SEMARAK ILMU
PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

SEMARAK ILMU
Journal of
ADVANCED RESEARCH IN APPLIED SCIENCES AND ENGINEERING TECHNOLOGY
ISSN 2462 - 1943

Video Structure Extraction using Shot Boundary Detection for Forgery Detection

Noraida Haji Ali[1,*], Fadilah Harun[1]

[1] Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia

ARTICLE INFO

Article history:
Received 9 May 2023
Received in revised form 10 July 2023
Accepted 18 November 2023
Available online 7 December 2023

Keywords:
Video extraction; shot boundary detection; video authentication detection

ABSTRACT

In the digital era, video and pictures are perceived as more reliable than words. Considering the popularity of the internet, forging digital videos is probably nothing new. The recording, archiving, and sharing images and videos on media platforms are frequently done with mobile phones and digital cameras. Additionally, anyone may easily adjust or edit an image because of the availability of software editors. Technology and technological advancements have made video authentication detection a prevalent issue. As a result, digital recordings are being used as visual evidence in many susceptible domains, including media, politics, civil or criminal trials, and military and intelligence activities. Digital videos are an essential source of information with a high degree of evidence, but they may also be easily manipulated on purpose. As a result, in situations where reliance on video evidence is required, it is crucial to authenticate the contents of the video evidence before accepting it as an accurate picture of reality. This research focuses on several strategies for detecting shot boundaries to determine whether video content has been altered or tampered with. Although shot boundary detection is difficult, it is commonly used in structured video because these techniques can detect changes brought on by a regular shot change rate. In the meantime, this research will provide a novel shot boundary detection-based video authentication algorithm essential for developing practical tools to evaluate the video's authenticity based on its fundamental structure.

1. Introduction

Digital video is a well-organized collection of photos taken using a digital camera. Additionally, there is audio and other additional data. Research conducted by Akhtar *et al.,* [1] showed people are increasingly dependent on multimedia content daily, especially digital videos. In study by Shelke and Kasana [2] explain that the surveillance camera is another treasure of modern technology widely utilized in homes, offices, and other public spaces as a reliable security precaution. This is because the video footage is evidence in most countries against such crimes. In addition, because of the easy access to advanced editing software and the use of the latest smartphone, it is easy for anyone to perform manipulations in digital video and tamper with its based stated by Girish and Nandini [3].

[*] *Corresponding author.*
E-mail address: aida@umt.edu.my (Noraida Haji Ali)

https://doi.org/10.37934/araset.34.2.187194

The intentional modification in the digital video for falsification is called video forgery. It may be hard for humans to decide the authenticity of those digital videos with the naked eye. Because of this, it is crucial to evaluate and determine whether the video content is authentic or has been changed before using it as evidence in court as explained by Wahab *et al.,* [4]. Therefore, digital false detection methods are required to confirm the reliability and validity of digital videos. However, it is challenging and complex to maintain the integrity of video data and verify its authenticity and its clarified by Upadhyay and Singh [5]. The goal is to guarantee that the video data and structure offered are authentic and exact reproductions of those recorded. Numerous reports of unlawful activity connected to the video have surfaced recently. For instance, video editing software damages the original video's structure and content. However, in some situations, such as forensic and police investigations, legal proceedings, and patent right of possession, the video's veracity is essential. For instance, Birajdar and Mankar [6] ensuring that video content is reliable, hasn't been changed, and comes from a legitimate court source. As a result, was conducted by Girish and Nandini [3], video authentication is a process that checks to see if the data and fixed structure of the tested video are genuine, free of interference, and identical to what is needed. Video is challenging and has several related issues because of the enormous number of frames that must be handled as asserted in Wahab *et al.,* [4]. There are several methods and strategies for identifying video content and its structure in different fields, such as judicial investigations, law enforcement, or evidence in a court of law, according to past studies. The techniques used include digital signatures and watermarks was conducted in Upadhyay and Singh [5], but this is only possible if embedded pre-processing is not done. Our research focuses on the identification of shot boundaries for video authentication in order to retrieve the video structure.

Recognizing the edges of a video shot is the first and most important step in indexing, browsing, and retrieving video content introduced by Verma and Raman [7]. According to the research, many different algorithms have been used to make software that can find video shooting restrictions. Akhtar *et al.,* [1] stated in existing applications, for instance, use shot boundary recognition methods, shot change detection algorithms and features to represent video frames. Colour, movement, histograms, the boundaries of the image, the rectangular block of a structure, the entire frame, a single pixel, etc., are some properties utilized in video representation explained by Rathod and Nikam [8]. In Tekalp [9], shot boundary identification requires decomposing the fundamental structure into its essential pieces and identifying the connections between its units. For instance, a shot boundary detection was made when the feature contrast displayed an abrupt transformation more substantial than the threshold. The visual system's foundational elements, however, are inconsistent and unstable. To better refine the content data, shot boundary detection techniques can help us discover the manipulation structure of a video. This research aims to develop a new algorithm model for video structure detection using practical shot boundary detection approaches to build useful tools for recognizing the video's authenticity and detecting irregularities due to modifications. Shot boundary detection is also ideal for extracting the entire video structure and ensuring that the content and video data obtained are correct and persistent, according to a study of the approaches collected by the proposed method Boreczky and Rowe [10].

2. Research Background

Because video data are such rich sources of information, they must be analyzed before modelling. As previously established, there are two steps to video analysis. In the first stage, the video sequence is divided into a collection of shots (shot boundary detection), and in the second stage, a frame is chosen to represent each shot. In general, the literature on segmenting video data shows two themes. The first executes in the uncompressed domain, while the second operates in the compressed domain. Several researchers have conducted a few earlier studies on video data. One of Queluz [11] earlier studies revealed the following:

(i) There is a lack of security protection for any video data.
(ii) A simple way to change the content and meaning of a video is by deleting, rotating, adding, or rearranging it using free software for image editing.

This study focuses on the issues expressed about modifying the content of videos and ensuring their authenticity based on their structure. Authentication verifies that the content is original and unaltered at its core. Given how simple it is to edit and manipulate video footage, the ability to spot changes in digital video is essential. Numerous authentication detection methods have recently been published in the literature, although most concentrate on images. Attacks on the video sequence, attacks on the video's content, or attacks on the temporal connections between frames are all indicators of video manipulation was explained by Sawant and Sabnis [12]. Attacks that tamper with video can change an object in various ways. Figure 1 depicts a typical instance of video editing in which an object is added in the next frame. In addition, existing things might also have their size, colour, and shape changed with bad intentions. The modification could alter the video's authenticity and meaning. Recent years have seen a lot of research activity in video authentication stated by Sawant and Sabnis [12]. Shot boundary detection methods can identified video structures like scenes, shots, and keyframes was conducted in Smeaton *et al.,* [13]. A hierarchical video representation by the employed shot boundary detection is suggested to visualize the entire structure. The use of shot boundary detection in video content architecture allows for efficient video abstraction and high-level video segmentation techniques mentioned in Browne *et al.,* [14].

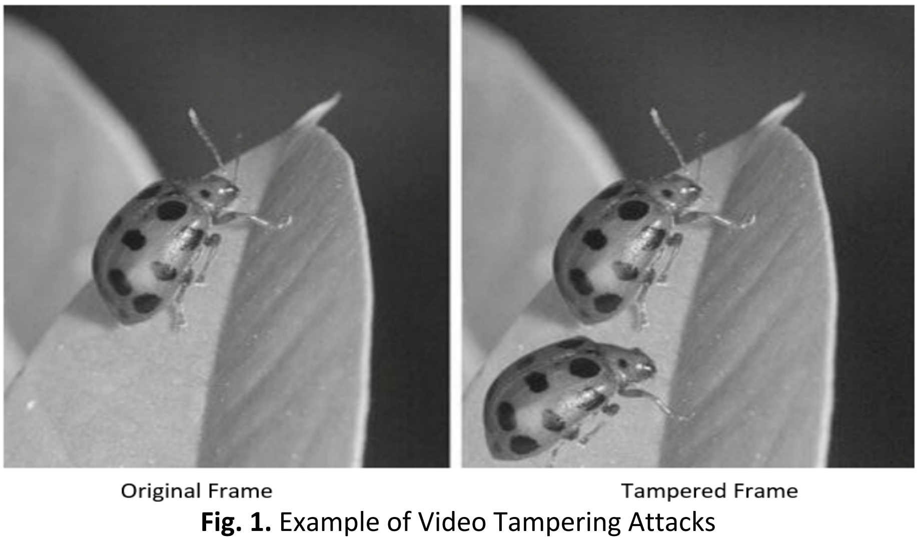


Fig. 1. Example of Video Tampering Attacks

Video authentication ensures that the material in a particular video is authentic and precisely the same as when it was recorded. Girish and Nandini [3] clarified two methods for detecting video authentication are active and passive. An active approach can be used if pre-processing insertions like watermarking and digital signatures are added to the video. The watermarking and signatures technique could restore the video if altered abnormally. In the meantime, a passive approach involves identifying changes by taking features from video footage. Recent times have seen the rapid evolution of a revolutionary technology for swiftly verifying the contents of digital images that do not require prior knowledge of the image and are hence referred to as passive. The technology, known as multimedia forensics mentioned by Gupta *et al.,* [15], is based on the observation that every step of the image history, from the process of acquiring the image to its storage in a compressed format to any post-processing operation, leaves a distinctive trace on the data, acting as a kind of digital fingerprint. By identifying the existence, lack, or incongruence of such elements organically linked to the digital content, it is, therefore, feasible to identify the source of the digital image or decide whether it is authentic or modified. Multimedia forensics is a branch of traditional forensic science that examines the application of rational thought to extract factual information from physical or digital evidence. By utilizing the already-existing expertise in digital photography and multimedia security research, multimedia forensic tools are responsible for exposing the traces in multimedia content at each stage of its life. Additionally, this study focuses on shot boundary detection to passively detect changes made to video data and confirm video content's originality.

Research from the previous Naturel and Gros [16], has shown that television streams are highly redundant and that it is necessary to find repetitions in the underlying structure of the video. However, retrieving the entire video programme with the precise bounds limit, from the first to the final frame, is still challenging. Shot matching was employed to locate the repeated shots and frames, but the entire video structuring was still not accomplished. Research by Boreczky and Rowe [10] analyzed the efficiency of several shot boundary detection methods and a debate that established and outlined several different methodologies for video shot boundary identification. The methods employed for shot boundary identification include motion vectors, edge differences, compression differences, statistical differences, histogram comparisons, and pixel differences. Each technique was unique in its benefits and drawbacks. Because it uses straightforward histogram methods to compute the two images' colour or gray level histograms, histogram comparison is more effective than other techniques. When the bin-wise difference between the two histograms exceeds a certain threshold, a shot boundary is recognized, and it is evident that the video has changed. This research by Ibrahim [17] employed the shot-matching method to identify repeated frames in a television stream. This shot-matching strategy is due to the structure and number of repetitions in television streams. However, retrieving only shots and frames with repetitions makes it harder to complete the entire framework. According to research by Zhou *et al.,* [18], video has two frequent problems: duration and an unstructured format, making it particularly challenging to access. The idea of a table of contents (TOC) was introduced for browsing and retrieving hierarchical structure, but it only returned the top three levels of structure, shots, groups, and scenes. However, by analyzing the video's visual content and dividing it into a fundamental unit of video called a frame, researchers can typically get a good result. The order of the frame numbers affects the frame order. A shot is a continuous section of a video frame sequence with either constant or static camera motion.

In most cases, combining a shot (scene) results in a single video clip. The detection of a scene's sequence of images (SBD) is based on the recognition of visual contrast brought on by transformations. SBD is also used to index and browse hierarchically organized videos. The problem of video authentication detection is widespread and crucial in many industries. Shot boundary detection is one method for dealing with video authentication, and it may assist in resolving the challenges of establishing and demonstrating the authenticity of video clips based on frame hierarchical structure.

3. Video Structure Detection

Similarity analysis is a valuable technique for finding problems with authentication since the original and duplicated sequences are similar. On the other hand, the earlier study was ineffective and time-consuming. In order to address this issue, this paper suggests a video structure extraction approach based on pixel similarity analysis. The algorithm will evaluate whether or not the video has been edited by looking for these video clips with many similarities, and it can easily locate source sequences and redundancy. The technique suggested in this paper is intended to improve video authentication's ability to identify forgery attempts using shot boundary detection. Shot boundary detection was conducted Lungisani *et al.,* [19] effectively identifies the relationship between the video structure's scene, shot, and frame components. The shot boundary may be noticed when the contrast between the character and their environment results in a net change greater than the threshold.

3.1 Shot Boundary Detection Algorithm

In general video processing, shot boundary detection is generally the initial step. A shot is a collection of frames taken from a single, continuous recording made by a camera. As a result, the same shot's subsequent frames display temporal continuity. Due to unique circumstances like flashlight events, quick lightning variations, fast camera motion, or huge object motions, the actual shot cut and the abrupt cut might significantly change the frame difference. As a result, each shot represents a single continuous motion, and the substance of each shot remains constant. Changes in content always take place where two shots meet. A video sequence can be divided into shots for video summary and indexing was clarified in Ko *et al.,* [20]. Shot boundary detection technique detects a shot cut if the scaled frame difference between two subsequent frames is more than a max-threshold (thmax), if the contiguous frame difference value is greater than a k-threshold (^{k}gloval), and if the Euclidian distance is greater than or equal to a global threshold (thgloval).

Following is a summary of the shot boundary detection algorithm;

(i) Step 1: Initially, the video is input into the extraction process.

(ii) Step 2: In order to identify each scene, the extraction process divides the collection of images into two blocks of two consecutive images. The remaining changes in each extracted image's pixel define the percentage of each image.

(iii) Step 3: To sum up based on Eq. (1), the difference between the corresponding pixels of two successive frames is calculated using pixel-based algorithms. A firing restriction must be assumed if a difference exceeds a specific threshold value.

$$FDk,r\ (x) = s(x,k) - s(x,r) \tag{1}$$

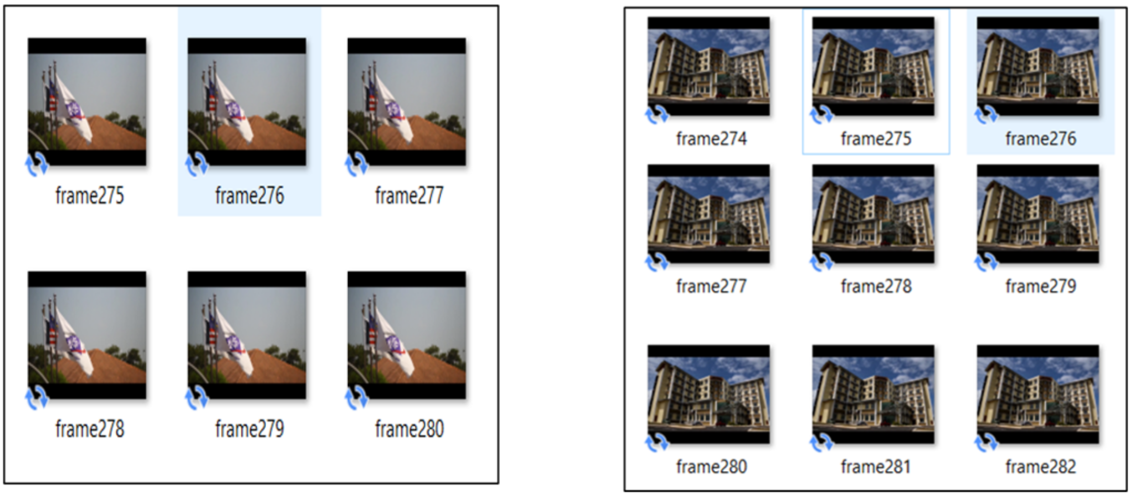
Where *s(x,k)* denotes the intensity per pixel for *x* inside the *k* frame and *x (x1, x2)* denotes the pixels' location. Between the current picture *(k)* and the reference image (*r*), the Image Difference displays the pixel-by-pixel difference. The reference image r can be considered a fixed-time image or the prior image *k - 1* (subsequent frame difference). When employing shot boundary detection for extraction, there are three steps involved. All of the frames on the video will first be read sequentially via the processing of the frames. Second, every frame sequence currently used will be converted to a block (shot). And lastly, a comparison between the corresponding shots to find the frame.

(iv) Step 4: Shot boundary detected.

4. Results and Discussion

To establish how to verify the video structure, the user must input two videos (original and altered) into the prototype. This application is crucial for spotting unlawful alterations to actual video footage. The quantity of the scene, shot, and frames are evidence of the extraction procedure. By comparing the number of existing structures, it may be understood. If there is a relationship between the structure of the video scene (*SC*), shot (*SH*), and frame (*F*), then the structure must be related to the occurrence of the function. Every structure should be connected to another structure initially, and the second structure should be related to a third structure via three joint list structures. The link between the scene, shot, and frames in the video are demonstrated. Relations composition describes the purpose-built into the video's structure. Each scene, shot, and frame may belong to one or more relationships.

Figure 2 below shows the different results for the same video input in the prototype. Video A showed the original location of frames after the extraction process, but Video B showed different findings with the different structures of the frame location. Based on this result, it can be stated that in Video B, an alteration (maybe a deletion or shuffle) occurs and doesn't have similar results to the original one.



Video A (Original) Video B (Edited)

Fig. 2. Results of Frame Extraction

As in Table 1, a video clips (original and edited) have been tested to prove the process of extracting video, and each video clip is about 10-15 minutes long. The prototype has been designed to extract the basic structure of a user-entered video. By comparing the number of frames and retrieved structures from the original video clip, it is possible to evaluate the authenticity of a video clip. The actual structure of the original video will be impacted by any alterations made to it. As a result, it can determine whether the video is real. Statistically comparing image differences in the pixel intensity domain is the most direct method of detecting temporal discontinuities. A shot may be reported if there are more noticeable changes between a certain number of pixels than a threshold value. By comparing the pixel width and height, Table 1 compares the variations in intensity or colour value of comparable pixels in two corresponding frames. It could signify the pixel's intensity level. According to the outcome, distinct pixels compare well to the original video when comparing pixels between frames 198 and 199. As a result, a modification process is shown to occur between frames 198 and 199. The same procedure was carried out on ten videos to discover the pattern of the original frame and edit the video.

Table 1
Example of Frame comparison between two videos

Original Video	Pixel Similarity	Edited Video	Pixel Similarity
Frame 197 to 198	100	Frame 197 to 198	100
Frame 198 to 199	100	Frame 198 to 199	99.89
Frame 199 to 200	100	Frame 199 to 200	100
Frame 200 to 201	100	Frame 200 to 201	99.73
Frame 201 to 202	100	Frame 201 to 202	100
Frame 202 to 203	100	Frame 202 to 203	99.81
Frame 315 to 316	100	Frame 315 to 316	99.18
Frame 316 to 317	100	Frame 316 to 317	100
Frame 317 to 318	100	Frame 317 to 318	100
Frame 318 to 319	100	Frame 318 to 319	100
Frame 319 to 320	100	Frame 319 to 320	99.91
Frame 320 to 321	100	Frame 320 to 321	99.97
Frame 321 to 322	100	Frame 321 to 322	100

5. Conclusion

The extraction approach separates an image's collection into two blocks followed by one another to identify each scene. The remaining changes in each extracted image's pixel are then used to establish each image's percentage. A detailed breakdown of scenes, shots and essential images was later produced due to the extraction of the video structure. The extraction process splits an image. The remaining changes in each extracted image's pixel are then used to establish each image's percentage. A detailed inventory of scenes, shots, and images was later produced as a consequence of extracting the video structure. Various approaches have been presented to find shot boundaries, including edge detection, motion compression, statistical analysis, histogram analysis, and pixel analysis. This study concentrated on finding pixel-based techniques that compare the relevant pixels of two subsequent frames. The assumption that a shot boundary exists and that there are differences between the two frames is made if the results between the two frames are higher than the threshold. Due to the threshold for detecting changes in video clips, the advantage of pixel-based shot boundary detection is that it is rapid and straightforward.

Additionally, the methods included in this study weren't efficient for videos with fast shifting. For instance, less well-known skills like object identification and analysis should be further studied to enhance methods. Hope that this study can improve video quality and assist with malicious video editing detection.

The most effective approach moving forward to resolve this issue is to implement passive video authentication detection approaches. A comparison to the original video is required to identify the authenticity of the present work. To demonstrate tampering, video forgery detection evaluates the originality of video evidence and divides it into active and passive approaches. Without additional information and hardware needs, a passive technique is independent of intrinsic information (pattern). The passive approach will produce accurate and detailed findings using organized video. The second goal of this study is to uncover and identify harmful assaults in videos. This research retrieved the material more effectively by employing a passive strategy. It is also demonstrated that the passive approach may be utilized to identify the originality of videos using no references by studying the strategies produced by the provided method.

Acknowledgement
This research was not funded by any grant.

References

[1] Akhtar, Naheed, Mubbashar Saddique, Khurshid Asghar, Usama Ijaz Bajwa, Muhammad Hussain, and Zulfiqar Habib. "Digital video tampering detection and localization: review, representations, challenges and algorithm." *Mathematics* 10, no. 2 (2022): 168. https://doi.org/10.3390/math10020168

[2] Shelke, Nitin Arvind, and Singara Singh Kasana. "A comprehensive survey on passive techniques for digital video forgery detection." *Multimedia Tools and Applications* 80 (2021): 6247-6310. https://doi.org/10.1007/s11042-020-09974-4

[3] Girish, N., and C. Nandini. "A review on digital video forgery detection techniques in cyber forensics." *Science, Technology and Development* 3, no. 6 (2019): 235-239.

[4] Wahab, Ainuddin Wahid Abdul, Mustapha Aminu Bagiwa, Mohd Yamani Idna Idris, Suleman Khan, Zaidi Razak, and Muhammad Rezal Kamel Ariffin. "Passive video forgery detection techniques: A survey." In *2014 10th International Conference on Information Assurance and Security*, pp. 29-34. IEEE, 2014. https://doi.org/10.1109/ISIAS.2014.7064616

[5] Upadhyay, Saurabh, and Sanjay Kumar Singh. "Video authentication: Issues and challenges." *International Journal of Computer Science Issues (IJCSI)* 9, no. 1 (2012): 409.

[6] Birajdar, Gajanan K., and Vijay H. Mankar. "Digital image forgery detection using passive techniques: A survey." *Digital Investigation* 10, no. 3 (2013): 226-245. https://doi.org/10.1016/j.diin.2013.04.007

[7] Verma, Manisha, and Balasubramanian Raman. "A hierarchical shot boundary detection algorithm using global and local features." In *Proceedings of International Conference on Computer Vision and Image Processing: CVIP 2016*, Volume 2, pp. 389-397. Springer Singapore, 2017. https://doi.org/10.1007/978-981-10-2107-7_35

[8] Rathod, Ganesh I., and Dipali A. Nikam. "An algorithm for shot boundary detection and key frame extraction using histogram difference." *International Journal of Emerging Technology and Advanced Engineering* 3, no. 8 (2013): 155-163.

[9] Tekalp, A. Murat. *Digital video processing*. Prentice Hall Press, 2015.

[10] Boreczky, John S., and Lawrence A. Rowe. "Comparison of video shot boundary detection techniques." *Journal of Electronic Imaging* 5, no. 2 (1996): 122-128. https://doi.org/10.1117/12.238675

[11] Queluz, M. Paula. "Authentication of digital images and video: Generic models and a new contribution." *Signal Processing: Image Communication* 16, no. 5 (2001): 461-475. https://doi.org/10.1016/S0923-5965(00)00010-2

[12] Sawant, Rohini, and Manoj Sabnis. "A review of video forgery and its detection." *Journal of Computer Engineering (IOSR-JCE)* 20, no. 2 (2018).

[13] Smeaton, Alan F., Paul Over, and Aiden R. Doherty. "Video shot boundary detection: Seven years of TRECVid activity." *Computer Vision and Image Understanding* 114, no. 4 (2010): 411-418. https://doi.org/10.1016/j.cviu.2009.03.011

[14] Browne, Paul, Alan F. Smeaton, Noel Murphy, Noel E. O'Connor, Seán Marlow, and Catherine Berrut. "Evaluating and combining digital video shot boundary detection algorithms." In *Proceedings of the 4th Irish Machine Vision and Information Processing Conference*, Queens University Belfast, 2000. 2000.

[15] Gupta, Ankita, Shilpi Gupta, and Anu Mehra. "Video authentication in digital forensic." In *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)*, pp. 659-663. IEEE, 2015. https://doi.org/10.1109/ABLAZE.2015.7154945

[16] Naturel, Xavier, and Patrick Gros. "Detecting repeats for video structuring." *Multimedia Tools and Applications* 38 (2008): 233-252. https://doi.org/10.1007/s11042-007-0180-1

[17] Ibrahim, Zein Al Abidin. "TV Stream table of content: a new level in the hierarchical video representation." *Journal of Computer Sciences and Applications* 7, no. 1 (2018): 1-9. https://doi.org/10.12691/jcsa-7-1-1

[18] Zhou, Xiang Sean, Yong Rui, Thomas S. Huang, Xiang Sean Zhou, Yong Rui, and Thomas S. Huang. "Constructing table-of-content for videos." *Exploration of Visual Data* (2003): 53-73. https://doi.org/10.1007/978-1-4615-0497-9_5

[19] Lungisani, Bose, Edwin Thuma, and Gabofetswe Malema. "A Classification Approach to Video Shot Boundary Detection." *International Journal of Signal Processing, Image Processing and Pattern Recognition* 10, no. 12 (2017): 103-118. https://doi.org/10.14257/ijsip.2017.10.12.08

[20] Ko, Kyong-Cheol, Young Min Cheon, Gye-Young Kim, Hyung -Il Choi, Seong-Yoon Shin, and Yang-Won Rhee. "Video shot boundary detection algorithm." In *Computer Vision, Graphics and Image Processing: 5th Indian Conference, ICVGIP 2006*, Madurai, India, December 13-16, 2006. Proceedings, pp. 388-396. Springer Berlin Heidelberg, 2006. https://doi.org/10.1007/11949619_35