An Enhanced Spatio-Temporal Human Detected Keyframe Extraction

Original Scientific Paper

Rajeshwari D
Research Department of Computer Science
Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women,
Affiliated to University of Madras, Chennai, India.
rajeshwari.d@sdnbvc.edu.in

Victoria Priscilla C
PG Department of Computer Science,
Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women,
Affiliated to University of Madras, Chennai, India.
aprofvictoria@gmail.com

Abstract *– Due to the immense availability of Closed-Circuit Television surveillance, it is quite difficult for crime investigation due to its huge storage and complex background. Content-based video retrieval is an excellent method to identify the best Keyframes from these surveillance videos. As the crime surveillance reports numerous action scenes, the existing keyframe extraction is not exemplary. At this point, the Spatio-temporal Histogram of Oriented Gradients - Support Vector Machine feature method with the combination of Background Subtraction is appended over the recovered crime video to highlight the human presence in surveillance frames. Additionally, the Visual Geometry Group trains these frames for the classification report of human-detected frames. These detected frames are processed to extract the keyframe by manipulating an inter-frame difference with its threshold value to favor the requisite human-detected keyframes. Thus, the experimental results of HOG-SVM illustrate a compression ratio of 98.54%, which is preferable to the proposed work's compression ratio of 98.71%, which supports the criminal investigation.*

Keywords*: Histogram of Oriented Gradients-Support Vector Machine, Keyframe Extraction, Spatio-temporal feature Extraction, Content-Based Video Retrieval*

1. INTRODUCTION

The use of Closed-circuit television (CCTV) surveillance for specific safety measures has increased incrementally in the majority of public areas in recent years. Surveillance plays an essential part in crime scene investigation by actively monitoring the circumstances inside a specific, stationary region. In the field of investigation, many investigators still struggle to identify the victim in the cases. Here are some of the most common issues, such as (1). Videos of poor quality (2) Videos with low frame rates lose detail between frames. (3). Analyzing and evaluating larger datasets in videos requires a significant amount of time and effort by the investigators.

Content-Based Video Retrieval (CBVR) is widely regarded as a crucial step in video analysis and Key frame extraction. It retrieves the desired video from a massive video storage database. Keyframe Extraction is the process of extracting a significant segment of a video by exploring its content to generate a condensed and semantically rich summary.

Moreover, if these keyframes for crime investigation reports are highlighted with humans, it is easier to suspect those responsible for the crime. To efficiently quote the sequences, it is necessary to determine an algorithm for human-detected keyframes in particular.

The Histogram of Gradients-Support Vector Machine (HOG-SVM) approach can identify people in the surveillance footage, although it occasionally fails in certain frames. As a result, the suggested work uses HOG-SVM with background subtraction to report human detection in all pertinent frames. Additionally, a Visual Geometry Graph (VGG-16) pre-trained these frames for the categorization report of human-detected frames. Finally, the frames (images) are pre-processed with the Canny-Edge detection method for enhanced structural information, and the desired Keyframes are extracted using the inter-frame difference method with its threshold value. These Keyframes play a crucial role in the investigation of crimes by substantially reducing the temporal and spatial complexities of the process.

The documentation is systematically structured as outlined below: Section 2 provides a concise summary of the current study on CBVR with human motion recognition and keyframe extraction techniques. In Section 3, the recommended approach of employing the HOG-SVM technique along with background subtraction is discussed in detail. The details regarding the implementation and the experimental findings can be found in Section 4, while the conclusion is provided in Section 5.

2. RELATED WORK

This literature probes the study of detecting humans through various algorithms. Consigning humans to other existing objects is very complicated in CCTV surveillance. Also, a person prolongs a long or short stay in a place to represent a certain action [1].

In most instances, humans are identified by their motion. Currently, the frame subtraction method, the background subtraction method, and the optical flow method are the most frequently utilized techniques for motion detection.

Optical Flow: This method observes the moving object based on its maximal frame-to-frame deviation. Identifying human motion in a video stream using the optical flow method requires a great deal of computational time [2].

Background Subtraction: This method attempts to encapsulate information regarding background scene changes concerning the video frame sequence [3]. There are various methods for performing background subtraction. The most common approaches are (a) Adaptive Gaussian mixture, which uses motion analysis to distinguish the foreground from the complex background [4], (b) Kalman filter, which is used to enhance image quality through background elimination [5], (c) Temporal differencing, which uses pixels to calibrate the motion detection on the foreground [6], and (d) Clustering techniques, which look at groups of pixels that are similar [7]. This method merits high accuracy but demerits to have a static background. Frame Difference: The moving object is identified efficiently at a complex background by taking the difference between the two frames [8, 9] but reports with less accuracy.

The motion detection phase in the video can also be detected using combination approaches such as background subtraction with the optical flow. This reduces the noise effect and eliminates the shadow present in the frames [10, 11]. Another combination is background subtraction with the frame difference method. This combination's main advantage results in the fast elimination of shadows [12] and inexpensive detection of frames [13]. Thus, this combination supports speculating the appropriate motion detection phase to extract keyframes from the surveillance video.

Keyframe Extraction refers to the video's summary because it removes redundant frames and provides only the video's essential content. It is a probabilistic task to extract keyframes from video footage containing massive amounts of data [8]. Many scholars have classified keyframe extraction techniques using Shot boundary, Motion Analysis, Visually Segmented, and cluster-based analysis as depicted in Table 1

Table 1. Existing Methods & Techniques for Keyframe Extraction

METHODS & TECHNIQUES	ACCURACY & MERITS	DEMERITS
Shot-boundary Detection		
SIFT-point distribution Histogram [14]	94.36% accuracy with less computation	Selecting only the Salient segment from each segmented shot
Middle Range Binary Local Pattern (MRLBP) [15]	96.34% with high entropy measures	Only Abrupt shot boundary detection is performed
Adaptive Threshold [16]	91.93% with less computation	Less performance due to blurred frames.
SVD Pattern Matching [17]	85.5% with high detection speed	Less precision value for gradual detection
Hadamard Transform [18]	88.7% based on significant feature	Less accuracy level at gradual transition
Genetic Algorithm and fuzzy logic [19]	86.8% with increased iterations	Time Complexity is high when iteration increases
Multimodal techniques [20]	88.7% with the selection of candidate segment	Speed is not detected and gradual detection has to be improved
Motion-Analysis		
Discrete cosine coefficients and rough sets theory [21]	82% for visual representation	Enormous Space Complexity
Thresholding technique [22]	81% based on threshold value	Using Key-object to analyze with less precision
Color and Structure Based [23]	86% with high computation	Poor performance on complex transitions
Perceived Motion Energy Mode [24]	80% with motion and color based	Requires improvement in color variation
Convolutional Neural Network [25]	92% with improved frame difference method	High computation time
Visually Segmented		
Region of Interest-KNN, SVM [26]	90% of motion detection by pixel change	Concentrated more on noise reduction
Multiple Feature Analysis [27]	80% of motion detection by pixel-level classification	Performance at the static background
Region Of Interest-FCN with CNN [28]	97% of detecting multiple objects	Less Performance in crowded areas
Cluster-Based		
Weighted Multi-View Cluster [29]	81.53% for medoid frames	Fails to report the number of clusters
Dynamic Spatio-Temporal Slice Clustering [30]	92.68% with high accuracy	Proposed only on human action video dataset

The primary result of these methodologies clarifies the applicability of distinguishing objects and identifying events in keyframes with an appropriate level of complexity. This research proposes a faster, more accurate, and computationally efficient strategy for Video Keyframe Extraction.

3. PROPOSED METHODOLOGY

The proposed approach's general framework (Figure 1) consists of four sequential steps: (1) Pre-processing the video. (2) Human motion detection using Background Subtraction and Frame Difference method. (3) Spatio-temporal feature extraction - HOG-SVM. (4) VGG-16 pre-trained CNN and (5) the Keyframe extraction using Threshold value along with Canny Edge Detection Method.

3.1. VIDEO PRE-PROCESSING

The recorded CCTV surveillance footage is in the initial stage of pre-processing. In this phase, the video footage endures a conversion to gray scale and is also resized to 640*480 for faster detection.

3.2. BACKGROUND SUBTRACTION TECHNIQUE

Background subtraction (Equation 1) is widely used for motion (Human) detection in video surveillance of static cameras. The detection of motion is achieved by calculating the disparity between the present frame and the reference frame [32].

$$|Frame_i\text{-}Background_i|>Threshold \qquad (1)$$

Whereas the frame difference method (Equation 2) calibrates the difference between the two frames by the pixel variation.

$$|Frame_{i-1}\text{-}Frame_i|>Threshold \qquad (2)$$

The background subtraction and frame difference algorithms' discrete performance are subject to false detection. To overcome it, the combination of background subtraction and frame difference assists surveillance video to detect motion more accurately.

Converting the video to gray scale frames simplifies the background subtraction process and facilitates the detection of humans. Calibration is performed by capturing the non-moving pixel specks in the first frame. If the pixel has changed in the subsequent frame, motion is detected. Then, these frames are subjected to the frame difference method, which identifies differencing structures to eradicate redundant frames. Still, the researchers have some limitations, as mentioned in Table 2.

Table 2. Merits and Demerits of Combination Factors

METHODS	ADVANTAGES	DISADVANTAGES
Background subtraction with frame difference using Running Gaussian Average [13]	Shadows are more efficiently removed	Once motionless, the whole part is considered background
Background with frame difference [33]	Eliminate the noise efficiently	Represent video with static background
Background and consecutive frame difference method [34]	Efficient method for surveillance datasets	The Dynamic background is not supported
Background subtraction with frame difference [11]	Rectangular contour for moving objects with noise elimination	Too many detections of moving objects
Background subtraction and frame difference using correlation coefficient [35]	Highly correlated with background image for speed and detection accuracy	The Shape and edge on each frame have to be concentrated more.
Background subtraction using pixel intensity [36]	Deduction of the person by pixel change	The speed of the process is slightly slow

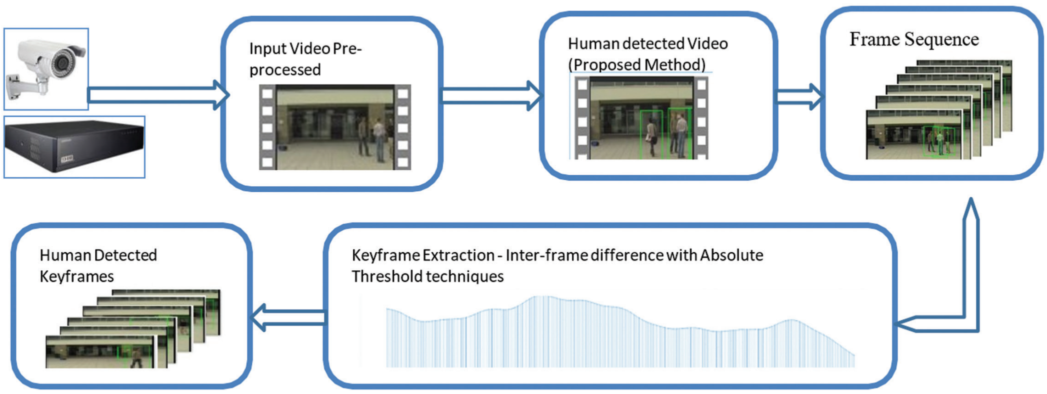


Fig. 1. The Overall Framework of Proposed Approach

3.3. SPATIO-TEMPORAL FEATURE EXTRACTION

3.3.1. Histogram Of Oriented Gradients

The Spatio-temporal feature extraction method supports human detection techniques using HOG, which was developed by Dalal and Trigg [31]. HOG represents the human shape and regional appearance based on the local histograms of image gradients in a dense grid. Here, the selected frame is partitioned into a small, connected area called cells. These cells contain several pixels, which unite to make a histogram of gradients. The computed gradients from the detector window are tiled like a grid of overlapped blocks, in which the HOG is extracted with normalized cells. The normalized cells give better accuracy on the variation through illumination and intensity.

3.3.2. Support Vector Machine

Support Vector Machine (SVM), is a supervised Machine Learning Algorithm that represents the most accurate image classification. Here, the resultant descriptors are fed into the linear SVM [37] for human/non-human classification.

3.4. VGG-16 CONVOLUTIONAL NEURAL NETWORK

In this proposal, the human-detected frames are trained by the VGG-16 pre-trained CNN (convolutional neural network) [38] model. Using a multi-class classification problem, the frames are categorized into three classes: 0 for False human predicted (FHP), 1 for True human predicted (THP), and 2 for Without human identification (WH) frames. All of these frames are resized so that the input image has dimensions of 224*224*3 and then sent to the input layer. The concealed layer is then convoluted three times with a dropout of 0.5, and the output layer is established using Softmax. By compiling the model with Adam optimizer, the accuracy reported for human-detected HOG-SVM in Table 3 and for the proposed work in Table 4 is significantly improved.

Table 3. VGG-16 trained HOG-SVM Human Detected Frames

Human detection using HOG-SVM						
Surveillance dataset	Detection	Total frames	Precision	Recall	F1-score	Accuracy
CCTV1	FHP	520	0.94	0.94	0.94	98.33
	THP		0.99	0.99	0.99	
	WH		1.00	1.00	1.00	
CCTV2	FHP	579	1.00	0.53	0.69	97.38
	THP		0.95	1.00	0.97	
	WH		0.00	0.00	0.00	
CCTV3	FHP	643	1.00	0.79	0.88	98.50
	THP		0.98	0.99	0.99	
	WH		0.98	1.00	0.99	
CCTV4	FHP	629	0.58	0.54	0.56	98.08
	THP		0.97	0.97	0.97	
	WH		0.00	0.00	0.00	
CCTV5	FHP	584	0.91	1.00	0.95	99.18
	THP		1.00	0.93	0.96	
	WH		0.99	1.00	1.00	

Table 4. VGG-16 trained Human Detected Frames for Proposed Work

Human detection using HOG-SVM						
Surveillance dataset	Detection	Total frames	Precision	Recall	F1-score	Accuracy
CCTV1	FHP	435	0.95	0.83	0.88	98.33
	THP		0.99	0.99	0.99	
	WH		0.73	1.00	0.84	
CCTV2	FHP	537	0.83	0.56	0.67	98.80
	THP		0.97	0.99	0.98	
	WH		0.00	0.00	0.00	
CCTV3	FHP	580	0.96	1.00	0.98	98.61
	THP		1.00	0.98	0.99	
	WH		1.00	1.00	1.00	
CCTV4	FHP	629	0.58	0.54	0.56	98.08
	THP		0.97	0.97	0.97	
	WH		0.00	0.00	0.00	
CCTV5	FHP	534	0.95	0.95	0.95	99.21
	THP		0.97	0.97	0.97	
	WH		1.00	1.00	1.00	

3.5. KEYFRAME EXTRACTION

After the preceding stages have been completed, the frames are fine-tuned using a Canny-edge detector to obtain a clear image. Now, the keyframe must be extracted from frames that differ significantly from one another. The average inter-frame difference greater than the threshold value, as calculated by equation (3), yields the keyframes.

$$|Average\ Inter\text{ - }frame\ difference|=0.6 \tag{3}$$

Here, the proposed method is combined with the threshold range to identify the ideal human-detected keyframes, as shown in Fig. 2.

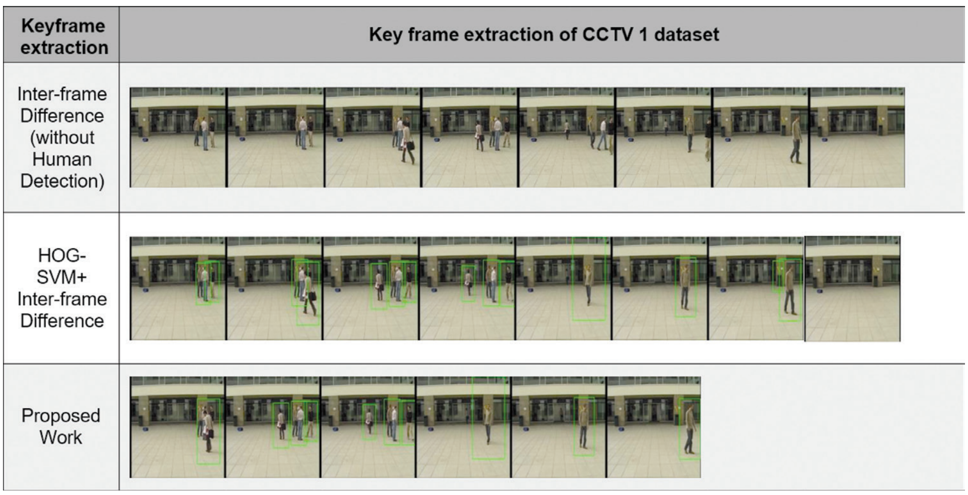


Fig. 2. Performance Analysis of Keyframe Extraction

For instance, the first CCTV1 surveillance system proposal included 435 video frames, from which 6 keyframes were extracted. The keyframes keyframe_99, keyframe_144, keyframe_178, keyframe_284, keyframe-336, and keyframe_375 are chosen based on their abrupt pixel change and difference from the overall frames. As depicted in Fig. 2, the performance evaluation of HOG-SVM with background subtraction reveals a reduction from 8 to 6 keyframes. The proposed task, in contrast, extracts only the required keyframes with perfect human detection. Consequently, the proposed result of HOG- SVM, along with background subtraction and inter-frame differences with the necessary threshold values, demonstrates the most accurate detection.

4. RESULTS AND DISCUSSION

The proposed task is carried out using Python Open CV image processing. The performance measurement derived from the obtained frames resulting in the keyframes is processed for the evaluation of metrics such as average frame per second (Equation 4) and frame per second (Equation 5).

4.1. AVERAGE FRAME PER SECOND

$$Avg\ FPS = (Total\ frames\ per\ second)/(Current\ frame) \qquad (4)$$

The average frame per second is determined by comparing the total frames per second to the current frame's frame rate. The frames per second are the unit of measurement for the video's performance.

4.2. FRAMES PER SECOND

$$FPS = 1/((end\ time\text{-}start\ time)) \qquad (5)$$

The frame rate is the number of frames displayed every second. In this instance, the average frame rate of CCTV 4 and CCTV 5 in the study under consideration exhibits an increase in Table 6 relative to Table 5. This may be attributed to the utilization of densely annotated surveillance footage, which facilitates the discovery of optimal keyframes that accurately show human activity.

4.3. COMPRESSION RATIO

This is used to determine the compression level achieved by the keyframes depicted in the video sequence. (Equation 6).

$$CR = 1\text{-}\{N_k / N_f\} * 100 \qquad (6)$$

Where N_k represents the number of extracted keyframes and N_f represents the total number of frames obtained.

4.4. PRECISION

This reveals the extraction accuracy, which is used to analyze the actual keyframe extracted (Equation 7).

$$Precision = N_c / (N_c + N_f) * 100\% \qquad (7)$$

Here N_c refers number of human-detected frames and N_f with total frames obtained.

4.5. RECALL

A sensitivity producer reveals the relationship between the obtained keyframe extractions to that of the actual number of required keyframes (Equation 8).

$$Recall = N_c / (N_c + N_m) * 100\% \qquad (8)$$

Here N_c is the number of human-detected frames, and N_m is the number of human-detected frames that were not detected.

Table 5. Accuracy Determination for Human Detected Keyframes using HOG-SVM

Surveillance dataset	Avg. Fps	Frames	HOG- SVM key frame extraction	Precision	Recall	CR
CCTV1	6.892	520	8	85.61	87.50	98.462
CCTV2	6.806	579	7	99.36	93.83	98.791
CCTV3	7.043	643	10	77.02	90.78	98.445
CCTV4	6.718	629	9	100	93.12	98.569
CCTV5	6.473	583	9	77.74	68.25	98.456
					Average	**98.54**

Table 6. Accuracy Determination for Human Detected Keyframes Using Proposed Method

Surveillance dataset	Avg. Fps	Frames	HOG- SVM key frame extraction	Precision	Recall	CR
CCTV1	6.809	435	6	92.60	78.13	98.621
CCTV2	6.698	537	5	100	90.17	98.883
CCTV3	6.922	580	7	77.06	72.94	98.793
CCTV4	6.918	629	9	100	93.12	98.569
CCTV5	6.749	534	5	77.46	67.24	98.699
					Average	**98.71**

Therefore, reports of average frames per second with a smaller time deduction achieve the demonstration of time complexity. Additionally, the attainment of space complexity is determined by comparing the keyframe obtained in Table 6 to that of Table 5, where the former is found to be superior.

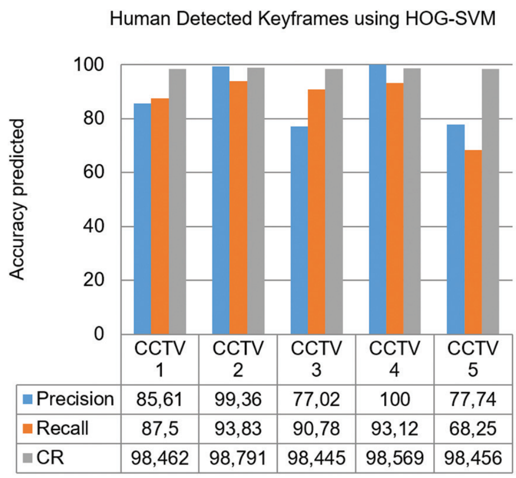
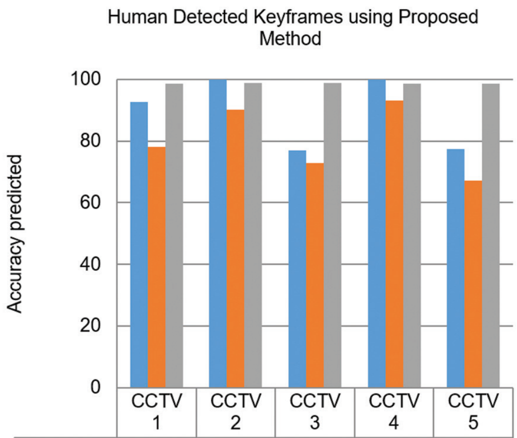


Fig. 3. Accuracy Metrics of Human Detected Keyframes using HOG-SVM



	CCTV 1	CCTV 2	CCTV 3	CCTV 4	CCTV 5
Precision	92,6	100	77,06	100	77,46
Recall	78,13	90,17	72,94	93,12	67,24
CR	98,621	98,883	98,793	98,569	98,69

Fig. 4. Accuracy Metrics of Human Detected Keyframes using the proposed method

Figs. 3 and 4 illustrate the precision and recall calibrations, demonstrating that the proposed work obtains the highest level of differentiation in comparison to prior work. The accuracy metrics of the proposed method yield an average compression ratio of 98.71%, which is superior to the prior method's maximum compression ratio of 98.54%, as shown in Tables 5 and 6 for human-detected keyframes. Consequently, the complexity of performance analysis reports is reduced in terms of both time and space.

5. CONCLUSION AND FUTURE WORK

Human-detected keyframes are categorized in this paper based on the progression of research in content-based video retrieval of surveillance video. The background subtraction method and frame difference facilitate the classification of human motion via pixel change. The human is highlighted as a rectangular segment by the Spatio-temporal feature extraction using HOG-SVM. Experiments utilizing the aforementioned combination algorithm demonstrate that the proposed work enhances the human detection accuracy of keyframe extraction, thereby reducing the time complexity of criminal investigations. The proposed method eliminates the maximal redundancy of frames and demonstrates the space complexity. In future work, the video footage will be fine-tuned under all circumstances to explicitly report human detection at crime scenes.

6. REFERENCES:

[1] M. P. J. Ashby, "The Value of CCTV Surveillance Cameras as an Investigative Tool: An Empirical Analysis", European Journal on Criminal Policy and Research, Vol. 23, No. 3, 2017, pp. 441-459.

[2] A. Ranjan, D. T. Hoffmann, D. Tzionas, S. Tang, J. Romero, M. J. Black, "Learning Multi-human Optical Flow", International Journal of Computer Vision, Vol. 128, No. 4, 2020, pp. 873-890.

[3] B. Garcia-Garcia, T. Bouwmans, A. J. R. Silva, "Background subtraction in real applications: Challenges, current models and future directions", Computer Science Review, Vol.35,2020.

[4] P. Karpagavalli, A. V. Ramprasad, "An adaptive hybrid GMM for multiple human detection in crowd scenario", Multimedia Tools & Applications., Vol.76, No. 12, 2017, pp. 14129-14149.

[5] S. Abdul, R. Shaikh, L. R. Wadekar, E. Engineering, T. Engineering, "Object Detection And Classification Using Sparsity Regularized Pruning On Low-Quality Image / Video", International Journal of Creative Research Thoughts, Vol. 10, No. 6, 2022, pp. 977-985,

[6] N. Paul, A. Singh, A. Midya, P. P. Roy, D. P. Dogra, "Moving object detection using modified temporal differencing and local fuzzy thresholding", Journal of Supercomputing, Vol. 73, No. 3, 2017, pp. 1120-1139.

[7] H. S. G. Supreeth, C. M. Patil, "Efficient multiple moving object detection and tracking using combined background subtraction and clustering", Signal, Image Video Processing, Vol. 12, No. 6, 2018, pp. 1097-1105.

[8] M. S. Zaharin, N. Ibrahim, T. M. A. T. Dir, "Comparison of human detection using background subtraction and frame difference", Bulletin of Electrical Engineering and Informatics, Vol. 9, No. 1, 2020, pp. 345-353.

[9] L. Maddalena, A. Petrosino, "Background subtraction for moving object detection in RGBD data: A survey", Journal of Imaging, Vol. 4, No. 5, 2018.

[10] M. N. Chapel, T. Bouwmans, "Moving objects detection with a moving camera: A comprehensive review", Computer Science Review, Vol. 38, 2020, p. 100310.

[11] S. K. Singh, "A Comparative Study Of Different Motion Detection Algorithms In Computer Vision Applications", International Research Journal of Modernization in Engineering Technology and Science, Vol. 4, No. 6, 2022, pp. 4173-4184.

[12] R. Zhong, R. Hu, S. Member, Z. Wang, S. Wang, "Using Compressed Video", IEEE Signal Processing Letters, Vol. 21, No. 7, 2014, pp. 834-838.

[13] V. Ghait, S. Karekar, N. Lagad, K. Mohare, M. Thorat, "Survey On Key Frame Extraction And Object Detection", International Research Journal of Engineering and Technology, Vol. 7, No. 12, 2020, pp. 2002-2005.

[14] R. Hannane, A. Elboushaki, K. Afdel, P. Naghabhushan, M. Javed, "An efficient method for video shot boundary detection and keyframe extraction using SIFT-point distribution histogram", International Journal of Multimedia Information Retrieval, Vol. 5, No. 2, 2016, pp. 89-104.

[15] B. S. Rashmi, H. S. Nagendraswamy, "Effective Video Shot Boundary Detection and Keyframe Selection using Soft Computing Techniques", International Journal of Computer Vision and Image Processing, Vol. 8, No. 2, 2018, pp. 27-48.

[16] S. Chakraborty, D. M. Thounaojam, "SBD-Duo: a dual-stage shot boundary detection technique robust to motion and illumination effect", Multimedia Tools and Applications, Vol. 80, No. 2, 2021, pp. 3071-3087.

[17] Z. M. Lu, Y. Shi, "Fast video shot boundary detection based on SVD and pattern matching", IEEE Transactions on Image Processing, Vol. 22, No. 12, 2013, pp. 5136-5145.

[18] P. G. G. Lakshmi, S. Dominic, "Walsh-Hadamard transform kernel-based feature vector for shot boundary detection", IEEE Transactions on Image Processing, Vol. 23, No. 12, 2014, pp.5187-5197.

[19] D. M. Thounaojam, T. Khelchandra, K. M. Singh, S. Roy, "A Genetic Algorithm and Fuzzy Logic Approach for Video Shot Boundary Detection", Computer Intelligence and NeuroScience, Vol. 2016, 2016.

[20] S. Tippaya, S. Sitjongsataporn, T. Tan, M. M. Khan, K. Chamnongthai, "Multi-Modal Visual Features-Based Video Shot Boundary Detection", IEEE Access, Vol. 5, No. C, 2017, pp. 12563-12575.

[21] P. Aigrain, H. Zhang, D. Petkovic, "Content-based representation and retrieval of visual media: A state-of-the-art review", Multimedia Tools and Applications, Vol. 3, No. 3, 1996, pp. 179-202.

[22] H. J. Zhang, J. Y. A. Wang, Y. Altunbasak, "Content-based video retrieval and compression: A unified solution", Proceedings of International Conference on Image Processing, Santa Barbara, CA, USA, 26-29 October 1997, pp. 13-16.

[23] U. Gargi, R. Kasturi, S. H. Strayer, "Performance characterization of video-shot-change detection methods", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, No. 1, 2000, pp. 1-13.

[24] T. Liu, H. J. Zhang, F. Qi, "A Novel Video Key-Frame-Extraction Algorithm Based on Perceived Motion Energy Model", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, No. 10, pp. 1006-1013, 2003.

[25] U. Gawande, K. Hajari, Y. Golhar, "Deep Learning Approach to Key Frame Detection in Human Action Videos", Recent Trends in Computational Intelligence, 2020 pp. 1-16.

[26] S. S Gornale, A. K. Babaleshwar, P. L. Yannawar, "Detection and Classification of Signage's from Random Mobile Videos Using Local Binary Patterns", International Journal of Image, Graphics and Signal Processing, Vol. 10, No. 2, 2018, pp. 52-59.

[27] D. Asha, Y. Madhavee Latha, V. S. K. Reddy, "Content-Based Video Retrieval System Using Multiple Features", International Journal of Pure and Applied Mathematics, Vol. 118, No. 14, 2018, pp. 287-294,

[28] M. Jian, S. Zhang, L. Wu, S. Zhang, X. Wang, Y. He, "Deep key frame extraction for sports training", Neurocomputing, Vol. 328, 2019, pp. 147-156.

[29] A. Ioannidis, V. Chasanis, A. Likas, "Weighted multi-view key-frame extraction", Pattern Recognition Letters, Vol. 72, 2016, pp. 52-61,

[30] M. Sima, "Key frame extraction for human action videos in dynamic spatio-temporal slice clustering", Journal of Physics: Conference Series, Vol. 2010, No. 1, 2021.

[31] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 7509, 2005, pp. 94-102.

[32] A. J. Lipton, H. Fujiyoshi, R. S. Patil, "Moving target classification and tracking from real-time video", Proceedings Fourth IEEE Workshop on Applications of Computer Vision, Princeton, NJ, USA, 19-21 October 1998, pp. 8-14.

[33] J. Guo, J. Wang, R. Bai, Y. Zhang, Y. Li, "A New Moving Object Detection Method Based on Frame-difference and Background Subtraction", IOP Conference Series: Materials Science and Engineering, Vol. 242, No. 1, 2017.

[34] D. S. Suresh, M. P. Lavanya, "Motion Detection and Tracking using Background Subtraction and Consecutive Frames Difference Method", International Journal of Research Studies in Science, Engineering and Technology, Vol. 1, No. 5, 2014, pp. 16-22.

[35] P. Ramya, R. Rajeswari, "A Modified Frame Difference Method Using Correlation Coefficient for Background Subtraction", Procedia Computer Science, Vol. 93, 2016, pp. 478-485.

[36] T. Mahalingam, M. Subramoniam, "A robust single and multiple moving object detection, tracking and classification", Applied Computing and Informatics, Vol. 17, No. 1, 2017, pp. 2-18.

[37] V. N. Vapnik, "An overview of statistical learning theory", IEEE Transactions on Neural Networks, Vol. 10, No. 5, 1999, pp. 988-999.

[38] H. Aung, A. V. Bobkov, N. L. Tun, "Face detection in real-time live video using yolo algorithm based on VGG16 convolutional neural network", Proceedings of the International Conference on Industrial Engineering, Applications and Manufacturing, Sochi, Russia, 17-21 May 2021, pp. 697-702.

[39] H. Yang, Q. Tian, Q. Zhuang, L. Li, Q. Liang, "Fast and robust key frame extraction method for gesture video based on high-level feature representation", Signal, Image and Video Processing, Vol. 15, No. 3, 2021, pp. 617-626.