Video shot boundary detection based on frames objects comparison and scale-invariant feature transform technique

Noor Khalid Ibrahim, Zinah Sadeq Abduljabbar
Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq

Article Info

Article history:

Received Dec 12, 2023
Revised Feb 24, 2024
Accepted Mar 4, 2024

Keywords:

Frames correlation
Object comparison
Shot boundary
Video analysis
Video segmentation

ABSTRACT

The most popular source of data on the Internet is video which has a lot of information. Automating the administration, indexing, and retrieval of movies is the goal of video structure analysis, which uses content-based video indexing and retrieval. Video analysis requires the ability to recognize shot changes since video shot boundary recognition is a preliminary stage in the indexing, browsing, and retrieval of video material. A method for shot boundary detection (SBD) is suggested in this situation. This work proposes a shot boundary detection system with three stages. In the first stage, multiple images are read in temporal sequence and transformed into grayscale images. Based on correlation value comparison, the number of redundant frames in the same shots is decreased, from this point on, the amount of time and computational complexity is reduced. Then, in the second stage, a candidate transition is identified by comparing the objects of successive frames and analyzing the differences between the objects using the standard deviation metric. In the last stage, the cut transition is decided upon by matching key points using a scale-invariant feature transform (SIFT). The proposed system achieved an accuracy of 0.97 according to the F-score while minimizing time consumption.

Corresponding Author:

Noor Khalid Ibrahim
Department of Computer Science, College of Science, Mustansiriyah University
Baghdad, Iraq
Email: noor.kh20@ uomustansiriyah.edu.iq

1. INTRODUCTION

The vast amount of video content on the internet makes it challenging to develop effective indexing and search strategies for managing video data. Content-based video retrieval is emerging as a trend in video retrieval systems, while conventional methods like video compression and summarizing aim for minimal storage requirements and maximum visual and semantic accuracy [1]. Given that video is the most sophisticated sort of multimedia data, it includes information about the target's mobility within the scene as well as information about the objective world changing with time [2].

Two modules can be approximately regarded in video segmentation which are video object (foreground/background) segmentation, and video semantic segmentation [3]. Video segmentation, also known as shot boundary detection (SBD), involves breaking the video up into meaningful scenes so that the essential feature(s) may be found in each scene through analysis [4]. A cut is a sudden change in the shot that takes place inside a single frame. A fade is a gradual alteration in brightness that often begins or ends with a completely dark frame. Frames inside the transition show one image overlaid on the other during a dissolve, which happens as the images of the first shot go darker and the images of the second shot get brighter [1]. The primary difficulties in shot boundary recognition are movements of the camera and objects since these can significantly change the video content, producing an effect akin to transition effects and leading to inaccurate shot transition detection [5].

Numerous studies have addressed video segmentation, Hong Shao *et al*. [6] utilized a combination of a color histogram with Hue Saturation Value (HSV) and features of histogram of gradient (HOG) to effectively detect abrupt shot changes in videos. In [3] This work proposes a shot boundary detection approach based on the scale-invariant feature transform (SIFT). Using a top-down search strategy, the initial phase of this approach compares the ratio of matched features derived by SIFT for each RGB channel of video frames to locate transitions. The boundaries' locations are shown in the overview stage. Second, to ascertain the kind of transition, a moving average computation is made.

In [7] The research aimed to use a multi-modal visual features-based SBD framework; the behaviors of the visual representation are analyzed concerning the discontinuity signal. This used a candidate segment selection strategy that does not compute the threshold; instead, it utilizes the discontinuity signal's cumulative moving average to determine the shot boundary locations while disregarding the non-boundary video frames. To differentiate between a candidate segment that is a cut transition and one that is a gradual transition, including fade in/out and logo occurrence, the transition detection is carried out structurally.

In [8] the proposed temporal video segment representation formalizes video scenes as temporal motion change data, determining motion modifications and cuts between scenes through optical flow character changes. This reduces the issue to an optical flow-based cut detection problem, enhancing a pixel-based representation. The proposed video segment representation divides temporal video segment points into cuts and non-cuts.

In [9] the bag of visual word (BoVW) model, which splits the video into shots and keyframes, is the basis for the segmentation model for videos that the study suggested. The BoVW model is employed in two variants: the traditional BoVW and an expansion known as the vector of linearly aggregated descriptors (VLAD). Keyframe feature vectors inside a sliding window of length L are used to calculate similarity. In [10] The study presents a method for feature fusion and clustering technique (FFCT)-based video shot boundary detection, which involves converting interval frames into grayscale images, extracting fingerprint and speed-up robust features, fusion, and clustering them using a K-means algorithm. Linear discriminant analysis (LDA) is introduced for cluster mapping, and features are chosen using density computation based on frame correlation.

In [2] a novel algorithm for camera detection based on SIFT features was introduced in this study. The proposed method involves the analysis of multiple frames of images in a sequential manner. Initially, the images are converted into grayscale and divided into blocks. Subsequently, the dynamic texture of the film is computed, and the correlation between the dynamic texture of adjacent frames and the matching degree of SIFT features is determined. Based on these matching results, pre-detection outcomes are obtained.

Idan *et al*. [11] proposed a fast video processing method for SBD. To reduce computing costs and disturbances, the proposed SBD framework makes use of candidate segment selection with frame active area and separable moments. Inequality criteria and adaptive threshold are used to exclude non-transition frames and maintain candidate segments. Cut transition detection is done using machine learning statistics.

In [12] a practical SBD method was presented in the study, which uses average edge information for gradual transition detection and gradient and color information for abrupt transition detection. Processing only transition regions yield an average edge frame and reduces computational complexity. In [5] The proposed method comprises two distinct stages. In the initial stage, projection features were employed to differentiate between non-boundary transitions and candidate transitions that potentially encompass abrupt boundaries. Consequently, only the candidate transitions were retained for further analysis in the subsequent stage. This approach effectively enhances the speed of shot detection by minimizing the detection scope. In [13] An effective SBD approach with several invariant properties was presented in this work. With the right mix of invariant features, such as edge change ratio (ECR), color layout descriptor (CLD), and scale-invariant feature transform (SIFT) key point descriptors, the accuracy level of SBD was increased.

According to the literature, many applications have been created to address the issue of shot boundary detection in videos. These applications are performed based on various techniques to process the challenges in SBD. This proposed SBD system has been achieved in three stages to improve its performance and try to reduce the problem of object and camera motion, wherein the first stage the redundancy frames in the same shots are reduced based on correlation value comparison, this stage yields minimizing time-consuming and computation complexity. Then in the second stage candidate transition is determined by comparing the objects of sequential frames, final stage the decision of the cut transition is made based on key points matching of SIFT method. This proposed method aims to find the boundary frame of a shot with a cut transition between consecutive shots accurately. The rest of the paper is organized as follows, section 2 explains the proposed method, the experimental result, and the analysis demonstrated in section 3, followed by a conclusion in section 4.

2. SBD PROPOSED METHOD

This proposed SBD system has been achieved in three stages, in the first stage, multiple images are read in temporal sequence and transformed into grayscale images. Based on correlation value comparison, the number of redundant frames in the same shots is decreased, and then, in the second stage, a candidate transition is identified by comparing the objects of successive frames using the proposed method to extract frame image objects. In the last stage, the cut transition is decided upon by matching key points using the SIFT approach. The details of these stages are explained as follows:

2.1. Reduces redundancy stage

The multiple frames of input video are extracted as the first step, then converted into grayscale and resized into 256*256. Some pre-processed operations are achieved on these frames to improve their quality when the noise is removed by the wiener filter [14], and contrast is enhanced by histogram equalization [15]. The resulting frames are normalized in the range [0-1].

In one shot the consecutive frames have a very high similarity, and achieving the SBD process on each pair of frames will be very time-consuming and computationally complex. So, to minimize this time and complexity the redundancy frames in one shot have been reduced based on the measure of their correlation value. The correlation value (r) of frames Fr(i) and Fr(i+1) and based on the threshold value (Th) identified experimentally the frame Fr(i) is passed to the next stage, otherwise, frame Fr(i) is discarded as demonstrated in (1). Where the correlation value is calculated as explained in (2) [16].

$$\mathrm{Fr(i)} \begin{cases} \text{Passed to next stage} & r < \ \mathrm{Th} \\ \text{discard} & \text{otherwise} \end{cases} \tag{1}$$

$$r = \frac{\sum_i (x_i - x_m)(y_i - y_m)}{\sqrt{(\sum_i (x_i - x_m)^2} \sqrt{(\sum_i (y_i - y_m)^2}} \tag{2}$$

where x_i denotes the pixel intensity in order ith of the first image, and y_i demarcated the ith pixel intensity of the second image, additionally, x_m and y_m is the mean intensity of first and second images sequentially.

2.2. Selection of candidate transition stage

Candidate transition selection is performed based on comparison made on consecutive frame objects, that means on frame image content. This image content extraction is achieved based on the proposed extraction method as explained in Figure 1in this stage. As seen in the figure, the objects of the frame have been extracted in two steps, which are the generation of the feature template, and extract the object, these steps are detailed as following:

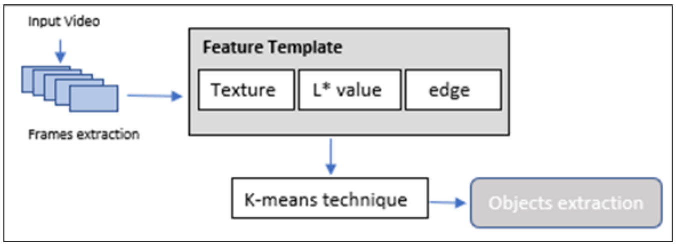


Figure 1. Frame objects extraction flowchart

2.2.1. First step (generate features template)

For each consecutive frame passed to this stage the template of features is generated when multiple features are extracted and combined from each frame image. The selection of these multiple features must be able to extract the objects of a frame image accurately, so in this proposed extraction method of this proposed SBD algorithm, the multiple features are represented by the texture characteristics that yield information about the local variability of the pixel's intensity values are recovered using the standard deviation filter (SD) [17] of the 3-by-3 neighborhood around the consistent pixel. The value luminance grayscale of these processed frames is represented by channel L* in the L* a* b* color space [18] used as second feature. The L*a*b* typically appears to be able to depict the colors to human vision. Additionally, because the RGB representation includes a transition color between blue and green, the L*a*b* color representation compensates for the diversity in the color distribution in the RGB color model [19]. For this reason, L*a*b* is taken into account along with its L* value. These two feature matrices are then merged with the edge of the detected frame by a canny operator which has the ability to recognize object boundaries in an image and object appreciation to create a feature template. The following is how SD is calculated [20].

$$\mu_j = \frac{1}{N}\sum_{i=1}^{N} x_{ji} \tag{3}$$

$$\sigma_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{ji} - \mu_j)^2} \tag{4}$$

2.2.2. Second step (objects extracted)

Utilize the k-means [21] algorithm with this created template to extract the objects from these successive frames. A k-number group of data is gathered in order to use K-means. kmeans method consists of two stages. In the first, the centroid is initialized, and in the second, the distance to the closest centroid is used to identify which cluster the data point belongs to. Because of its ease of use and quick calculation, the k-means clustering approach is widely utilized in clustering processes [22], which is the reason that it was chosen for this phase. Consequently, the frame image object has been identified based on this proposed extraction method with generated features template and K-means technique.

The frames' similarity has been measured based on the objects' comparison by dividing images of objects of related sequential frames into 8×8 blocks, and then the entropy value of each block is calculated, in turn, these entropies values are arranged into vectors of the length 64, which represent similarity measurement vectors as explained in Figure 2, and then the standard deviation is calculated to differences between these two entropies vectors of object images of consecutive frames when the value of standard deviation is nearest to zero normal transition has been distinguished. According to the threshold (Thr) value perceived experimentally, the abrupt transition has been a candidate, otherwise, normal transition has been detected as demarcated in (5). Entropy value is determined as in (6) [23]. In turn, these candidate frames are passed to the third stage to make a decision of abrupt transition.

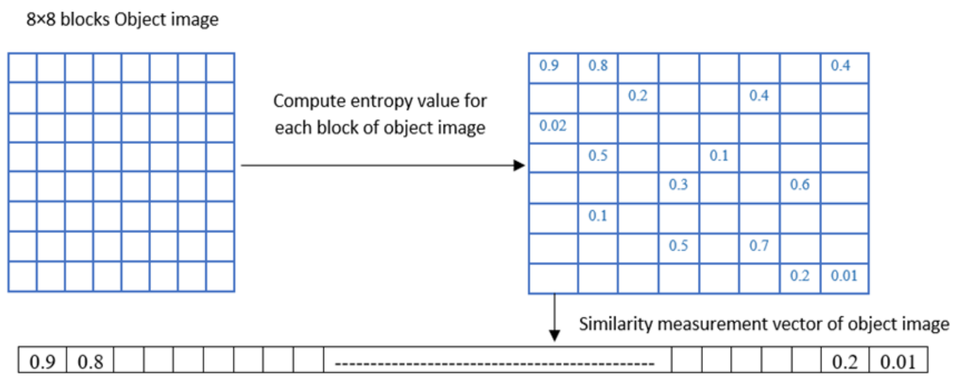


Figure 2. Construct similarity measurement vector of object image

$$Fri \begin{cases} \text{Abrupt transition candidate} & sd > Thr \\ \text{Normal transition} & \text{otherwise} \end{cases} \tag{5}$$

Let Fri represent the video frame with index i

$$H_r = -\sum {}_{g_r}^{-k} \log_2(g_r^{-k}) \tag{6}$$

Where g_r^{-k} denote distribution of assumed color space.

2.3. Transition decision stage

Making the right choice when deciding how to divide a video sequence into shots is mostly dependent on selecting the right method. David adopted a scale-invariant feature transform SIFT [24]. The SIFT feature has been used in this stage to determine the frame transition and its boundary because, given an image as input, the SIFT descriptor generates a wide range of local feature vectors that are independent of image scaling and rotation. SIFT is capable of precisely correlating two images [13]. In situations of abrupt transitions, when the matching degree of the SIFT feature between the frames is low, neighboring frames are recognized as belonging to different shots, which can better discern the moving objects in successive frames.

3. EXPERIMENTAL RESULTS AND ANALYSIS

Eight distinct videos from the standard dataset, TRECVid 2001 test data made existing on the open video project and accessible at https://open-video.org, are used to assess the suggested method in this research. These videos are referred to as Vid1 through Vid8. A comprehensive description of those input videos is provided in Table 1. The ground truth value is determined by observing abrupt changes as seen by people. The chosen videos contain a variety of aberrations, including lighting variations, viewpoint shifts, scaling, zooming, rotation, and more.

Table 1. Description of input videos

Input video	Video name	Time Duration In sec.	Frames number	Abrupt transition
Vid1	Free-for-all race at Charter Oak Park (Historical)	26	853	3
Vid2	New Indians, Segment 101 (Documentary)	131	3953	14
Vid3	New Indians, Segment 01 (Documentary)	56	1687	15
Vid4	Winning Aerospace, Segment 02 (Documentary)	65	1970	11
Vid5	Hidden Fury, segment 10 (Documentary)	33	1002	1
Vid6	Hurricanes, Segment 05 (Documentary)	115	3448	32
Vid7	The Miracle of Water, segment 05 (Documentary)	83	2314	1
Vid8	Winning Aerospace, Segment 04 (Documentary)	110	3318	18

3.1. Reduces redundancy stage

When the multiple frames of input video have been extracted, the frames images in the same shot have a high similarity degree and when performing features extraction to extract objects from each frame image results in time-consuming and computing complexity, so reducing the redundancy frames stage results in time-consuming minimization as seen in Table 2 and Figure 3, for instance, the execution time was equivalent to (111.4 seconds) when the second stage was applied to all of the vid1's frames, that means without similarity frames reduction. as opposed to the execution time (44.41 seconds) when vid1 advanced to the lower redundancy level, and so on to others videos as explained in this table that shows how much time each utilized video takes.

3.2. Selection of candidate transition stage

Based on the motion of the object and/or the camera, shots may be categorized into four types: static objects with static cameras, dynamic objects with static cameras, dynamic objects with dynamic cameras, and dynamic objects with dynamic cameras [25]. Candidate transition selection is performed based on comparison made on consecutive frames objects. This stage is achieved by comparison made to the extracted objects of frames images based on the created features template by combining multiple features texture, edges, and L* value of L*a*b* color space applied to the k-means technique.

The stage starts choosing potentially cut transition frames by examining the standard deviation to the differences of vectors created from frames object blocks for similarity comparison after the frames Fri and Fri+1 pass the first stage based on their correlation value measure. Table 3 and Figure 4 describe how the block size of the frame image object is determined empirically, where Figure 4(a) explains the block size effect on execution time and Figure 4(b) demonstrates the effect of block size on F-score. Vid3, Vid4, and Vid8 are taken as examples to demonstrate that block size affects execution time and accuracy in this table. For this investigation, 8*8 blocks with a 32*32 block size are more appropriate in this study.

Table 2. Time consuming comparison

Videos	Execution time in Sec. (with reduction)	Execution time in Sec. (without reduction)
Vid1	44.41	111.42
Vid2	224.94	785.83
Vid3	99.75	314.75
Vid4	129.07	363.44
Vid5	70.178	177.52
Vid6	320.34	566.40
Vid7	111.47	421.15
Vid8	201.23	679.02

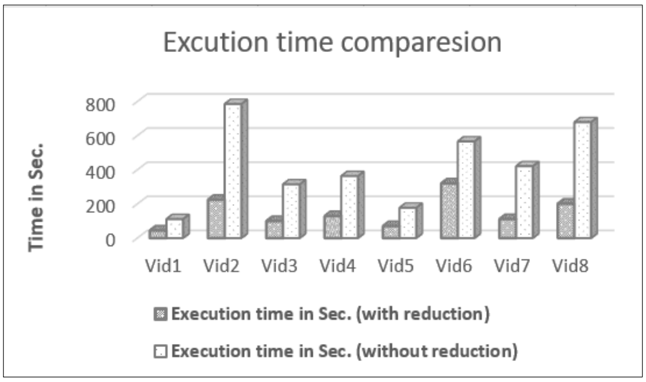
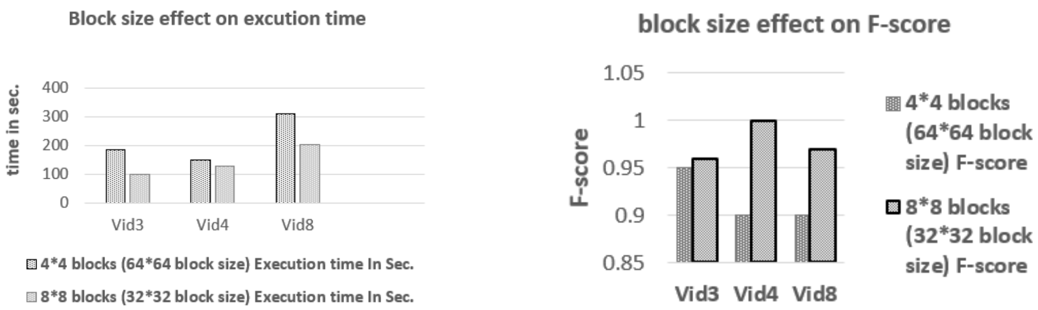


Figure 3. Comparison in execution time



(a) (b)

Figure 4. Block size effect, (a) on execution time and (b) on F-score

Table 3. The effect of block size within 256*256 frame size

videos	4*4 blocks (64*64 block size)		8*8 blocks (32*32 block size)	
	Execution time In Sec.	F-score	Execution time In Sec.	F-score
Vid3	183.58	0.95	99.75	0.96
Vid4	148.92	0.90	129.07	1.00
Vid8	311.36	0.90	201.23	0.97

To explain the frame’s object extraction, for example with samples of frames that explained in Figure 5, the frame objects extraction method steps demonstrate in Figure 6. The recovered combined features (Texture, frame edge, and L* value of L*a*b* color space) from frames i and i+1 create the template features for each one. The frame objects are then extracted for the frame similarity comparison using the k-means approach. If identical objects are found in two consecutive frames, they are likely associated with the same shot; if not, a cut shot transition is a possibility. The significant problem of object and camera movements can be addressed by similarity discovery based on object comparison because the frame object is recognized where it should be in the image of succeeding frames.

Figure 5. transitions examples

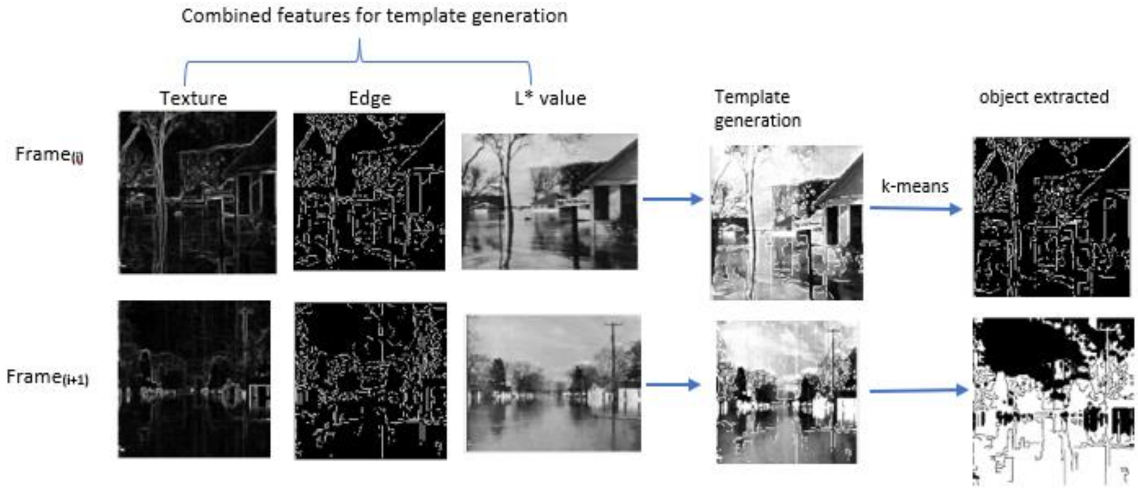


Figure 6. Example of consecutive frames objects extracted

This proposed object extraction method has been assessed for adopting in this proposed SBD algorithim. According to Table 4 and Figure 7, which describe the information content as determined by the entropy value that means the accuracy of extracting objects by the proposed extraction method of frame, in this table some frames that apply extraction its objects from some different used videos are selected as samples for evaluation. As a result of this evaluation explained in this table, and from the analysis of this evaluation, this proposed object extraction operation has been adopted in this stage of the proposed SBD algorithm.

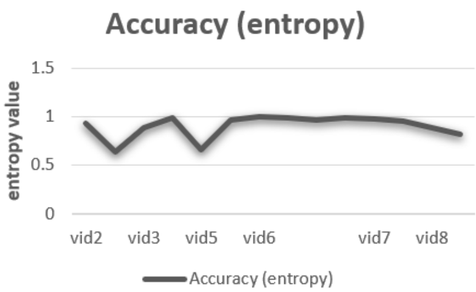


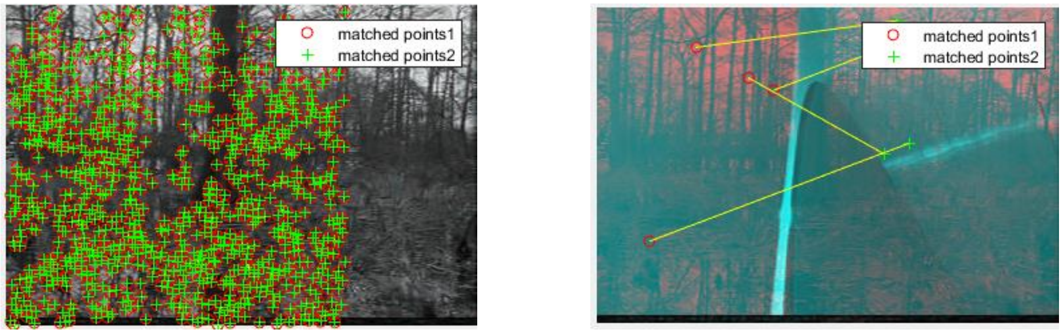
Figure 7. Object extraction accuracy using entropy

Table 4. Object extraction evaluation using entropy measure (Ent)

	Vid2		Vid3		Vid5		Vid6			Vid7		Vid8	
Fr.	397	398	262	263	432	432	82	83	568	569	1375	1376	517
Ent	0.926	0.634	0.890	0.985	0.660	0.969	0.998	0.989	0.968	0.992	0.979	0.956	0.884

3.3. Transition decision stage

The SIFT properties are adopted in this stage for shot transition decision-making because when it comes to rotation, and zoom, SIFT characteristics remain unaffected and it able to reflect the local variation of moving object efficiently, and may be used to impartially characterize the image [2]. SIFT key points are detected, features are extracted from candidate frames of video results from the previous stage, then feature matching is performed. In features matching two features' matrices of frame i, frame i+1 have been matched using distance calculation results in a p-by-1 vector, where p represents the key point number that is detected. And from the matched features shot transition decision-making, when the matching degree of the SIFT feature between the frames is low, neighboring frames are recognized as belonging to different shots. Figure 8(a) demonstrates features key point matching for frames in same shot, and Figure 8(b) from different shot.



(a) (b)

Figure 8. Frames shots feature key points matching, (a) frames in the same shot and (b) frames in a different shot

As seen in the figure, due to comparable visual features, the similarity matching between two frames in the same shot is typically high. Frames from diverse shots, however, lack visual uniformity. They therefore have either little or no similarity matching.

Recall and precision are the key performance metrics of the suggested system that are typically employed in the SBD process. The F1 score, which is the harmonic mean of precision and recall, is used in this paper's evaluation along with these metrics [2]. The following formula can be used to compute these metrics [5]:

$$R = \frac{true}{true + miss} \tag{7}$$

$$P = \frac{true}{true + false} \tag{8}$$

$$F - score = \frac{2 * P * R}{P + R} \tag{9}$$

where True denotes accurate transition detection, False denotes inaccurate transition detection, and Miss denotes missed transition detection. Table 5 demonstrates the accuracy with these metrics of this proposed SBD algorithm.

Table 5. Efficiency of the proposed method

Video	Recall	Precision	F-score
Vid1	1.00	1.00	1.00
Vid2	1.00	1.00	1.00
Vid3	0.93	1.00	0.96
Vid4	1.00	1.00	1.00
Vid5	1.00	1.00	1.00
Vid6	0.87	0.96	0.91
Vid7	1.00	1.00	1.00
Vid8	0.94	0.94	0.94
Average	0.96	0.98	0.97

3. CONCLUSION

By comparing frame image objects and using a scale-invariant feature transform SIFT feature with the discard to the redundant frames of the same shot, the suggested SBD approach has been realized. Three stages are involved in implementing this proposed system: first, the redundancy frames are reduced based on their correlation value; this reduces computation complexity and time consumption; second, the candidate shot transition and boundary are identified based on object comparison using proposed extraction method; this stage can identify objects that where should be in the image of subsequent frames. The last step then uses the SIFT feature to choose which of these candidate frames to select. The research demonstrates that this approach minimizes false positives by utilizing SIFT matching key points, which are independent of the scale and rotation of the image. Our method yields a 97% F1 score, which is high result while requiring a lesser amount of time and complexity.

ACKNOWLEDGEMENTS

Authors thank the Department of Computer Science, College of Science, Mustansiriyah University, Baghdad-Iraq for supporting this present work.

REFERENCES

[1] 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 (2088-8708)*, vol. 7, no. 5, 2017.
[2] L. Kong, “SIFT feature-based video camera boundary detection algorithm,” *Complexity*, vol. 2021, pp. 1–11, 2021.
[3] T. Zhou, F. Porikli, D. J. Crandall, L. Van Gool, and W. Wang, “A survey on deep learning technique for video segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 6, pp. 7099–7122, 2022.
[4] D. M. Thounaojam, T. Khelchandra, K. M. Singh, and S. Roy, “A genetic algorithm and fuzzy logic approach for video shot boundary detection,” *Computational intelligence and neuroscience*, vol. 2016, 2016.
[5] E. Hato, “Temporal video segmentation using optical flow estimation,” *Iraqi Journal of Science*, pp. 4181–4194, 2021.
[6] H. Shao, Y. Qu, and W. Cui, “Shot boundary detection algorithm based on HSV histogram and HOG feature,” in *2015 International Conference on Advanced Engineering Materials and Technology*, Atlantis Press, pp. 951–957, 2015.
[7] S. Tippaya, S. Sitjongsataporn, T. Tan, M. M. Khan, and K. Chamnongthai, “Multi-modal visual features-based video shot boundary detection,” *IEEE Access*, vol. 5, pp. 12563–12575, 2017, doi: 10.1109/ACCESS.2017.2717998.
[8] S. Akpinar and F. Alpaslan, “A novel optical flow-based representation for temporal video segmentation,” *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 25, no. 5, pp. 3983–3993, 2017.
[9] M. Haroon, J. Baber, I. Ullah, S. M. Daudpota, M. Bakhtyar, and V. Devi, “Video scene detection using compact bag of visual word models,” *Advances in Multimedia*, vol. 2018, pp. 1–9, 2018.
[10] F.-F. Duan and F. Meng, “Video shot boundary detection based on feature fusion and clustering technique,” *IEEE Access*, vol. 8, pp. 214633–214645, 2020.
[11] Z. N. Idan, S. H. Abdulhussain, B. M. Mahmmod, K. A. Al-Utaibi, S. A. R. Al-Hadad, and S. M. Sait, “Fast shot boundary detection based on separable moments and support vector machine,” *IEEE Access*, vol. 9, pp. 106412–106427, 2021.
[12] N. Kumar, “Shot boundary detection framework for video editing via adaptive thresholds and gradual curve point,” *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 11, pp. 3820–3828, 2021.
[13] J. T. Jose, S. Rajkumar, M. R. Ghalib, A. Shankar, P. Sharma, and M. R. Khosravi, “Efficient shot boundary detection with multiple visual representations,” *Mobile Information Systems*, vol. 2022, 2022.
[14] K. A. Akintoye, N. A. F. B. Ismial, N. Z. S. B. Othman, M. S. M. Rahim, and A. H. Abdullah, “Composite median Wiener filter based technique for image enhancement.,” *Journal of Theoretical & Applied Information Technology*, vol. 96, no. 15, 2018.
[15] S. H. Majeed and N. A. M. Isa, “Adaptive entropy index histogram equalization for poor contrast images,” *IEEE Access*, vol. 9, pp. 6402–6437, 2020, doi: 10.1109/ACCESS.2020.3048148.
[16] A. M. Neto, A. C. Victorino, I. Fantoni, D. E. Zampieri, J. V. Ferreira, and D. A. Lima, “Image processing using Pearson’s correlation coefficient: Applications on autonomous robotics,” in *2013 13th International Conference on Autonomous Robot Systems*, IEEE, pp. 1–6,2013.
[17] N. K. Ibrahim, A. H. Al-Saleh, and A. S. A. Jabar, “Texture and pixel intensity characterization-based image segmentation with morphology and watershed techniques,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 3, pp. 1464–1477, 2023. doi: 10.11591/ijeecs.v31.i3.
[18] N. khalid, “Hybrid features of mask generated with gabor filter for texture analysis and sobel operator for image regions segmentation using K-Means technique,” *Journal La Multiapp*, vol. 3, no. 5, pp. 250–258, 2022, doi: 10.37899/journallamultiapp.v3i5.743.
[19] X. Zheng, Q. Lei, R. Yao, Y. Gong, and Q. Yin, “Image segmentation based on adaptive K-means algorithm,” *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, pp. 1–10, 2018.
[20] U. Petronas, “Mean and standard deviation features of color histogram using laplacian filter for content-based image retrieval,” *Journal of Theoretical and Applied Information Technology*, vol. 34, no. 1, pp. 1–7, 2011.
[21] R. Sammouda and A. El-Zaart, “An optimized approach for prostate image segmentation using K-means clustering algorithm with elbow method,” *Computational Intelligence and Neuroscience*, vol. 2021, 2021.
[22] N. Dhanachandra and Y. J. Chanu, “A new approach of image segmentation method using K-means and kernel based subtractive clustering methods,” *International Journal of Applied Engineering Research*, vol. 12, no. 20, pp. 10458–10464, 2017.
[23] N. M. Kwok, Q. P. Ha, and G. Fang, “Effect of color space on color image segmentation,” in *2009 2nd International Congress on Image and Signal Processing*, IEEE, pp. 1–5, 2009.
[24] L. David, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, pp. 91–110, 2004.
[25] S. H. Abdulhussain, A. R. Ramli, M. I. Saripan, B. M. Mahmmod, S. A. R. Al-Haddad, and W. A. Jassim, “Methods and challenges in shot boundary detection: a review,” *Entropy*, vol. 20, no. 4, p. 214, 2018.

BIOGRAPHIES OF AUTHORS

Noor Khalid Ibrahim is lecturer at college of college of science, Mustansiriyah University, Iraq. Received the B.Sc. degree in computer science from Department of Computer, College of Science, Mustansiriyah University, Iraq. She holds a master degree in computer science at 2015, with specialization in multi-media. Her research areas in image processing. She can be contacted at email: noor.kh20@uomustansiriyah.edu.iq.

Zinah Sadeq Abduljabbar is lectuter at collage of science, Mustansiriyah university, Iraq. Received the B.Sc. degree in computer science from department of computer, collage of science, Mustansiriyah university, Iraq. She holds a master degree in computer science at 2014, with specialization in multi-media. she can be contacted at email: zinahsadeq@uomustansiriyah.edu.iq.