Static video summarization approach using Binary Robust Invariant Scalable Keypoints

Eman Morad
Information technology department, Faculty of Computers and Information, Menoufia University
Menoufia, Egypt
eman.morad@ci.menofia.edu.eg

Khalid M. Amin
Information technology department, Faculty of Computers and Information, Menoufia University
Menoufia, Egypt
k.amin@ci.menofia.edu.eg

Sameh Zarif
Information technology department, Faculty of Computers and Information, Menoufia University
Menoufia, Egypt
sameh.shenoda @ci.menofia.edu.eg

***Abstract*— The constant demand and generation of digital video information have recently resulted in an increase in the growth of digital video content. Due to the rapid browsing of large amounts of data, content retrieval and indexing of video require an effective and advanced analysis technique. For quickly browsing, indexing, and accessing massive video archives, video summarizing approaches have been proposed. This research presents a new binary descriptor-based method for video summarization. The proposed method extracts key points and descriptors using a Binary Robust Invariant Scalable Key point (BRISK). For matching the binary descriptors between two successive frames, we employ a Brute-force method. And keyframes are extracted from each shot as the middle frame. Experiments were carried out using open video project data sets containing videos of various genres. The Comparison of user summaries (CUS) evaluation metric is used to assess the proposed method by calculating the accuracy and error rates and comparing it to other methods. As demonstrated by the experimental results, the proposed method gives good results when compared with other methods.**

***Keywords*—** *Video summarization, shot boundary detection, keyframe extraction, Binary Robust Invariant Scalable Keypoints (BRISK).*

I. Introduction

Due to advancements in multimedia technologies such as smartphones, tablets, digital cameras, and so on, there has been a fast increase in the number of digital videos. There are several difficulties in manipulating large amounts of video data, such as limited memory size for storing video and the time required to watch the video to understand its contents. To provide users with fast video indexing, retrieval, and browsing, while also reducing storage, an effective video management method was required. Many researchers have recently become interested in video summarization due to its benefits in a variety of applications such as video retrieval, information browsing, video indexing, video-on-demand, distance education, digital video libraries, and geographical information systems, etc.[1].

The video summary is a shortened version of the original video, and its purpose is to provide the user with an easily understood summary of the video sequence. There are two types of video summaries: static video summaries (storyboards) and dynamic video skimming. Dynamic video skimming (or a moving storyboard) selects the most pertinent small dynamic portions of video and audio to create the video summary. While a static video summary, or storyboard, contains a series of still images from the video sequence. The main advantage of video skimming is that it is more entertaining and expressive because the final summary contains video and audio; however, the user must watch a small version of a video sequence to understand its content. Otherwise, a static video summary is better suited for indexing, retrieval, and browsing. It is not constrained by synchronization or timing issues, and it allows the user to get a quick overview of the video [2-5].

We concentrate on a static video summarization method that extracts a small but significant number of silent frames (keyframes). These frames represent the entire video's content and are dissimilar to one another as possible. There are two types of transitions between shots: abrupt (cut) shot transitions and gradual shot transitions with video editing special effects (fade-in, fade-out, dissolving, wipe). A cut transition is an abrupt change in visual content between two shots. Lighting variations, object, and camera movement, and so on are some of the challenges that affect transition detection.

In this paper, we present a keyframe video summarization method based on the binary descriptor. The proposed method extracts video frame key points and descriptors using a Binary Robust Invariant Scalable Key points (BRISK)[9]. Then, for feature matching, we use the brute-force method. The Hamming distance is used to specify the distance as a measure of similarity between two descriptors. To detect shot boundaries, the ratio of matched number of key points to the total number of consecutive frames is measured. This ratio prevents false detection caused by too few key points generated from a simple frame or frame with few colors. The middle frame of a shot is chosen to be the keyframe, for generating the final summary.

The remainder of the paper is structured as follows: in section II, we review some video summarization methods. The proposed method for video summarization is presented in Section III. Section IV displays the datasets and evaluation metrics that were used to evaluate the video summary. Section V discusses the experimental results of our proposed method. Section VI shows the paper's conclusion.

II. Related work

A significant amount of research effort has been devoted to video summarization in general. To generate a summary from a long video sequence, various techniques are used. These techniques can be classified based on visual features such as (pixel, edge, motion, etc.); some methods use clustering techniques, while others rely on descriptors. Two types of descriptors are global descriptors like a histogram and local descriptors like scale invariant features transform (SIFT) and speed up robust features (SURF). A brief summary of them is provided below.

A. Pixel based methods:

Pixel based methods refer to Pixel-wise comparison that calculates the difference between the corresponding pixels of intensity or color values in two consecutive frames and compares it to a predefined threshold. Several methods for pixel-wise comparison were used, including the total sum of pixel differences [10], the absolute sum of pixel differences [11], total partial differences of pixels [12], a likelihood ratio [14], and the weighted sum of pixel differences [15]. Pixel-based methods are sensitive to camera and object movement and have a high false alarm rate. The technique [16] is sensitive to global motion because of its dependence on spatial location. Missed detections have occurred [17], although the technique is very sensitive to motion. These methods based on threshold do not take into account the temporal relation of similarity/dissimilarity signal.

B. Edge based methods:

Edge based methods are frame low-level features. The shot boundary is determined by a large difference between the current frame edges and the previous frame edges that have vanished. Edge-based methods such as a Canny edge detector [18], a wavelet transform [19], and a Robert edge detector [20]. Because they are more stable to different lighting changes, these methods can remove false positives caused by the occurrence of a flashlight (sudden lighting changes) [21,22]. Furthermore, edge-based methods are much more computationally expensive, and their performance is inferior when compared to histogram-based methods [23,24].

C. Motion based methods

Motion based methods compute the vectors of motion between consecutive frames by employing a block matching algorithm to differentiate between camera operations and transitions such as zooming and panning. A block matching algorithm [25] and a linear motion [26] extract motion vectors from compressed video sequences. However, these techniques are inconvenient for uncompressed video sequences that require significant computational power to estimate motion vectors [27].

D. Clustering based methods:

The clustering algorithms cluster video frames and then select the centroids of each cluster to generate the final summary. The k-means clustering method [28], a graph-based technique called "modularity" [29], Delaunay Triangulation [30], Farthest Point-First (FPF) algorithm [31], and Density-Based Spatial Clustering (DBSCAN) [32] are all used in video summarization.

E. Descriptor based methods:

The features are represented using two types of descriptors: global descriptors like the histogram and local descriptors like SIFT and SURF. The global descriptors are simple computation, fast execution, and a small amount of memory required. However, it is sensitive to significant motion, cannot distinguish the image foreground from the background, and is susceptible to occlusion and clutter. Otherwise, noise, scale, rotation, and illumination have no effect on the local features. It is useful for several application such as object recognition and images matching, but it requires a large amount of memory.

1) Global descriptor:

Color histograms or gray level of successive frames are computed using histogram-based methods. If the bin-wise difference between the adjacent frames histograms exceeds a certain threshold, the shot boundary is detected. There are several methods for comparing the two histograms, including an absolute difference between the corresponding bins (with or without a color similarity matrix [33]), chi-square comparison [35], histogram intersection [36], and the absolute difference of the histogram of consecutive frames [46]. Histogram based methods are resistant to object and camera motion, but when comparing two images with similar histograms, a shot change is frequently lost.

2) Local descriptors:

Local descriptors are divided into two types: floating-point descriptors and binary descriptors. SIFT, SURF, and KAZE were floating-point descriptors that computed Euclidean distance. Oriented Fast and Rotated BRIEF (ORB), Binary Robust Invariant Scalable Key points (BRISK), and Fast Retina Key point (FREAK) are all binary descriptors. The use of Hamming distance calculation is one of the key advantages of binary descriptors. Because it relies on XOR calculation, it is extremely quick.

Several recent approaches to shot boundary detection and keyframe extraction in the literature are based on SIFT [37,38,44] and SURF [39,40,45]. These methods are employed to extract key points and descriptors from a video sequence. The Euclidean distance between key points is calculated to match the descriptors of adjacent frames, then this distance is compared to a predefined threshold and the keyframe is extracted. The authors of [41] summarized the performance evaluations of SIFT and SURF in lecture videos. According to the experimental results, the SIFT detected more key points than the SURF, but at a higher computational cost. The SURF a good performance, fast and requires less time.

Despite this success, there are still significant challenges in segmenting video and extracting keyframes for effective video summarization due to the complexities caused by camera operations, lighting changes, object motions, and high computational cost. As a result, we propose a new method for detecting video shot boundaries based on the ratio of the matched number of key points to the total number of key points in two adjacent frames. This ratio is used to compare the similarity of two consecutive frames. Because of the efficiency and representativeness of the middle frame of the shot, it is chosen as a keyframe.

III. THE PROPOSED FRAMEWORK

In this method, we relied on BRISK[9], which is used for the first time in summarizing the video. The method is distinguished with simple, fast computation since it depends on hamming distance for comparing the bit-string binary descriptors. It is unaffected by rotation, scale, or simple lighting changes..

BRISK detects corners based on the AGAST algorithm and uses the FAST Corner score to filter them while searching for maxima across both the image and scale dimensions. The BRISK descriptor is a bit-string binary descriptor that estimates the patch's orientation by identifying the characteristic direction of each feature, and the concatenation of simple brightness tests. The BRISK algorithm is unaffected by rotation, scale, simple brightness, or affine changes.

As shown in Fig. 1, the proposed video summarizing approach consists of multiple phases. First, the frames from the input video are retrieved and transformed to grayscale images. Second, using BRISK to derive the descriptor feature vectors, the key points and descriptors are generated for all grayscale frames. Finally, we employ hamming distance to execute Brute-force feature matching between two consecutive frames to find the best correspondence features in the subsequent frames from the set of feature vectors (descriptors). Fourth, we compute the ratio of the matched number of key points to the total number [38,42] to determine the similarity between two consecutive frames as given in equation(1). This ratio is resistant to noise and camera rotation and can avoid false detection caused by a few key points generated from a simple frame or frame with few colors.

$$avg(t) = \frac{2matches(t)}{KP(t-1) + KP(t)} \qquad (1)$$

Where $KP(t-1)\ and\ KP(t)$ is the total number of key points generated from consecutive frames (frame t and frame $t-1$), $matches(t)$ are the matched key points number between consecutive frames.

Sixth, moving average value at frame t is computed in equation (2):

$$movavg(t) = 1/k \sum_{i=t-k}^{t-1} avg(i) \qquad (2)$$

where k is the length of frames that are used to compute the moving average value.

The visual features for the frames surrounding the boundary are more different from the frames within a shot. We measure the change in the similarity as the difference between $avg(t)$ and $movavg(t)$ in equation (3). Cut transition is detected when the difference is greater than a predefined threshold.

$$dif = movavg(t) - avg(t) \qquad (3)$$

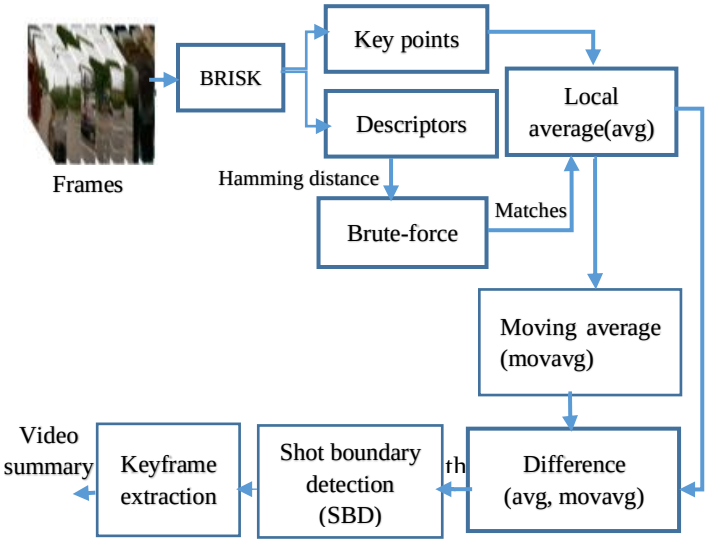

Fig. 1. Block diagram of the proposed method

In the last step, a single frame of a shot is choosing as a keyframe to generate the video summary. In our method, we chooses a shot middle frame as a keyframe because it provides an appropriate balance between efficiency and representativeness.

IV. DATASET AND EVALUATION METRIC

This section presents an Open Video Project (OVP) dataset [43] and a Comparison of User Summaries (CUS) evaluation metric[5] that are used to evaluate the proposed method.

All of the videos are MPEG-1 encoded and are divided into several genres (educational, documentary, historical, lecture, ephemeral). The videos cover a variety of colors, motion characteristics, and lengths, with durations ranging from 1 to 4 minutes. The user summaries (ground truth) were created by 50 users, with 5 summaries created by 5 different users for each video. The proposed method's final video summary is compared to other techniques available in the literature based on the same dataset. The proposed method is assessed using a quantitative metric known as CUS. Several users manually create the video summary from the sampled frames. The user summaries are used as the ground truth against which the summaries of different methods are compared. The proposed summary's quality is evaluated using two metrics, accuracy rate (CUSA) and error rate (CUSE), as defined in equations (4 and 5).

$$CUS_A = \frac{n_{mAS}}{n_{US}} \qquad (4)$$

$$CUS_E = \frac{n_{\bar{m}AS}}{n_{US}} \qquad (5)$$

where n_{mAS} defines the matched keyframe numbers from automatic summary (AS), $n_{\bar{m}AS}$ represents the non-matched keyframe numbers of from AS, and n_{US} are the keyframe numbers from the user summary (US).

V. EXPERIMENTAL RESULTS

All Experiments were conducted on an Intel(R)Core(TM) i5-2410M CPU 2.30GHz processor and 4 GB of memory. The proposed method was implemented with Python 3.7.6 and OpenCV 4.2.0 platform. We use the OVP dataset and CUS evaluation metric described in section IV to evaluate the proposed method.

We compare the proposed method against 5 different video summarization methods which also employ the OVP dataset. These methods are Delaunay Triangulation (DT) [30], Still and Moving Video Storyboards (STIMO) [31], VSUMM1 [5], VSUMM2[5], and OVP summaries. DT algorithm used to cluster the video frames based on the Delaunay Triangulation method and picked each cluster's centroid to generate the summaries. STIMO generated static and dynamic summaries by using a fast-clustering algorithm called Farthest Point-First (FPF) [43]. VSUMM1 and VSUMM2 were very simple mechanisms that used the K-means clustering algorithm and HSV color histograms. The only difference between VSUMM1 and VSUMM2 is that VSUMM1 chose one keyframe per cluster, whereas VSUMM2 chose one keyframe per key cluster.

Figures 2 and 3 show the video summaries generated by the proposed method for set videos, as well as summaries generated by other summarization algorithms. The summaries are created by five different users manually. In both examples, the proposed method summary included all user-defined content.

Fig. 2 shows that both the proposed method, OV and STIMO generate summary covered all summaries generated manually by 5 different users than other methods. The same happens in the example shown in Fig. 3, where the summary of the proposed method, VSUMM1, and OV is better than the summary of other methods when compared with the users' summaries. Table I shows the accuracy rate and error rate for the proposed method summary and the other compared video summarization methods.

In experiment America's New Frontier, segment 03, the proposed method achieves the best results where increases the accuracy rate (1.0) and decreases the error rate (0.0). The OV and STIMO give very well accuracy rates but increase the error rate. The accuracy rate and the error rate of other methods are mentioned in the table. The same happens in the Voyage of the Lee, segment 15 experiment that is indicated in Table I. The proposed method, VSUMM1, and OV gives the best accuracy rate, however the other methods obtained worse results due to their respective summaries size. The summary of OV covered the user summaries content but gave a bigger size of the final summary led to an increase in the error rate. DT and VSUMM1 generated shorter summaries with less satisfying content. Table I shows that, when the proposed method is compared with some video summarization methods, it produces good results.

Fig. 2. a) User summaries and b) methods summaries of the *America's New Frontier, Segment 03 video.*

TABLE I. EVALUATION METRICS of the methods summaries for the OVP database.

Experiment name	Evaluation metrics		
	Methods	***CUSA***	***CUSE***
▪ The America's New Frontier, segment 03 ▪ Numbers of Frames: 2,166 ▪ Duration: 1:12m	VSUMM1	0.96	0.04
	VSUMM2	0.76	0.04
	OV	1.00	0.40
	DT	0.96	0.04
	STIMO	1.00	0.40
	PROPOSED METHOD	**1.00**	**0.0**
▪ The Voyage of the Lee, segment 15. ▪ Numbers of Frames: 2,094 ▪ Duration: 1:15m	VSUMM1	0.93	0.60
	VSUMM2	0.79	0.13
	OV	1.00	1.62
	DT	0.81	0.29
	STIMO	0.89	0.86
	PROPOSED METHOD	**0.97**	**0.56**

Fig. 3. a) User summaries and b) methods summaries of the *Voyage of the Lee, Segment 03* video.

More tests on the OVP dataset and the use of the CUS assessment metric to obtain the accuracy rate are shown in Fig. 4. The accuracy rate findings from analyzing several categories of videos and comparing them to other video summarizing methods are shown in Fig. 4.

In the Great Web of Water, segment 02, the proposed method gives accuracy 0.8 better than the accuracy rate of VSUMM1, VSUMM2, and DT but less than OV and STIMO.

According to the experiment of Exotic Terrane, segment 08 which is indicated in the bar chart, the proposed method gets the best accuracy (0.9) when compared with the accuracy of other methods.

VI. Conclusion

This paper describes a method for static video summarization that employs BRISK to extract key points and descriptors from video frames. The methodology pipeline has been designed in such a way that the approach outperforms previous methods and is easier to integrate into a wide range of applications. Experiments were carried out on the OVP dataset using various video types. Fair comparisons are carried out using the same dataset and CUS metric to compare the final video summaries produced by our proposed method with those produced by other methods. Because the proposed method is based on the middle frame of a shot, the keyframes provide adequate coverage of video content.

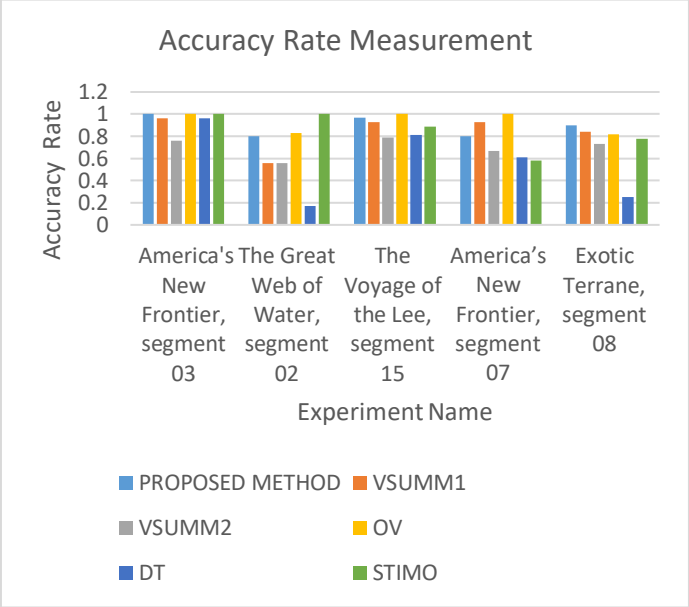

Fig. 4. Accuracy Rate evaluation of the proposed method compared with some video summarization methods.

References

[1] I. Koprinska, and S.Carrato, "Temporal video segmentation: A survey." Signal processing: Image communication 16.5, pp. 477-500, 2001.

[2] Z. Li, G. M. Schuster, and A. K. Katsaggelos, "MINMAX optimal video summarization," IEEE Trans Circuits Syst. Video Technol., vol.15, no.10, pp.1245-1256, 2005.

[3] C. Panagiotakis, A. Doulamis, and G. Tziritas, "Equivalent keyframes selection based on iso-content principles," IEEE Trans. Circuits Syst. Video Technol., vol.19, no.3, pp.447-451, 2009.

[4] G. Guan, Z. Wang, S. Lu, J. D. Deng, and D. D. Feng, "Keypoints-based keyframe selection," IEEE Trans. Circuits Syst. Video Technol., vol.23, no.4, 2013.

[5] S. E. D. Avila, A. B. P, Lopes, L. J. Antonio, and A. d. A. Araujo, "VSUMM: a mechanism designed to produce static video summaries and novel evaluation method," Pattern Recognition Letter, vol.32 (1), pp.56-68, 2011.

[6] H. Yoo, H. Ryoo, and D. Jang. Gradual shot boundary detection using localized edge blocks. Multimedia Tools and Applications, 28(3):283–300, 2006.

[7] R. Zabih, J. Miller, and K. Mai. A feature-based algorithm for detecting and classifying scene breaks. ACM Multimedia 95, pages 189–200, 1995.

[8] R. Lienhart. Comparison of automatic shot boundary detection algorithms. Proc. IS&T/SPIE Storage and Retrieval for Image and Video Databases VII, 3656:290–301, 1999.

[9] S. Leutenegger, C. Margarita, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints." International conference on computer vision. Ieee, 2011.

[10] M.S. Lee, Y.M. Yang, S.W. Lee, "Automatic video parsing using shot boundary detection and camera operation analysis", Pattern Recognit, 34, pp. 711–719, 2001.

[11] T. Kikukawa, S. Kawafuchi, “Development of an automatic summary editing system for the audio-visual resources”, Trans. Inst. Electron. Inf. Commun. Eng, 75, pp. 204–212, 1992.

[12] A. Nagasaka, Y. Tanaka, “Automatic video indexing and full-video search for object appearances”, In Visual Database Systems II; North-Holland Publishing Co.: Amsterdam, The Netherlands, pp. 113 127, 1992.

[13] H. Zhang, A. Kankanhalli, , S.W. Smoliar, “Automatic partitioning of full-motion video”, Multimedia Syst, 1, pp. 10–28, 1993,.

[14] R. Kasturi, R. Jain, "Computer Vision: Principles, ch. Dynamic vision", IEEE Computer Society Press, Washington DC, pp. 469-480, 1991.

[15] S. Lian, “Automatic video temporal segmentation based on multiple features”, Soft Comput., pp. 469–482, 2011.

[16] Z.H.Z. Huan, L.X.L. Xiuhuan, Y.L.Y. Lilei, “Shot Boundary Detection Based on Mutual Information and Canny Edge Detector” , Int. Conf. Comput. Sci. Softw. Engineering, 2, pp. 1124–1128, 2008.

[17] I. Koprinska, S. Carrato, “Temporal video segmentation: A survey”, Signal Process. Image Commun, 2001, 16, 477–500

[18] R. Zabih, J. Miller, K. A. Mai, “Feature-Based Algorithm for Detecting and Classifying Scene Breaks”, In Proceedings of the Third ACM International Conference on Multimedia Multimedia 95; San Francisco, CA, USA, 5–9, Volume 95, pp. 189–200, Novenber 1995.

[19] J. Nam, A.H. “Tewfik, Combined audio and visual streams analysis for video sequence segmentation”, In Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP-97), Munich, Germany, Volume 4, pp. 2665–2668, 21–24 April 1997.

[20] J. Zheng, F. Zou, M. Shi, “An efficient algorithm for video shot boundary detection”, In Proceedings of the 2004 IEEE International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, China, pp. 266–269, 20–22 October 2004.

[21] W.J. Heng, K.N. Ngan, “High accuracy flashlight scene determination for shot boundary detection”, Signal Process. Image Commun, 18, pp. 203–219 2003.

[22] S.H. Kim, R.H. Park, “Robust video indexing for video sequences with complex brightness variations”, In Proceedings of the lnternational Conference on Signal and Image Processing, Kauai, HI, USA, pp. 410–414, 12–14 August 2002.

[23] A. Dailianas, R.B. Allen, P. England, “Comparison of automatic video segmentation algorithms”, In Proceedings of SPIE–The International Society for Optical Engineering; SPIE: Philadelphia,PA, USA; Volume 2615, pp. 2–16, 1996.

[24] R.W Lienhart, “Reliable transition detection in videos: A survey and practitioner’s guide”, Int. J. Image Graph, 1,pp. 469–486, 2001.

[25] B. Shahraray, “Scene change detection and content-based sampling of video sequences”, In IS&T/SPIE’s Symposium on Electronic Imaging: Science & Technology; International Society for Optics and Photonics: San Jose, CA, USA, pp. 2–13, 1995.

[26] E. Bruno, D. Pellerin, “Video shot detection based on linear prediction of motion”, In Proceedings of the 2002 IEEE International Conference on Multimedia and Expo (ICME’02), Lausanne, Switzerland, Volume 1, pp. 289–292, 26–29 August 2002.

[27] I.A. Zedan, K.M. Elsayed, E. Emary, “Abrupt Cut Detection in News Videos Using Dominant Colors Representation”, In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016; Hassanien, A.E., Shaalan, K., Gaber, T., Azar, A.T., Tolba, M.F., Eds.; Springer International Publishing: Cham, Switzerland, pp. 320–331, 2017.

[28] X. Wenzhu, X. Lihong, “A novel shot detection algorithm based on clustering”, In: 2nd international conference on education technology and computer, vol 1, pp 570–572, 2010.

[29] A. Choudhury, G. Medioni, “A framework for robust online video contrast enhancement using modularity optimization”, IEEE Trans Circ Syst Video Technol 22(9):1266–1279, 2012.

[30] P. Mundur, Y . Rao, Y. Yesha, “Keyframe-based video summarization using delaunay clustering”, Int J Dig Libr 6:219–232, 2006.

[31] M. Furini, F. Geraci, M. Montangero, M. Pellegrini, “STIMO: STIll and MOving video storyboard for the Web scenario”, In: Multimedia Tools and Applications, vol 46. Kluwer Academic Publishers, MA, USA, pp 47–69, 2010.

[32] K. Mahmoud, MA. Ismail, NM. Ghanem, “VSCAN: An Enhanced Video Summarization Using Density-Based Spatial Clustering”, In: Lecture Notes in Computer Science, vol 8156. Springer, pp 733– 742, 2013.

[33] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin, “The QBIC project: query images by content using color, texture, and shape,” Proc. SPIE, vol. 1908, pp. 173–181, 1993.

[34] B. M. Mehtre, M. S. Kankanhalli, A. D. Narasimhalu, and G. C. Man,“Color matching for image retrieval,” Pattern Recognit. Lett., vol. 16, pp. 325–331, 1994.

[35] M. J. Swain and D. H. Ballard, “Color indexing,” Int. J. Comput. Vis., vol. 26, no. 4, pp. 461–470, 1993.

[36] R. Hannane, A. Elboushaki, K. Afdel, P. aghabhushan, and M. Javed, “An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram”. International Journal of Multimedia Information Retrieval, 5(2), 89 104, 2016.

[37] G. Liu, X. Wen, W. Zheng, and P. He, “Shot boundary detection and keyframe extraction based on scale invariant feature transform”, In 2009 Eighth IEEE/ACIS International Conference on Computer and Information Science, pp. 1126-1130, June 2009.

[38] J. Baber, N. Afzulpurkar, and M. Bakhtyar,”Video segmentation into scenes using entropy and SURF”, Emerging Technologies (ICET), 7th International Conference on pp.1–6, 2011.

[39] J. Baber, N. Afzulpurkar, and S. Satoh, “A framework for video segmentation using global and local features”. International Journal of Pattern Recognition and Artificial Intelligence, 27(05), 1355007-1 - 1355007-29, 2013.

[40] S. Athani, , & C. Tejeshwar, “Performance analysis of key frame extraction using SIFT and SURF algorithms”, IJCSIT Int. J. Comput. Sci. Inf. Technol, 7(4), 2136-2139, 2016.

[41] Z. El Khattabi, Y. Tabii, and A. Benkaddour. "Video Shot Boundary Detection Using The Scale Invariant Feature Transform and RGB Color Channels." International Journal of Electrical & Computer Engineering , pp. 2088-8708, 2017.

[42] The open video project, http://www.open-video.org

[43] T. Gonzalez, “Clustering to minimize the maximum intercluster distance”, Theor Comput Sci 38:293–306, 1985.

[44] S. Jadon, & M. Jasim, “Unsupervised video summarization framework using keyframe extraction and video skimming”. In 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA) (pp. 140-145), October 2020.

[45] B. Liang, N. Li, , Z. He, Z.Wang, Y. Fu, & T. Lu, “News Video Summarization Combining SURF and Color Histogram Features”, Entropy, 23(8), 982, 2021.

[46] H. B. Taher, & A. H. Awadh, “Video Summarization for Surveillance System Using key-frame Extraction based on Cluster”. Journal of Education for Pure Science-University of Thi-Qar, 11(1), 54-65, 2021.