	-	2	-	-
FlashPreset	398	7	-	5	2	-	6	1	-
anni006	1590	5	5	8	2	-	8	2	
anni007	2775	2	12	11	3	1	13	1	1

Indi114	1813	5	2	7	-	-	7	-	-
---------	------	---	---	---	---	---	---	---	---

Table 4: Performance Evaluation

Wavelet Transform		Contourlet Transform	
Recall	Precision	Recall	Precision
80.00%	96.96%	90.00%	97.29%

4. Conclusions

In this paper, we proposed a new method for a frame quality improvement by using logarithm transform and denoising by wavelet transform compared contourlet transform, which can to suppress the flash problem and abrupt change illumination within video. Then, frame differences were calculated by gray-scale histogram difference with edge change ratio. We proposed adaptive threshold algorithm to detect abrupt transition and gradual transition of the boundary. After performing some experimenting to evaluate the performance, we found that the precision and recall value of the boundary position in using wavelet transform less than contourlet transform. However, both methods are not still absolutely successful because some shots are missing and false positive occurs. As a result of this, in the future, we will improve this method by combining multiple features such as motion feature within frame, and statistics so as to reduce missed shot detection and false positive.

Conflict of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This study was supported by Department of Information Technology, Faculty of Science and Technology, Southeast Bangkok College. We are grateful to website www.open-video.com for supported video database.

References

- [1] H. Zhang, R. Hu, L. Song, "A Shot Boundary Detection Method Based on Color Feature", International Conference on Computer Science and Network Technology, Computer Science and Network Technology (ICCSNT), 2011, pp.2541-2544. <https://doi.org/10.1109/ICCSNT.2011.6182487>
- [2] J.Yuan, H. Wang, L. Xiao, W. Zheng, J. Li, F. Li and B. Zhang, "A Formal Study of Shot Boundary Detection", IEEE Transection on Circuits and Systems for Video Technology, Vol.17, No.2, February 2007, pp.168-185. <https://doi.org/10.1109/TCSVT.2006.888023>
- [3] J. S. Borecky, L. A. Rowe, "Comparison of video shot boundary detection techniques", In Proceedings of SPIE, vol. 2670, pp. 170-179, 1996. <https://doi.org/10.1117/12.238675>
- [4] N. Hirzalla, "Media processing and retrieval model for multimedia documents", PhD thesis, Ottawa University, Jan. 1997
- [5] R. Kasturi, R. Jain, Dynamic vision, in Computer Vision: Principles (R. Kasturi and R. Jain, eds.), IEEE Computer Society Press, 1991, pp. 469-480.
- [6] T. Kikukawa, S. Kawafuchi, "Development of an automatic summary editing system for the audiovisual resources", IEEE Trans. on Electronics and Information, J75-A, pp. 204-212, 1992.
- [7] A.Nasreen, S. G., "Key Frame Extraction Using Edge Change Ratio for Shot Segmentation", International Journal of Advanced Research in Computer and Communication Engineering Vol.2 Issue 11, November 2013, pp.4421-4423.
- [8] J. Mann, N. Kaur, "Key Frame Extraction from a Video using Edge Change Ratio", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 5, Issue 5, May 2015, pp.1228-1233.

- [9] Z. Wu, P. Xu, "Shot Boundary Detection in Video Retrieval", Electronics Information and Emergency Communication (ICEIEC), pp.86-89. <https://doi.org/10.1109/ICEIEC.2013.6835460>
- [10] L. Hongqiao, W. Shengqian, "A New Image Denoising Method Using Wavelet Transform", Information Technology and Applications, 2009, pp.111-114. <https://doi.org/10.1109/IFITA.2009.47>
- [11] Do, M. N. and Vetterli, M., "The Contourlet Transform: an efficient directional multiresolution image representation", IEEE Transaction on Image Processing 14(12), 2005, pp.2091-2106. <https://doi.org/10.1109/TIP.2005.859376>
- [12] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classifying scene breaks", Proc. ACM Multimedia '95, San Francisco, CA, pp. 189-200, Nov. 1995. <https://doi.org/10.1145/217279.215266>