International Journal on Emerging Technologies **11**(2): 205-212(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Perspective Study on Content Based Video Retrieval

C. Victoria Priscilla[1] and D. Rajeshwari[2]
[1]Associate Professor, Department of Computer Science,
S.D.N.B. Vaishnav College for Women, University of Madras, Chennai (TamilNadu), India.
[2]Research Scholar, Department of Computer Science,
S.D.N.B. Vaishnav College for Women, University of Madras, Chennai (TamilNadu), India.

(Corresponding author: C. Victoria Priscilla)
(Received 14 December 2019, Revised 08 February 2020, Accepted 15 February 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The Closed-Circuit Television (CCTV) footages plays a vital role in criminal investigations which helps to reduce cost, time and effort but still it has many challenges to face such as monitoring multiple cameras simultaneously, missing pre-eminent details or the object captured on video surveillance, excess storage of video data and spending huge time on watching the entire suspect video to collect the affirmation for investigations. The Motion Detection, Facial recognition, Automatic number plate recognition through CCTV streams a live report. From this, the identification of suspicious behaviour, like public inebriation or attaining thievery from the entire video content becomes very critical to deter the criminals. To meliorate all these situation, Content Based Video retrieval (CBVR) is efficiently used to analyze the video content of the CCTV Footages. The CBVR detects the shots and frames obtained from the CCTV Footages where it analyzes the Color, texture, shape and inter-frame relations to detect the similarity between the frames. Still CBVR lacks to analyze a huge storage of video data content, which is frustrating the person for long time extraction of particular crime scene detection. With this objective, the paper reveals the study on feature extraction methods of Shot Boundary Detection (SBD) and Key-frame extraction methods.

Keywords: Closed-circuit television, Content Based Video Retrieval, Convolutional Neural Network, Key Frame Extraction, Recurrent Neural Network, Shot Boundary Detection.

Abbreviations: CBVR, Content Based Video Retrieval; CCTV, Closed-circuit Television; SBD, Shot Boundary Detection; CNN, Convolutional Neural Network; RNN, Recurrent Neural Network.

I. INTRODUCTION

In recent trends, it has been predicted that the use of smart CCTV technology has been rapidly increased to judge the current circumstances and it is immediately informed to the administrator to take any action for security reasons. CCTV has been identified as a "situation of interest" [1] where it feeds all the records that have to be documented for further investigations. Nowadays CCTV has been installed in most of the public places where it records thousands of scenes, especially in crowded areas which become more peculiar to gather particular information from the very large video database [2]. For crime investigation, this CCTV footage is used to suspect a guilty scene, where the examiner have to view all the shots and scenes to spot the frames. In these cases they have to spend more time to view all the scenes without missing any evidences such as motion detection, facial identification and face detection. Also the obstacles arises with poor image resolution from the footages results more stress to detect the scenes. The survey on CCTV footage reveals that the stress and time management are very weird with low resolution video dataset. To avoid such distractions CBVR is one of the best of all the methods to analyze those video databases to recover the particular content from large collections through shots and scenes [3]. CBVR method is incorporated in two different ways: (a) Video segmentation. (b) Key Frame Selection:

Video segmentation: The video is fragmented into shots by the feature extraction method related to Color, texture, shape, and movement determines the Shot Boundary Detection

Key Frame Selection: The shot constitutes the frames in which the selection of frames using the frame extraction method is affiliated to Motion-based, Content-based, and reference-based to gather the beneficial frames referred to key-frame.

Then the implementations of retrieving the required video sequences are indexed [4] to produce the video summarization. The advantage of CBVR is to pre-process all the frames even in a low resolution pixel such as CCTV footages to sustain better results.

This paper is configured as follows. The following section II addresses about the CBVR Structure, section III describes about Shot Boundary Detection which depict the shots from the frames through Histogram Analysis of various methods, section IV explains the survive methods of Key Frame Extraction with its advantages and disadvantages, section V explains the subsist approach of deep learning techniques and its methods, section VI elaborates the various datasets used in key frame extraction to identify the meaningful key frames, section VII is concluded with the observations.

II. CBVR STRUCTURE

The CBVR structure of CCTV Surveillance involves the following steps as depicted in Fig. 1.

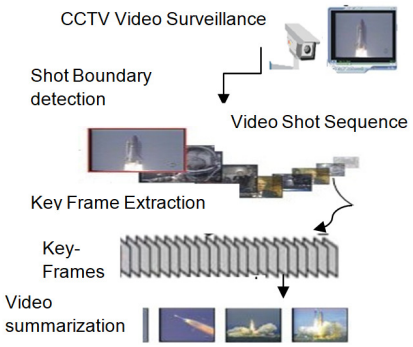

Fig. 1. Framework of CBVR.

(a) Input Data: CCTV footages are taken as input video sequences to convert the video sequence into frames using CBVR methods.
(b) Detecting the Shot boundary: The cuts and the gradual emergence constructs the shots from the frames are referred to SBD.
(c) Key frame Extraction: The shots are represented as frames. One or several frames are extracted to form a particular frame by eliminating and circumvent the redundancy. This is referred as Key Frame Extraction.
(d) Video Summarization: The extracted Key-frames are evaluated to produce a summarized output video. These retrieved video contents are finally evaluated by its precision and accuracy level [5].
Thus the Shot Boundary Detection and Key frame Extraction of CBVR methods and techniques redeem the required content from the CCTV footages.

III. SHOT BOUNDARY DETECTION

Shot Boundaries can be detected through the dissimilarity that occurred due to transition. Identification of the transition between consecutive shots is referred as Shot Boundary Detection. This shot transition [6] is classified into two types: abrupt (determined as Hard-cut) and gradual (determined as dissolve, fade in, fade out, wipe).
Abrupt can be easily determined by a sudden discontinuation between the frames. Kundu & Mondal [7] used conventional pixel based comparison between two consecutive frames to determine the abrupt change made between the frames using the evaluation of statistical measures. Pal *et al.*, [8] identified that the moderate change of brightness that occurs between two consecutive video frames which is referred as Fade. In case, if the pixels of the first shot and the second shots are replaced in a sequential way then the transition is referred as wipe. If the first shot is enhanced more than the second shot then it referred as Dissolve.

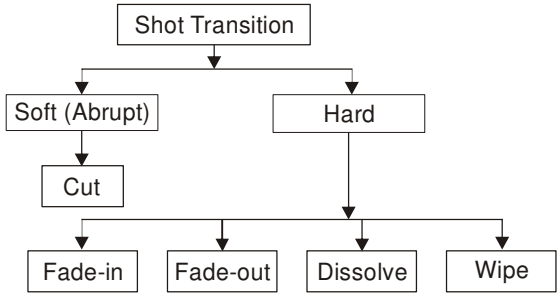

Fig. 2. Shot Transition Methods.

A. Shot Classification Methods
Apart from transition, the SBD between frames is determined through Spatio-temporal domain feature extraction.
Spatial Domain Feature [6]: Feature extraction methods are performed on the same size shaped objects in the frames to identify the discontinuities. Another method to determine the spatial domain feature is applying the luminance extraction on every pixel in the frame through which the shot change can be easily detected. One of the disadvantages is identifying the difference between two homogenous shots through the spatial domain is difficult.

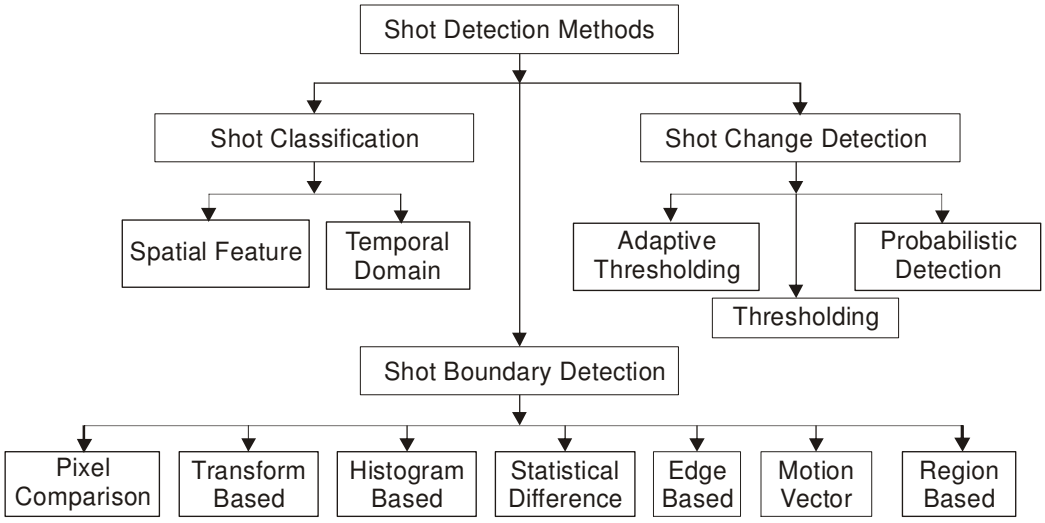

Fig. 3. Classification of Shot Detection Methods.

Temporal Domain Feature [6]: The problem to identify the dissimilarity between frames by predicting the highest metric value which is adjacent is said to be temporal. The metric value can be calculated using dynamic thresholding to recognize the discontinuity between frames easily. It can also be calculated using the statistical measures where the whole shot has to be taken and compared with the next consecutive frame to identify the discontinuity. By using various methods along with the temporal domain, the feature is extracted for the whole video to determine the change in the shots.

B. Shot Change detection

(a) Static Thresholding: This measures the difference between the frames by assigning a constant threshold value. If the predicted value is greater than the threshold value then the frame dissimilarity is identified easily by using SBD measures.

(b) Adaptive Thresholding: The Adaptive Threshold method follows the problem occurred in static thresholding and predicts the value with the statistic measurement to gain more accuracy.

(c) Probabilistic Detection: Probabilistic Detection is one of the tremendous ways to detect the shot change in which a particular pattern is generated and compared with the frames to obtain absolute result [9].

C. Various Shot Difference Measurement

(a) Pixel Comparison: Pixel values generated by each shot are compared to predict the difference between two successive shots. Yong *et al.,* [10] referred Pixel Comparison as template matching where it predicts the difference in intensity value of successive frames for both color and gray level images in the shots. Patel *et al.*, [11] explains that the adaptive thresholding method along with the pixel comparison results the difference in gradual boundaries for uncompressed video which is better than that for compressed video.

(b) Transform-Based: Kekre *et al.*, [12] has proposed the feature extraction through Transform-based by taking the higher coefficients of Discrete Cosine, Walsh transform, Haar, Kekre Transform, Discrete Sine transform, Slant transform, and Discrete Hartley transform methods are compared to minimize the vector size using fractional Co-efficient of frames to obtain the shot boundary.

(c) Edge-Based: This method is used to minimize the problem of undeviating shots formed due to interpretation and motion through edges obtained in each shot. Adjeroh *et al.,* [13] made a clear output of the difference in shots by using adaptive threshold detection on multi-level edge partitioning to make a time consuming on motion estimation. Zabih *et al.,* [14] also used video indexing to exhibits the exit and entry level of edges to classify the difference in shots.

(d) Motion Vector: The video frames differ by single changes with time especially through the movement of certain objects from one place to another. Volkmer *et al.,* [15] analyzed one of the techniques as the motion capture of pre-frame to the current frame along with the post frame that denotes a gradual change through the measures of adaptive thresholding method which results better.

(e) Region-Based: The region of interest to detect a particular object is implemented in this method. Russakoff *et al.*, [16] proposed a new method for the collection of common information through region mutual information for medical analysis. This method collects the nearby regions of the proportionate pixels that lead to the similarity between the frames.

(f) Statistical Difference: The Statistical difference is calculating the pixel difference occurred through color, region, edge, etc. Liu & Chen [17] have proved that the feature extraction of shot boundary through the statistical modeling model via Eigen-values is better when compared to direct differencing method. This method is useful in detecting the frame change in facial expression accurately.

(g) Histogram-Based: The shot transition detection can be easily identified by the difference between the histograms of the two consecutive frames. Also, the difference in color and gray levels of consecutive frames can be determined using Histogram prediction.

D. Performance Analysis of Shot Boundary Detection

The following three classifiers denominate the cut detection from one shot to another.

Recall and Precision exhibits the field of retrieving the information. Some complication arises between algorithms to detect the boundary. Recall and Precision are the best solution to determine the boundary effortlessly [18].

Recall: It exhibits the existing cuts of the shots. Here the recall betrays the percentage of how many Shot boundary were detected correctly with the number of shot boundary missed [18].

$$\text{Recall} = \frac{C}{C + D}$$

Precision: It represent the probability of assumed cuts are considered as shots. Here precision betrays the percentage of number of Shot boundary detected correctly with the number of shot boundary detected falsely [18].

$$\text{Precision} = \frac{C}{C + F}$$

F1 is the combined measure of both precision and recall only if the resulting value is high,

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where C, number of correctly detected shots, D, number of not detected shots, and F, number of falsely detected shots. The experimental analysis of various methods of SBD using the histogram based methods results the recall and precision value using the above metrics has been manifested in Table 1. If the trace difference between two histograms found to be greater than the threshold value then the shot boundary cuts can be easily determined.

Table 1: Experimental Results of Shot Boundary Detection using Various Methods.

Histogram based Method Evaluation	Reference/Author	Frames	Shots (Cuts/Gradual)	X=Precision (%)	Y=Recall (%)
Statistical Based	Liu *et al.*,(1998) [17]	Not mentioned	Not mentioned	83	23
Motion Vector & Pixel Based	Yong *et al.*, (2002) [10]	10000	70	75	88.73
Color & Edge Based	Ling *et al.*, (2008) [18]	89211	481	86.2	91.1
Region Based & Pixel Based	Engin & Bayrak (2010) [19]	2500	18	100	92.5
Statistical Based Adaptive Threshold	Meine *et al.*, (2010) [20]	344	241	93.8	65.7
Motion Vector through B & P Frame	Chunmei *et al.*, (2012) [21]	6482	129	95.81	95.6
Pixel Based - Fisher Criterion	Zhang & Wang (2012) [22]	144737	783	89	87.75
Color & Pixel Based	Fu *et al.*, (2013) [23]	1410	10	90	90
Color & Pixel Based	Patel *et al.*, (2013) [11]	90000	974	94.74	96.3
Motion Based & Pixel Based	Lungisani *et al.*, (2017) [24]	637805	2463	88.75	69.1

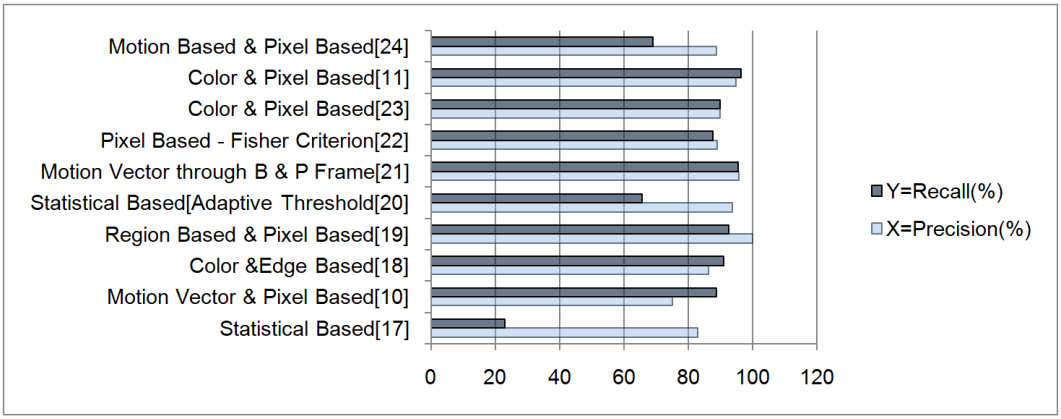

Fig. 4. Experimental results of various methods by Histogram Prediction.

Shot Boundary Detection using Histogram analysis with methods such as Region Based, Color based, Motion Based, Statistical Based, etc., has shown its excellence to determine the gradual and dissolve cuts from Fig. 4.

IV. KEY FRAME EXTRACTION

Representation of the video content through frames retrieved from the CCTV footages is represented as Key frame extraction. The CCTV camera has been installed at many places where it records thousands of scenes. To detect a crime scene, a huge amount of frames from the shots has to be analyzed through the frames that hold more data. All these frames contain eventful information hence it is referred as Key frames [25]. To retrieve a particular content from such a huge frame it can be compressed without the loss of information. At the same time reducing the redundancy in frames provide a better performance [26]. By analyzing the video with meaningful frames the results should be with the vigilant information for crime detection.

Key frames are generally classified as (i) Sequential Comparison-based (ii) Reference frame-based (iii) Global Comparison-based (iv) Clustering based (v) Curve simplification-based (vi) Object or event based [19].

1. Sequential Comparison-Based: In this comparison is made between the existing frame with the already extracted key frames in a sequential way resulting in the high position leads to form a new key frame.

As in movies most of the scenes are repeated to form the same set of key frames. Aoki *et al.*, [27] proposed a new method of identifying the redundancy in frames by using chromatic histogram where he classified the key frames as patterns and acts by which the redundancy has been reduced.

2. Global Comparison-based: This refers to a global difference between frames. It is classified as

(a) Even Temporal Variance: The temporal value of each shot from the frames is collected and their sum of dissimilarity of the variance computes the progressive change between the frames. Rautiaenen *et al.*, [28] identified temporal correlations at the gradient edges using Gradient Correlogram of the frames to reduce the redundancy.

(b) Minimum Correlation: To identify the repetition in the frames the sum of minimum correlation occurred between the frames is depicted through this method. A* algorithm was used by Zhang *et al.*, [29] for finding the shortest path of the vertices to construct the minimum correlation that occurred in each frame from which select the particular frame to avoid the redundancy.

(c) Minimum reconstruction error: This method evaluates the minimization of the excess frames through interpolation. Divakaran *et al.,* [30] suggested a Key Frame Selection with interpolating algorithm to identify the motion of each frame in the shots by which the maximum number of times of the same motion is recorded. This method results in the minimization of the repeated frames through the proposed algorithm.

3. Reference Frame-Based: This method explains the fact of having some reference to distinguish the common factor between the frames. Shi *et al.,* [31] mentioned two ways to predict the key frames.
One of the ways is using the color attribute which is used as a reference to predict the difference between the video sequences of the occurred frame and neighboring frames. Another reference is predicting the structural characteristics from one frame to another. From these two methods finally, the key frame is obtained.
4. Clustering: The Key frame is obtained by choosing the appropriate cluster center. The cluster center is explained by Lei *et al*., [32] where he used three types of filters to construct high dimensional features. The sparse transition matrix is used to obtain a low dimension feature from the shots. The unsupervised clustering method is applied to these shots to classify it into sub-shots. Finally using the Bhattacharya coefficient the cluster frame is identified from which the Key Frame has resulted.

5. Curve-simplification method: It proposes the Key frames present in the shot as by the point representation obtained from the curve. These curves are represented as a high valid complexity. Lim & Thalmann [33] implemented this method to establish the highest value of the curve obtained from the motion being captured in the video shots. The extreme point that occurred during the curve simplification method is represented as Key Frames.
6. Objects or Event – Based: These methods examine the object or event present in the given frame so that the extracted key frame contains valid information.

Table 2: Advantages and disadvantages of Key Frame Extraction methods [35].

Key Frame Classification Methods	Advantages	Disadvantages
Sequential Comparison-based.	Low computational complexity	Repeated comparison to reduce redundancy
Global Comparison-based	Extracted Key frames are controllable when compared to sequential	More computational complexity
Reference frame-Based	Easy implementation by placing some reference to represent a frame	Some of the references may be missed and it doesn't represent the shot appropriately.
Clustering based	Key frame through cluster centre forms a global feature.	High computation cost
Curve simplification-based	Best representation through curve	High computational complexity
Object or event based	Detection of object or event from a collection of frames is easier	The object has to be chosen appropriately else leads to misclassification.

Makandar & Mulimani [34] elaborates the object detection in sports such as Kabaddi game. Using the morphological operations such as elimination of background with the foreground highlights only the objects present in the frames. Applying the threshold method the object of the Kabaddi player is accurately classified. A comparative analysis of each key frame classification methods results a high computational complexity has its disadvantage which is a major drawback to move on to other competence such as deep learning.

V. DEEP LEARNING TECHNIQUES OF KEY FRAME EXTRACTION METHOD

Deep learning, a subset of machine learning, exploits the hierarchical level of ANN (artificial neural network) for the process of exhibiting the machine learning. Deep Learning method is implemented in many areas such as image recognition, video recognition, medical research tools, etc., to understand and process the major defeats. Video surveillance in many public areas has increased dramatically due to the wide cause of traffic management, crime detection or for making a view on crowded area and so on. These videos for an assumption consist of a frame of 25fps (frame per second) which leads to 3600*25 = 90000 frames for per hour video. From these frames, the SBD accords the shot frames. Key frame extraction is one of the methods to extract the key frame from these shot frames [36].
In, deep learning methods such as RNN(Recurrent Neural Network), LSTM (Long Short Term Memory), CNN(Convolutional Neural Network) are implemented for the extraction of key frame from huge dataset. Krizhevsky *et al*., [37] examined a huge data by deep convolutional neural network to estimate the contest of 1000 different classification from 1.2 million high resolution images. Speech and text recognition are also designated by the application of CNN [38]. Inoue *et al.,* proposed the identification of high-performance of semantic indexing system using deep CNN and GMM super vectors of audio and visual video shots [39].

VI. DATASET USED FOR KEY FRAME EXTRACTION

TRECVID 2007: TREC Video Retrieval Evaluation.
This dataset is sponsored by National Institute of Standards and Technology. The dataset provides a video track for the research of segmentation, indexing and Content Based video retrieval process.
MFC18: Media Forensics Challenge 2018.
These dataset contains the image/video data implemented for the detection of region and the types of correlation that is modified by the image/video data's.
P2ES5 & P2LS5: It consists of higher resolution pixel than the standard video data. High definition video holds the frames of 50-60 frames per seconds.
UCF 101: This is action recognition dataset of feasible action videos. It contains of about 101 action classes, over 13000 clips of about 27 hours videos. The most challenging is to describe the action sequence from those large numbers of clips for the researchers.

Table 3: Detection of Key Frames from Various Dataset.

Methods	Author	Dataset	Detection	Techniques	SBD	Key frame Extraction	
						Actual Number of Frames / Video clips Instances	Number of Key frames Detected
Deep Convolutional Neural Network (DCNN)	Sanglae *et al.*, [40]	TRECVID 2007	Video Concept	SVM(Support Vector Machine)	Edge- Based	111 video frames	180
	Long *et al.*, [41]	MFC18	Frame Duplication	C2F-DCNN	-	1036 videos(104 400 frames)	25200 duplicate frames
Convolutional Neural Network (CNN)	Khan *et al.*, [42]	Movie	Movie Tags Detection	Inception V3, Softmax Classification	HSV(Hue, Saturation, Value) Histogram	42	Predicted key frames as Exercise/ Drama/ Fashion
	Xuan *et al.*, [43]	Live Video Surveillance P2ES5	Face Detection	GPU(Graphic Process Unit)	—	898	57
		P2LS5				755	54
	Deshmukh *et al.*, [44]	UCF 101	Video Concept Detection	SVM(Support Vector Machine)	Sequential Based	537	205
	Qi *et al.*, [45]	HD Videos a. Corridor b. Pathway	Face Detection	—	—	a.292 b.1508	a.24 b.134
Others	Luo *et al.*, [46]	Video Surveillance a. Person boxing b. Laboratory raw c. Campus raw	Moving Object Detection	SURF(Speed-Up Robust Features)	—	a. 424 b. 2659 c. 3535	a. 3 b. 10 c. 16
	Clinton *et al.*, [47]	Sports Video	Frame Duplication	Unsupervised Clustering	Histogram Based	26640	1734
	Janwe Bhoyar [48]	Standard Video Library Set	Near Duplication	Unsupervised Clustering	—	384	7

VII. CONCLUSION

The comprehensive study on Content Based video retrieval has been reported in this paper. The experimental analysis of SBD through histogram analysis invoked the best result when compared to others to form the key frame Extraction. The merits and demerits of Key frame extraction expose its lack to represent a huge video database from the traditional methods. To conquer this fact, the latest approach is to classify the video sequence through deep learning methods which can process millions of videos in a very short time. One of the advantages of using the Deep-learning model is said to be a state-of-art accuracy. The extension of work on deep learning method for the extraction of key frame and shot boundary detection can give more precious results in a very huge dataset.

VIII. FUTURE SCOPE

In the future, the retrieval of required content over large video dataset using deep learning methods makes easier to analyze the video for crime investigations.

Conflict of Interest. The authors declare that there is no conflict of interest in this work.

REFERENCES

[1]. Sidhu, R. S., & Sharad, M. (2016). Smart surveillance system for detecting interpersonal crime. *2016 International Conference on Communication and Signal Processing (ICCSP).* https://doi.org/10.1109/iccsp.2016.7754524.

[2]. Yang, Y., Lovell, B. C., & Dadgostar, F. (2009). Content-Based Video Retrieval (CBVR) System for CCTV Surveillance Videos. *2009 Digital Image Computing: Techniques and Applications.* https://doi.org/10.1109/dicta.2009.36.

[3]. Aasif, M., & Vasishtha, H. (2015). Enhanced Video Retrieval and Classification of Video Database Using Multiple Frames Based on Texture Information. *International Journal of Computer Science and Information Technologies*, *6*(2), 1740-1745.

[4]. Goncalves, V., Santos Nunes, F. (2016). A Systematic Review on Content-Based Medical Video retrieval. *Journal of Health Informatics, 8*:799-808. https://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis.

[5]. Gornale, S. S., Babaleshwar, A.K., & Yannawa, P.L. (2019). Analysis and Detection of Content Based Retrieval. *Journal of Image, Graphics and Signal Processing*, *11*(3), 43-57.

[6]. Patel, D. H. (2015). Content Based Video Retrieval: A Survey. *International Journal of Computer Applications*, *109*(13), 1-5.
[7]. Kundu, M. K., & Mondal, J. (2012). A novel technique for automatic abrupt shot transition detection. *2012 International Conference on Communications, Devices and Intelligent Systems (CODIS).* https://doi.org/10.1109/codis.2012.6422281.
[8]. Pal, G., Rudrapaul, D., Acharjee, S., Ray, R., Chakraborty, S., & Dey, N. (2015). Video Shot Boundary Detection: A Review. *Advances in Intelligent Systens and Computing Emerging ICT for Bridging the Future – Proceedings of the 49th Annual Convention of the Computer Society of India CSI. 2*, 119-127. https://doi.org/10.1007/978-3-319-13731-5_14.
[9]. Cotsaces, C., Nikolaidis, N., & Pitas, I. (2006). Video Shot Boundary Detection and Condensed Representation. A Review. *IEEE Signal Processing Magazine*, *23*(2), 28-37. https://doi.org/10.1109/ msp.2006. 1621446.
[10]. Yong, C., Xu, Y., & De, X. (2002). A Method for Shot Boundary Detection with Automatic Threshold, *2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering. TENCOM' 02. Proceedings*. https://doi.org/10.1109 /tencon.2002. 1181342.
[11]. Patel, U., Shah, P., & Panchal, P. (2013). Shot Detection Using Pixel wise Difference with Adaptive Threshold and Color Histogram Method in Compressed and Uncompressed Video. *International Journal of Computer Applications, 64*(4), 38-44. https://doi.org/10.5120/10625-5347.
[12]. Kekre, H. B., Thepade, S. D., & Akshay, M. (2011). Comprehensive Performance Comparison of Cosine, Walsh, Haar, Kekre, Sine, Slant, and Hartley transforms of CBIR with Fractional Coefficients of Transformed Image. *International Journal of Image Processing (IJIP), 5*(3), 336-351.
[13]. Adjeroh, D., Lee, M.C., Banda, N., & Kandaswamy, U. (2009). Adaptive Edge Oriented Shot Boundary Detection. *EURASIP Journal on Image and Video Processing*, 1-13. https://doi.org/10.1155/2009/859371.
[14]. Zabih, R., Miller, J., & Mai, K. (1999). A Feature-Based Algorithm for Detecting and Classifying Scene Breaks. *Multimedia Systems, 7*(2), 119-128. https://doi.org/10.1007/s005300050115,
[15]. Volkmer, T., Tahaghoghi, S. M. M., Williams, H. E., & Thom, J. A. (2003). The Moving Query Window for Shot Boundary Detection. *Proceedings of the TREC Video Retrieval Evaluation (TRECVID) Workshop. Gaithersburg, MD, USA,* 147-156.
[16]. Russakoff, D.B., Tomasi, C., Rohlfing, T., & Maurer, C.R. (2004). Image Similarity Using Mutual Information of Regions. *Lecture Notes in Computer Science Computer Vision-ECCV 2004,* 596-607. https://doi.org/10.1007/978-3-540-24672-5_47.
[17]. Liu, X., & Chen, T. (2002). Shot Boundary Detection using temporal statistics modeling. *IEEE International Conference on Acoustics Speech and Signal Processing.* https://doi.org/10.1109/icassp.2002.5745381.
[18]. Ling, X., Chao, L., Huan, L., & Zhang, X. (2008). A General Method for Shot Boundary Detection. *2008 International Conference on Multimedia and Ubiquitous engineering (mue 2008).* https://doi.org/10.1109/mue.2008.102.
[19]. Engin, M., & Bayrak, C. (2010). Shot Boundary Detection and Key Frame Extraction Using Salient Region Detection and Structural Similarity. *Proceedings of the 48th Annual Southeast Regional Conference on ACM SE'10.* https://doi.org/10.1145/1900008.1900096.
[20]. Meine, A., Hermes, T., Loannidis, G., Fathi, R., & Herzog, O. (2003). Automatic Shot Boundary Detection and Classification of Indoor and Outdoor Scenes. *Information Technology: The 11th Text Retrieval Conferences*.
[21]. Chunmei, M., Changyan, D., & Baogui, H. (2012). A New Method for Shot Boundary Detection. *2012 International Conference on Industrial Control and Electronics Engineering.* https://doi.org/10.1109/icicee.2012.49.
[22]. Zhang, C., & Wang, W. (2012). A Robust and Efficient Shot Boundary Detection approach based on Fisher Criterion. *Proceedings of the 20th ACM International Conference on Multimedia-MM'12* https://doi.org/10.1145/2393347.2396291.
[23]. Fu, Q., Zhang, Y., Xu, L., & Li, H. (2013). A Method of Shot Boundary Detection based on HSV space. *2013 Ninth International Conference on Computational Intelligence and Security.* https://doi.org/10.1109/cis.2013.53.
[24]. Lungisani, B., Thuma, E., and Malema, G., (2017). A Classification Approaches to Video Shot Boundary Detection. *International Journal of Signal Processing, Image Processing and Pattern Recognition, 10*(12), 103-118. http://doi.org/10.14257/ijsip. 2017.10.12.08.
[25]. Sahu, K., & Verma, S. (2017). Key Frame Extraction from Video Sequence: A Survey. *International Research Journal of Engineering and Technologies (IRJET), 4*(5): 1346-1350. https://www.irjet.net.
[26]. Patil, P. V., & Bodake, S. V. (2015). Video Retrieval by Extracting Key Frames in CBVR System. *International Journal of Innovative Research in Science, Engineering and Technology, 4*(12), 10121-10128. http://doi.org/10.15680/IJIRCCE.2017.0505320
[27]. Aoki, H., Shimotsuji, S., & Hori, O. (1996). A Shot Classification Method of Selecting Effective key-frame for Video Browsing. *Proceedings of the Fourth ACM International Conference on Multimedia–MULTIMEDIA'96.*
http://doi.org/10.1145/244130.244135.
[28]. Rautiaenen, M., Seppanen, T., Penttila, J., & Peltola, J. (2003). Detecting Semantic Concepts from Video Using Temporal Gradients and Audio Classification. *Lecture Notes in Computer Science Image and Video Retrieval*, 260-270. http://c10.1007/3-540-45113-7_26.
[29]. Zhang, H. J., Wu, J., Zhong, D., & Smoliar, S. W. (1997). An integrated system for content-based video retrieval and browsing. *Pattern Recognition*. Vol. *30*(4): 648-658. http://doi.org/10.1016/s0031-3203 (96)00109-4.
[30]. Divakaran, A., Radhakrishnan, R., & Peker, K. A. (2002). Motion Activity-based Extraction of Key-frames from Video Shots. *Proceedings International*

Conference on Image Processing. http://doi.org/10.1109/icip.2002.1038180.
[31]. Shi, Y., Yang, H., Gong, M., Liu, X., & Xia, Y. (2017). A Fast and Robust Key Frame Extraction Method for Video Copyright protection. *Journal of Electrical and Computer Engineering*, 1-7. https://doi.org/10.1155/2017/1231794.
[32]. Lei, P., Xin, S., & Zhang, M. (2015). A Key Frame Extraction Algorithm Based on Clustering and Compressive Sensing. *International Journal of Multimedia and Ubiquitous Engineering*, *10*(11), 385-396. http://dx.doi.org/10.14257/ijmue.2015.10.11.37.
[33]. Lim, S.L., & Thalmann, D. (2001). Key posture Extraction out of Human Motion Data. *2001 Conference Proceedings of the 23rd Annual International Conference of IEEE Engineering in Medicine and Biology Society.* https://doi.org/10.1109/iembs.2001.1020399.
[34]. Makandar, A., & Mulimani, D. (2016). Key Frame Extraction and Object Detection in the Sports Video. *International Conference on Recent Trends in Engineering, Science and Technology (ICRTEST). IET Inspec*.
[35]. Ali, I. H., & Al-Fatlawi, T. (2018). Key Frame Extraction Methods. *International Journal of Pure and Applied Mathematics. 119*(10-C), 485-490.
[36]. Wang, W., & Farid, H., (2007). Exposing digital forgeries in video by detecting duplication. *Proceedings of the 9th Workshop on Multimedia &Security – MM & Sec '07.* http://doi.org/10.1145/1288869.1288876
[37]. Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). Image Net Classification with Deep Convolutional Neural Network. *Advances in Neural Information Processing Systems.* 1097-1105. University of Toronto.
[38]. Bhandare, A., Bhide, M., Gokhale, P., & Chandavarkar, R. [2016]. Applications of Convolutional Neural Network. *International Journal of Computer Science and Information Technologies (IJCSIT).* Vol. *7*(5), 2206-2215. http://www.ijcsit.com.
[39]. Inoue, N., Shinoda, K., & Xuefeng, Z and Ueki K., (2014). Semantic Indexing using Deep CNN and GMM Supervectors, *Tokyo Institute of Technology Waseda University.*
[40]. Sanglae, R. D., Patil, N. S., & Sawarkar, S. D. (2019). Video Concept Detection Using Convolutional Neural Network. *International Journal Of Engineering Research and Development*, *15*(1), 79-86. http://www.ijerd.com.
[41]. Long, C., Basharat, A., & Hoogs, A. (2019). A Coarse to fine Deep Convolutional Neural Network Framework for Frame Duplication and Localization in Forged Videos. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).*
[42]. Khan, U. A., Martinez-del-Amor, M. A., Altowaijri, S.M., Ahmed, A., Rahman, A.U., Sama, N. U., Haseeb, K., & Islam, N. (2020). Movie Tags Prediction and Segmentation Using Deep Learning. *IEEE Access*. Vol. *8*: 6071-6086. https://doi.org/ 10.1109/access.2019.2963535.
[43]. Xuan Q., Liu, C., Schukers, S. (2018). CNN Based Key Frame Extraction for Face in Video Recognition. *IEEE 4th International Conference on Identity, Security and Behavior Analysis(ISBA)*. 978-1-5386-2248-3/18. Clarkson University.
[44]. Deshmukh, J.J., Patil, N.S., & Sawarkar, S.D. (2019). Video Concept Detection Using SVM and CNN. *International Journal of Engineering Development and Research, 7* (3). ISSN: 2321-9939.
[45]. Qi. X., Liu, C., & Schukers, S. (2018). Boosting Face in Video Recognition via CNN Based KeyFrame Extraction. *2018 International Conference on Biometrics (ICB).* https://doi.org/ 10.1109/icb2018.2018.00030.
[46]. Luo, Y., Zhou, H., Tan, Q., Chen, X., & Yun, M. (2017). Key Frame Extraction of Surveillance Video Based on Moving Object Detection and Image Similarity. *Pattern Recognition and Image Analysis. 28*(2), 225-231.
[47]. Clinton, P., Rao, N. R., & Rani, N. S. (2018). Key Frame Extraction from Video – A Comparative Study. *Journal of Advanced Research in Dynamical and Control Systems*, *10*(12), ISSN: 1943-023X.
[48]. Janwe, N. J., & Bhoyar, K. K. (2016). Video Key-frame Extraction using Unsupervised Clustering and Mutual Comparison. *International Journal of Image Processing, 10*(2), 73-84.

How to cite this article: Priscilla, C. V. and Rajeshwari, D. (2020). Perspective Study on Content Based Video Retrieval. *International Journal on Emerging Technologies*, *11*(2): 205–212.