Hybrid Method of Video Shot Segmentation Based on YCbCr Space Color Model

Sasmita Kumari Nayak[1*] and Jharna Majumdar[2]

[1]*Department of Computer Science and Engineering, Centurion University of Technology and Management, Odisha, India.*
[2]*Centurion University of Technology and Management, Odisha, India.*

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

Original Research Article

ABSTRACT

In this digital world, Video analysis is the most important and useful task. Currently, tremendous tasks have been done in video analysis like compressing the videos, video retrieval process and video database indexing, etc. For all these tasks, one common step is segmenting the video shots, which are referred to as Video Shots Segmentation (VSS). Video shots segmentation is used to segment the input videos into a number of frames sequentially where the scene changes occurred, i.e. called shots. In this article, segmenting the video shots follows a hybrid procedure. Here, we have introduced the moments of colors, distance metrics and threshold techniques. All the videos follow the above mentioned steps for segmenting the video shots. But, before that, the input video is converted into a specific color model i.e. YCbCr. Then, apply the color moments to extract the feature vectors of frames, which are differentiated based on the color features of frames. In every two frames of the video, distance metrics methods are applying to compute the similarity and dissimilarity of frames. And the dissimilarity of the frames can be computed by using the threshold technique to get the shots from the video. In this paper, we are using the adaptive threshold technique to segment the videos into various shots. In this step, we will get a true number of shots. By the experimental results, this proposed methodology can be evaluated with the sequence of videos based on the performance or evaluation metrics.

_____

Corresponding author: Email: nayaksasmita484@gmail.com;

1. INTRODUCTION

Video Shots Segmentation (VSS) is used to segment the input videos into number of frames sequentially where the scene changes occurred, i.e. called as shots. Generally, the shot will be used as an important unit, to analyze the videos, for compressing the videos, video retrieval process and video database indexing, and video summarization etc. [1], [2].

The key procedure of VSS is: i) extract one or more features from each frame of the video, ii) for subsequent frames, computing the feature differences, iii) compare these differences with the provided threshold value, and iv) finally, detect the shot boundary, when each time the threshold value is exceeded. Many methods are different depending on the use of feature types. The basic idea of all existing video segmentation as well as abstraction methods is the boundary detection of video shots. This procedure is the combination of color moments, distance metrics as well as the adaptive threshold technique. Hence it is treated as a hybrid approach of segmenting the shots from the input video.

Normally, a shot includes camera motions like zooms, tilts, or pans, and video modifications like wipes, dissolves, and fades [3]. Specifically, video shot transitions are categorized into two shot transitions. They are abrupt shot transitions and gradual shot transitions. The abrupt shot transition is also called as sharp shot transition or cuts. This cuts occurred when there is a sudden change from one frame to another. Similarly, the gradual shot transitions occurred after video editing effect. Depending on these effects, the gradual shot is categorized into two various types like, wipe, fade and cross dissolves.

The color space illustration has competed a significant role in committal to writing, compression, transmission, pattern recognition and digital transmission applications [4]. The RGB color space illustration has the foremost correlative elements, whereas the YCbCr color elements are the smallest amount correlated elements. The forward and backward transforms between RGB and YCbCr color areas are linear. The correlated RGB elements do not seem to be appropriate to embed the watermark. In RGB color area the perceived color quality of a picture depends on all elements. Thus, embedding watermark bits into one element independently of the opposite RGB elements is not the most effective alternative. On the opposite hand the YCbCr permits to extract uncorrelated elements and it favor the separation of the achromatic half from the chromatic elements of the colour image. To realize high robustness and enormous embedding capability, the projected scheme uses the smallest amount correlative YCbCr elements of the colour image. The colour image is pictured by Y, Cb and Cr elements. From these three parts, the modification within the intensity of chrominance parts is that the most sensitive to human eyes whereas for brightness level parts are least sensitive [4], [5].

In this work, we have used high order color moments for VSS [6]. In this article, a hybrid methodology is used to segment the video shots. This hybrid technique is the combination of three techniques like, color moments, distance metrics and adaptive threshold. Initially, the proposed technique uses three moments like, the mean, standard deviation, skewness, [7], at different layers of YCbCr color model in a block of MXM size for frames in video. Further, the outcomes of color moments are getting in the form of feature vector, which is utilized to get the similarity between subsequent frames [8]. Then, use 5 different distance metrics to construct continuity signal. The method of distance metrics are Euclidean, Chi-Square, Jeffrey Divergence, Canberra, and Hellinger distance metrics. Two different adaptive methods of thresholding are used, namely - Global adaptive and Local adaptive thresholds [9], [10]. The temporal position in the frame index where continuity measures cross, beyond a predefined threshold value, which shows a shot boundary.

The structure of this article is ordered in the following sections. The works, which are reacted to this paper has discussed in section 2. The explanation of proposed algorithm has given in section 3. After that the details of experimental outcomes are presented in section 4 and at the end, the paper has summarized in section 5.

2. LITERATURE SURVEY

The main focus of this section is to discuss the previous works did in the field of Video Shots Segmentation (VSS). Sometimes the VSS is also called as Shot Boundary Detection (SBD). In

paper [11], [12], authors discuss the varied shot boundaries techniques employed so far. The significant part of this paper is highlighting certain issues that require accurate and efficient SBD technique to solve varied video application problems.

The SBD is categorized into content based technique and temporal based technique. The content based technique detects the shot boundary by considering the contents present in the frame whereas the temporal based technique measures the differences in video frames based on time [13].

Similarly, a method proposed in [6], [14], utilized to estimate the matching difference between two subsequent frames of video with different weights. Here, automatic threshold technique is utilized to detect the shot boundaries. At the end, extracted the key frames form the videos by using reference frame-based method.

Huang, et al. [3], [15] proposed an automatic shot detection method that performed ineffectively because of detection of high false rates by reason of camera or object motion. To overcome this issue, a local key point matching of video frames is applied to detect the gradual changes as well as sudden changes proficiently.

Chakraborty, et al. [11], proposed varied algorithms, are clearly related to temporal video segmentation that is classified into compressed as well as uncompressed video domains [11], [16], [17]. Including this, for the same domain, some more algorithms have been proposed like, Histogram based, Block based and Pixel based comparisons [13], [18-21]. Whereas for content based video segmentation shape features [13], [20], [22], [23], texture [13], [20], [22-24] and color [13], [25-28] approaches are used.

Authors in [1], [29] exploit the color histogram features approaches like Hue Saturation Value (HSV) and Histogram of Gradient (HOG) for detecting the abrupt transition. This HSV color histogram and HOG feature approaches are utilized as the primary and secondary approach for the shot detection, which helps to increase the proficiency of these approaches. Then, use Chi square measurement method to get the difference of two adjacent frames that compared to the adaptive global threshold.

2.1 Description of Proposed Algorithm

This work is confined to YCbCr space color, which is represented in three color values Y, Cb

and Cr. Y, Cb and Cr represents the Luminance component, blue-difference and red-difference chroma components respectively. Each color channels of a frame computing the moments. Hence, each video frames are considered as 9 moments i.e. 3 moments for every 3 color channel. These three color moments are mean, standard deviation and skewness have defined in the next to this.

Mean is the average color value of the frame. The square root of the variance of the distribution is the standard deviation. Skewness is a measure of the degree of asymmetry in the distribution. All these three color moments i.e. mean, standard deviation and skewness are represented in equation (1), (2) and (3) respectively. These three color moments are defined in pixel form. Each pixel of the frames is represented as X_{ij}, which is the pixel of ith color channel ($i = 1, 2$ and 3 represents channel Y, Cb and Cr channel respectively) of jth frame.

$$Z_i = \frac{1}{M} \sum_{M}^{j=1} X_{ij} \tag{1}$$

$$SD_i = \sqrt{\left(\frac{1}{M} \sum_{M}^{j=1} \left(X_{ij} - Z_i \right)^2 \right)} \tag{2}$$

$$Sk_i = \sqrt[3]{\left(\frac{1}{M} \sum_{M}^{j=1} \left(X_{ij} - Z_i \right)^3 \right)} \tag{3}$$

Where, Z_i, SD_i and Sk_i are the Mean, standard deviation and skewness of ith color channel respectively.

Compute the sum of the weighted differences between the moments of the two distributions by using equation (4), which is used to get the dissimilarity between color distributions of two subsequent frames of a video. Then, a thresholding method is used to mark shot changes.

$$diss_{momnt}(H, I) = \sum \left(Z_i^1 - Z_i^2 \right) + \left(SD_i^1 - SD_i^2 \right) + \left(Sk_i^1 - Sk_i^2 \right) \tag{4}$$

Where,

$diss_{momnt}(H, I)$ =Comparison of two image distributions

$Z_i^1 Z_i^2$: mean (first moments) of the two image distributions

$SD_i^1 SD_i^2$: Standard deviation (second moments) of the two image distributions

$Sk_i^1 Sk_i^2$: Skewness (third moments) of the two image distributions

2.2 Proposed Algorithm

Step1: Convert each frame of the video sequence from RGB to YCbCr format.

Step2: Partition the frame into sub-blocks. Depending on the size of the frame, an optimal number of sub-blocks will be chosen. In this work, 8×8 is the optimal choice.

Step 3: Scan and calculate the three measures (Mean, Standard deviation and Skewness) for each sub-block, which is in Y, Cb, and Cr planes. Each block gives a feature vector, which is represented in equation (5).

$$Fv_m = \begin{bmatrix} Z_{mY} & SD_{mY} & Sk_{mY} & Z_{mCb} & SD_{mCb} & Sk_{mCb} & Z_{mCr} & SD_{mCr} & Sk_{mCr} \end{bmatrix} \tag{5}$$

Where,

$Z_{mY} \; SD_{mY} \; Sk_{mY}$ = mean, standard deviation and skewness values of Y color channel

$Z_{mCb} \; SD_{mCb} \; Sk_{mCb}$ = mean, standard deviation and skewness values of Cb color channel

$Z_{mCr} \; SD_{mCr} \; Sk_{mCr}$ = mean, standard deviation and skewness values of Cr color channel

Step 4: After calculating values for all frames the dissimilarity between frame F_i and F_{i+1} is calculated by a distance metric between each feature vector. Here, F_i and F_{i+1} are the feature vectors of frames in sub block wise moments of Y, Cb and Cr color channels.

In our implementation we have used Euclidean distance, Jeffrey divergence, Hellinger Distance, Chi-Square distance and Canberra Distance. The details of each distance metrics are as given below.

Euclidean Distance

$$d(f_i, f_{i+1}) = \sqrt{\sum \left(F_i(j,k) - F_{i+1}(j,k) \right)^2} \tag{6}$$

Canberra Distance:

$$d(f_i, f_{i+1}) = \sum \frac{\left| F_i(j,k) - F_{i+1}(j,k) \right|}{\left| F_i(j,k) \right| + \left| F_{i+1}(j,k) \right|} \tag{7}$$

Chi-Squared Distance:

$$d(f_i, f_{i+1}) = \sum \frac{(F_i(j,k) - F_{i+1}(j,k))^2}{F_i(j,k)} \tag{8}$$

Jeffrey Divergence:

$$d(f_i, f_{i+1}) = \sum \left(F_i(j,k) \log \frac{F_i(j,k)}{A} + F_{i+1}(j,k) \log \frac{F_{i+1}(j,k)}{A} \right) \tag{9}$$

Where, $A = \dfrac{F_i(j,k) + F_{i+1}(j,k)}{2}$

Hellinger Distance:

$$d(f_i, f_{i+1}) = \frac{1}{\sqrt{2}} \sqrt{\sum \left(\sqrt{F_i(j,k)} - \sqrt{F_{i+1}(j,k)} \right)^2} \qquad (10)$$

Step 5: Thresholding: An adaptive threshold is selected and dissimilarity value greater than threshold T indicates shot boundary. In our Implementation we have used 2 different Adaptive Thresholds the details of which are given below:

Global Adaptive Threshold:

$$T = \mu + (\alpha * \sigma) \qquad (11)$$

Where, µ is the mean, σ is the standard deviation of dissimilarity values, α is heuristically determined to be 3.

Local Adaptive Threshold: This is more reasonable than global threshold. A local window based threshold calculation method is verified.

$$T = \alpha \times (\mu \ of \ the \ window \ size \ m) \qquad (12)$$

Value of m is decided as 50, and α as 5, heuristically.

2.3 Work Flow of VSS

The work flow of VSS is shown in Fig. 1. This work flow follows the steps explained in the previous section i.e. in algorithm section.

3. RESULTS AND ANALYSIS

In this part, we have discussed the outcomes by analyzing the input videos with the help of our proposed methodology. Initially, the input video is loading and then run the videos for VSS i.e. for shot detection as shown in Fig. 2. After that, The

Frame Numbers of detected shots is being displayed on completion of video play, or on clicking the "pause" button as shown in Fig. 3.

Fig. 4 shows the navigation between the detected shots, by clicking the "SHOW SHOTS" button, and then "PREV" / "NEXT" button to fetch and display the frame before and frame after the detection.

The efficiency of the proposed algorithm is measured based on the recall and precision value. Recall is computed by taking the ratio of a number of detected cuts to the sum of undetected and detected cuts. Whereas the precision is measured by taking the ratio of a number of detected cuts to the sum of falsely detected and correctly detected cuts of video data.

$$\mathrm{Re}call = \frac{No.\ of\ correctly\ \det ected\ boundaries}{No.\ of\ true\ boundaries} \qquad (13)$$

$$\Pr ecision = \frac{No.\ of\ correctly\ \det ected\ boundaries}{No.\ of\ totally\ \det ected\ boundaries} \qquad (14)$$

In our implementation 5 different videos, with a variety of shot cuts were chosen and used for calculating the performance metrics. The details of videos used are given in Table 1 as shown in below:

Table 2 shows the performance metrics calculated for the videos using different Distance Measures and Thresholding Methods.

Table 1. Details of input videos

Video Name	Total Frames	Total Length	Dimensions	True Shots
Cricket 1	2251	1:30	320*240	35
Football 1	2760	1:50	320*240	26
Football 2	2989	1:59	320*240	41
Formula 1 2012 Australian Grandprix	5821	3:14	320*240	44
Movie 1	1294	0:43	640*480	05

Table 2. Performance of the Proposed Methodology Using Recall as well as Precision

Methods		Cricket 1		Football 1		Football 2		Formula 1		Movie 1	
Distance Measure	Thresholding	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
Euclidean	Global Adaptive	00.60	00.88	00.70	00.75	00.76	00.84	00.89	00.81	00.80	00.80
Euclidean	Local Adaptive	00.80	00.97	00.77	00.83	00.80	00.90	00.91	00.87	01.00	00.71
Canberra	Global Adaptive	00.66	00.96	00.81	00.91	00.93	00.88	00.93	00.79	00.80	00.80
Canberra	Local Adaptive	00.77	00.96	00.73	00.83	00.30	00.86	00.81	00.97	01.00	00.63
CHI-Squared	Global Adaptive	00.51	01.00	00.73	00.76	00.78	00.86	00.86	00.84	00.60	01.00
CHI-squared	Local Adaptive	01.00	00.83	00.85	00.70	00.78	00.71	00.95	00.72	01.00	00.42
Jeffrey Divergence	Global Adaptive	00.51	00.90	00.73	00.80	00.88	00.90	00.89	00.83	00.60	01.00
Jeffrey Divergence	Local Adaptive	01.00	00.80	00.81	00.64	00.78	00.73	00.95	00.70	01.00	00.40
Hellinger	Global Adaptive	00.74	00.93	00.73	00.73	00.88	00.82	00.91	00.78	00.80	00.67
Hellinger	Local Adaptive	00.80	00.97	00.77	00.87	00.85	00.93	00.86	01.00	01.00	00.56

INPUT VIDEO
Frames (0,1,2,...N)

Frame N-1

Frame N

Convert RGB to YCbCr

Convert RGB to YCbCr

Segment each frame into 8 X 8 blocks. i.e. 64 blocks

Segment each frame into 8 X 8 blocks. i.e. 64 blocks

For each Block and Each color (Y, Cb, Cr) calculate Mean, Standard Deviation, Skewness.
i.e. 64 blocks X 3 colors X 3 features = 576 values.

Each block is represented by feature vector with 9 elements. [Ymean, Ysd, YSkw, Cbmean, Cbsd, CbSkw, Crmean, Crsd, Crskw]

For each Block and Each color (Y, Cb, Cr) calculate Mean, Standard Deviation, Skewness.
i.e. 64 blocks X 3 colors X 3 features = 576 values.

Each block is represented by feature vector with 9 elements. [Ymean, Ysd, YSkw, Cbmean, Cbsd, CbSkw, Crmean, Crsd, Crskw]

Calculate dissimilarity DismN between frame N-1 and frame N between feature vectors of corresponding Blocks.

THRESHOLDING
After calculating Dism for each continuous pair of frames, adaptively set Threshold using either:

$$T = \mu + (3 * \sigma)$$

ORc
For every 50 frames   T= 5× (μ of 50 frames)

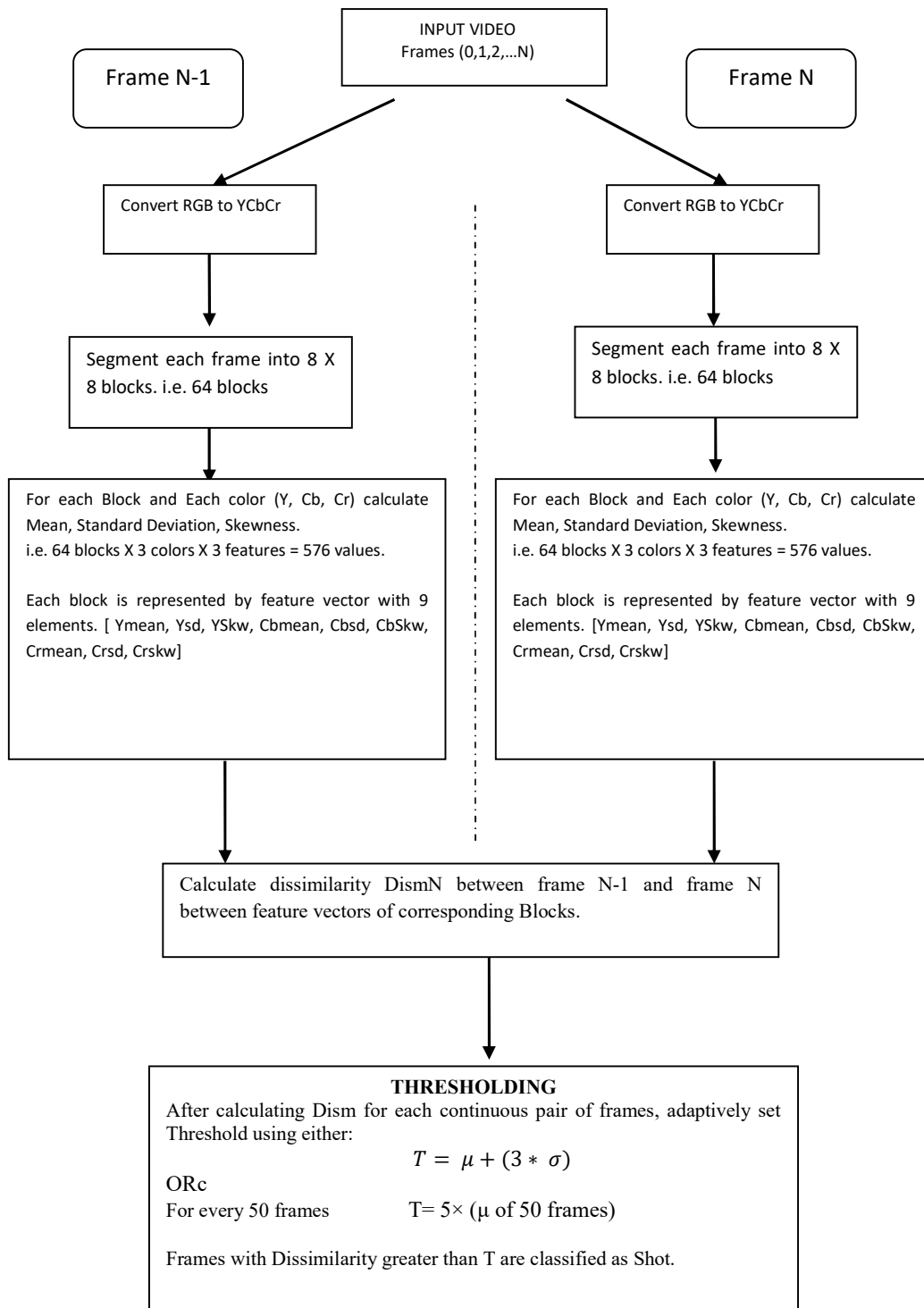Frames with Dissimilarity greater than T are classified as Shot.

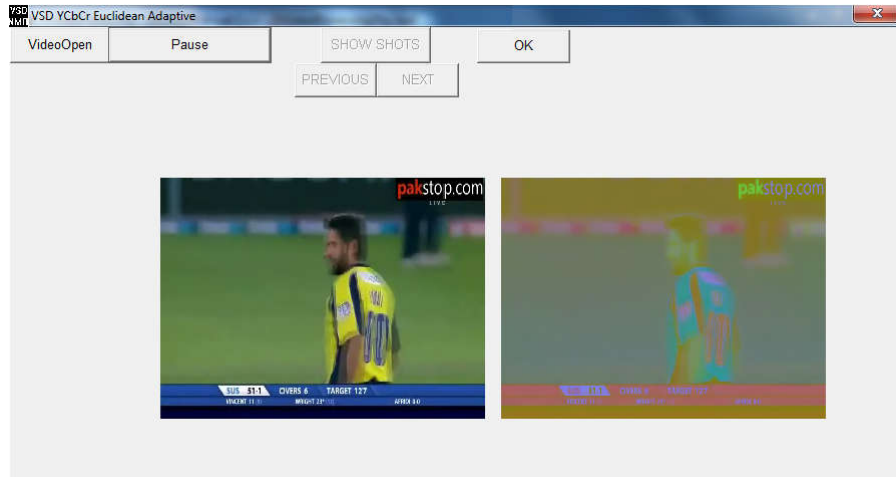**Fig. 1. A Graphic representation of work flow of the methods**

Fig. 2. The Interface for loading and running videos for shot detection
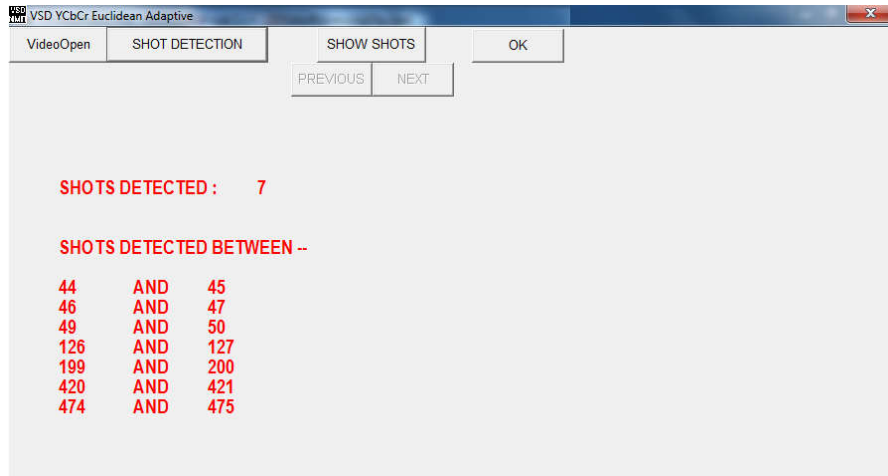

Fig. 3. Shots detection with the frame numbers

Fig. 4. Detection of shots with frames of before detection and after detection

4. CONCLUSION

The initial step of most of the video applications is video shot segmentation. In this process, find the changes in the scenes from the input videos. In this article, the proposed methodology finds the segementation of shots efficiently, which solves most of the video applications like video summarization, content indexing etc. Here, we have chosen the YCbCr color space model instead of RGB. After that applied the three color moments with the distance metrics to get the similairty and dissimilarity of frames to get the shot segments, after comparing the threshold value. Here we have applied the adative thresholding technique to get the segmented shot efficiently. From the analysis and performance table, we observe that though Chi-squared and Jeffrey distances with local adaptive thresholds perform well in some cases, Euclidean Distance with local adaptive threshold is verified to be more consistent and robust. In future, we will use a novel technique to get the video shot segmentation with the new video application.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCE

1. Hato E, Abdulmunem ME. Fast Algorithm for Video Shot Boundary Detection Using SURF features. In 2019 2nd Scientific Conference of Computer Sciences (SCCS). 2019;81-86. IEEE.
2. Singh RD, Aggarwal N. "Novel research in the field of shot boundary detection–a survey", in Advances in Intelligent Informatics, E.S. El-Alfy, S. Thampi, H. Takagi, S. Piramuthu, T. Hanne, Eds. Springer, Cham. 2015;457- 469.
3. Pal, G, Rudrapaul D, Acharjee S, Ray R, Chakraborty S, Dey N. Video shot boundary detection: a review. In Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India CSI Springer, Cham. 2015;2:119-127.
4. Jaba M, Elsghair M, Tanish N, Aburgiga A. Applying Edge Information in YCBCR Color Space on the Image Watermarking. Signal & image processing: An International Journal (sipij). 2017;8:2.
5. Kutter M, Jordan F, Bossen F. Digital signature of color images using amplitude modulation", in Proc. SPZE Int. Conf: Storage and Retrieval for Zmage and Video Database. 1997;3022:518-526.
6. Jadhav MPS, Jadhav DS. Video summarization using higher order color moments (vsuhcm). Procedia Computer Science. 2015;45:275-281.
7. Singh SM, Hemachandran K. Image retrieval based on the combination of color histogram and color moment. International Journal of Computer Applications. 2012;58(3).
8. Bekhet, S., Ahmed, A. Evaluation of similarity measures for video retrieval. Multimed Tools Appl. 2020;79:6265–6278. Available:https://doi.org/10.1007/s11042-019-08539-4
9. Senthilkumaran N, Vaithegi S. Image segmentation by using thresholding techniques for medical images. Computer Science & Engineering: An International Journal. 2016;6(1):1-13.
10. Davies ER. Chapter 4 - The role of thresholding. In E.R. Davies (Ed.). Computer Vision Fifth Edition, Academic Press. 2018;93-118.
11. Chakrabort S, Thounaojam DM, Sinha N. A Shot boundary Detection Technique based on Visual Colour Information. Multimedia Tools and Applications. 2021;80(3):4007-4022.
12. Abdulhussain Sadiq H, Ramli Abd Rahman Saripan MI, Mahmmod BM, Al-Haddad S AR, Jassim WA. Methods and challenges in shot boundary detection: a review. Entropy. 2018;20(4):1–42.
13. Guru DS, Suhil M, Lolika P. A novel approach for shot boundary detection in videos. In Multimedia Processing, Communication and Computing Applications. Springer, New Delhi. 2013; 209-220.
14. Zhao Guang-sheng. A Novel Approach for Shot Boundary Detection and Key Frames Extraction', IEEE; 2008.
15. Huang CR, Lee HP, Chen CS. Shot Change Detection via Local Key point Matching. IEEE Transactions on Multimedia. 2008;10(6):1097–1108.
16. Koprinska I, Carrato S. Temporal video segmentation: A survey'. Signal Processing:, Image Communication. 2001; 16(5):477–500.
17. Thounaojam DM, Trivedi A, Manglem-Singh KH, Roy S. A survey on video

segmentation Proc. Int. Conf. Advanced computing, networking, and informatics India. 2013;903–912.

18. Dengsheng Zhang and Guojun Lu, Segmentation of Moving Objects in Image Sequence: A Review. 2001;2(2):143-183.

19. Irena Koprinska and Sergio Carrato, Temporal Video Segmentation: A Survey Signal Processing. IEEE transactions. 2001;16(5):413-506.

20. Idris F, Panchanathan S. Review of Image and Video Indexing Techniques. Journal of visual communication and image representation. 1997;8(2):146–166.

21. Manjunath S, Guru DS, Suraj MG, Harish BS. A Non Parametric Shot Boundary Detection: An Eigen Gap based Approach, proceedings of fourth annual ACM Bangalore conference. 2011;1:1030-1036.

22. Hualu Wang, Ajay Divakaran, Anthony Vetro, Shih-Fu Chang and Huifang Sun, Survey of Compressed-Domain Features Used in Audio-Visual Indexing and Analysis, Jornal of Visual Communication and Image Representation. 2003;14: 150-183.

23. Patel BV, Meshram BB. Shah and Anchor Kutchhi Polytechnic, Content Based Video Retrieval Systems, International Journal of UbiComp (IJU). 2012;3(2).

24. Ying Liua,, Dengsheng Zhanga, Guojun Lua and Wei-Ying Mab, A Survey Of Content-Based Image Retrieval With High-Level Semantics, Pattern Recognition. 2007;40:262 – 282.

25. Yasira Beevi C P and Dr. S. Natarajan, An Efficient Video Segmentation Algorithm with Real Time Adaptive Threshold Technique. International Journal of Signal Processing, Image Processing and Pattern Recognition. 2009;2:4.

26. Reddy PVN, Satya prasad K. Color and Texture Features for Content Based Image Retrieval. 2011;2(4):1016-1020.

27. Quynh Nguyen Huu, Ha Nguyen Thi Thu, Tao Ngo Quoc, An efficient Content based Image Retrieval Method for Retrieving Images International Journal of Innovative Computing, Information and Control. 2012;8:4.

28. Yi Tao, William, Object-Based Image Retrieval Using Point Feature Maps. 1997; 40:12.

29. Shao H, Qu Y, Cui W. "Shot boundary detection algorithm based on HSV histogram and HOG feature", In Proceedings of 5th International Conference on Advanced Engineering Materials and Technology. 2015;951-957.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle4.com/review-history/71178