Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

A Hybrid Novel Methodology for Human-Detected Keyframe Extraction in Crime Scene Analysis

Rajeshwari D[1], Victoria Priscilla C[2]

[1]Research Department of Computer Science, Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women, University of Madras, Chennai- 600044
E-mail: rajeshwari.d@sdnbvc.edu.in
ORCID iD: https://orcid.org/0000-0002-6888-942X
[2]PG Department of Computer Science, Shrimathi Devkunvar Nanalal Bhatt Vaishnav College for Women, University of Madras, Chennai-600044
E-mail: victoriapriscilla.c@sdnbvc.edu.in
ORCID iD: https://orcid.org/ 0000-0003-4066-8032

ARTICLE INFO

Received: 30 Dec 2024

Revised: 05 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Introduction: Surveillance system research is currently growing rapidly. Residential, public transit, financial institutions, and other public significant locations where it is frequently employed. It is also crucial for safeguarding critical infrastructure. Due to its numerous uses, human identification in surveillance system video scenes has become more and more popular in recent years. For a number of application areas, including crime scene investigation, a video surveillance system's ability to accurately recognise humans is crucial.

Objectives: This paper presents the Human Detected Keyframe Extractor (HDKFE), a novel framework for efficient keyframe extraction tailored to crime scene investigation and surveillance applications.

Methods: HDKFE integrates Faster R-CNN with an optimized threshold tuning and Keyframe extraction using Local Maxima with Canny Edge Detection selectively capture frames containing high human activity, effectively reducing redundancy while preserving critical content.

Results: Comparative evaluations on existing and real-time surveillance datasets demonstrate HDKFE's superior performance, achieving compression ratios up to 99.22% with high accuracy and also a comparative with different techniques HDKFE method proves with 98.94% higher accuracy.

Conclusions: This efficiency reduces storage demands and enables streamlined analysis of extensive footage, positioning HDKFE as an effective tool for evidence gathering in dynamic surveillance environments. Future advancements may focus on real-time adaptability, integration with predictive analytics, and improved automation to support comprehensive monitoring and proactive security measures.

Keywords: Keyframe Extraction, Video Retrieval, Deep Learning techniques, Video Processing, Crime Scene Analysis.

INTRODUCTION

Surveillance footage has become an essential tool in crime scene investigations, offering valuable insights that can aid in the identification of suspects and the reconstruction of events. However, as the amount of video data collected through closed-circuit television (CCTV) systems grows, the process of manually analyzing this footage has become increasingly time-consuming and resource-intensive. This challenge necessitates the development of automated systems capable of efficiently extracting important information from video streams [1].Automated video analysis techniques, such as keyframe extraction, play a crucial role in reducing the burden on law enforcement while ensuring that critical information is preserved [2]

Keyframe extraction serves as a crucial technique for addressing the challenges of analyzing large video datasets. By condensing extensive video sequences into a limited number of keyframes, it allows investigators to focus on the most significant events within a video. Accurate keyframe extraction can significantly decrease the time required to review footage while maintaining the integrity of the visual data [3]. Despite these benefits, extracting keyframes with meaningful information remains a challenge due to complex scenes, varying lighting conditions, dynamic backgrounds, and the diversity of human actions [4].

Before extracting keyframes, it is necessary to enhance the human presence in order to easily identify the culprit responsible for the incident. In this context, the recent advances in deep learning have opened new avenues for automating the analysis of surveillance footage. Convolutional neural networks (CNNs)[5], especially region-based models like Faster Region-Convolutional Neural Network (Faster R-CNN), have shown significant potential in detecting objects and human activities within video frames [6]. These models have been effective in overcoming challenges such as precise detection of relevant objects in dynamic scenes [7]. By leveraging these capabilities, automated keyframe extraction can be improved, ensuring that only frames with detected human activities are selected for further review. Several methodologies for human detection and keyframe extraction have been proposed.

The **Histogram of Oriented Gradients (HOG)**and **Support Vector Machine (SVM)** approaches have been foundational in recognizing human contours and shapes by analyzing gradient orientations within images, yielding moderate accuracy but with notable limitations in crowded or complex backgrounds[8]. However, these methods may lack precision in dynamic or real-time applications due to their sensitivity to background noise.

Recent developments in **Convolutional Neural Networks (CNNs)**, such as **Faster R-CNN**, have shown promise in human detection by leveraging region proposal networks that can localize human presence with improved accuracy and speed. Faster R-CNN has been particularly useful for real-time applications, as it identifies regions of interest (ROIs) and applies feature extraction to optimize detection[9]. Despite this, the standard Faster R-CNN approach may still produce redundant frames, leading to inefficiencies in storage and analysis.

To enhance detection precision and reduce redundancy, this paper introduces the **Human-Detected Keyframe Extractor (HDKFE)[10],** which aims to address these challenges with different datasets by selectively extracting frames that contain human activity along with keyframe extractor using local maxima with **Canny Edge Detection** method is incorporated to further isolate frames with distinctive human contours, enhancing the model's ability to identify meaningful keyframes frames in complex scenes [11].

The framework of this written work is as follows: Section 2 offers a succinct summary of pertinent research investigations employed to gain insights into the current status of Keyframe Extraction technologies. Section 3 elucidates the specific comparison study of various human detection methodologies, while Section 4 presents the comparative results. The report concludes in Section 5, which provides a succinct summary of prospective future research initiatives.

REVIEW LITREATURE

Recent advancements in human detection with keyframe extraction have focused on improving accuracy and computational efficiency, especially for large-scale surveillance applications. Traditional approaches, such as **HOG-SVM** and **HAAR-like cascade classifier-based detectors**, have been widely used for object and human detection due to their simplicity and moderate accuracy. **HOG-SVM** captures shape and gradient information to distinguish humans from backgrounds, yet its sensitivity to environmental changes and background complexity often leads to false positives, limiting its reliability in dynamic settings [12]. Meanwhile, the **HAAR-like cascade classifier**, which relies on cascades of features to detect faces or full human forms, is computationally efficient but struggles with high accuracy in crowded environments or scenes with low lighting [13]

In recent years, **Deep Learning (DL)** models have emerged as effective tools for human detection in video surveillance, particularly with the development of CNN-based methods. **Faster R-CNN**, one of the most widely adopted DL models; combines region proposal networks (RPNs) with CNNs to perform both object localization and classification within a single framework, achieving high accuracy in diverse scenarios [14]. Faster R-CNN's ability to detect multiple objects per frame and handle complex backgrounds makes it a strong candidate for real-time human detection. However, without optimizations, Faster R-CNN may generate redundant frames that increase storage and processing costs, which can be a limitation in long-term video analysis [15].

Criminal investigations predominantly depend on the behavioural patterns of individuals observed in surveillance footage. To enhance encouragement of future proposals, periodic reviews of deep learning approaches for keyframe extraction that utilize human detection are essential. Several studies have focused on optimizing keyframe extraction to improve data compression and storage efficiency. **Region of Interest (ROI)** methods, for example, focus on identifying the most visually significant parts of a video frame, thereby reducing the need for storing full-frame sequences. While these methods achieve higher compression ratios, they may overlook key contextual details necessary for effective crime investigation [15].

All of the stated strategies can exhibit exceptional performance under suitable conditions, but they tend to overlook variations in the precise motion states of the objects in motion. The suggested methodology identifies humans that fall into the category of keyframe extraction based on motion. The suggested HDKFE may efficiently extract key frames that accurately describe humans, while minimizing unnecessary information, surpassing prior approaches[16].

METHODS

The suggested methodology comprises three processes for extracting the keyframes detected by humans. These phases are outlined as follows: The approach consists of three phases: (1) Pre-processing phase, (2) Human detection algorithm (3) Keyframe Extraction. In this context, pre-processing techniques are employed to improve the video quality, whereas the human detection is made with an comparative analysis of HOG-SVM, HAAR-like cascade classifier, HOG-SVM with background subtraction, Faster R-CNN and HDKFE method to prove where humans are recognized as the best in all the frames obtained, as shown in Figure 1. Thus the Human-detected video surveillance keyframes are extracted in support of the criminal investigation after the classification of humans is completed in contrast to other distinguishable moving objects from the existing method.

3.1 Video Pre-processing

Surveillance footage is initially converted to gray-scale. This step reduces the computational load by eliminating color information, which is unnecessary for detecting shapes and contours. This is followed by resizing the video to form uniformity in all frames. Here, the gray-scale images are resized to a standard resolution of 640×480 pixels. Resizing ensures that all frames maintain uniform dimensions, allowing the processing model to operate more efficiently across datasets and adapt to different camera resolutions. Then the video is adjusted by brightness and contrast to improve the visibility of human figures, especially in dim lighting or variable weather conditions. By tuning brightness and contrast, this step also aids subsequent detection algorithms in recognizing distinct shapes and edges more accurately.

3.2 Human Detection Techniques

Human detection presents a formidable challenge in machine vision due to the myriad of potential appearances resulting from variations in articulated position, attire, lighting, and background; yet, prior awareness of these constraints can enhance detection efficacy. In our previous work the human detection is performed by five methodologies in order to attain the best selection of bounded box humans in all frames from the pre-processed video.

3.2.1. HOG-SVM (Histogram of Oriented Gradients-Support Vector Machine)

HOG captures human contours based on gradient orientations, effectively highlighting edges and shapes relevant to human detection. This method divides the image into small regions and calculates gradient orientations, focusing on regions with significant shape variations typical of humans. SVM is applied to the gradient features to classify detected regions as "human" or "non-human," filtering out irrelevant objects.

3.2.2. HAAR-Like Cascade Classifier

A HAAR-like cascade classifier is trained on images with human features to detect shape patterns that resemble human forms, such as heads, torsos, or limbs. This classifier employs a series of scaled features to recognize and filter out human shapes, while suppressing irrelevant structures in the frame.

3.2.3. HOG-SVM with Background Subtraction(HOG-SVM with BS)

Background subtraction helps isolate moving objects (e.g., humans) from static backgrounds. The model identifies static pixels as part of the background and discards them, minimizing noise and focusing on regions with active movement. Hence the human detection becomes easier when compared to prior methods. The amalgamation of Background-Subtraction and Frame-Differencing possesses certain characteristics, including the capability to reduce noise within the frame and effectively address gaps created by Background-subtraction scenarios through its fusion technique[17]. This method eliminates the noise present in the frame and can greatly fill the holes obtained by the Background-Subtraction at certain situation due to its fusion method. This system also executes image restoration and employs several morphological processing techniques to accurately identify moving objects. Also this method performs the image repair and some morphological processing methods to detect the moving objects definitely

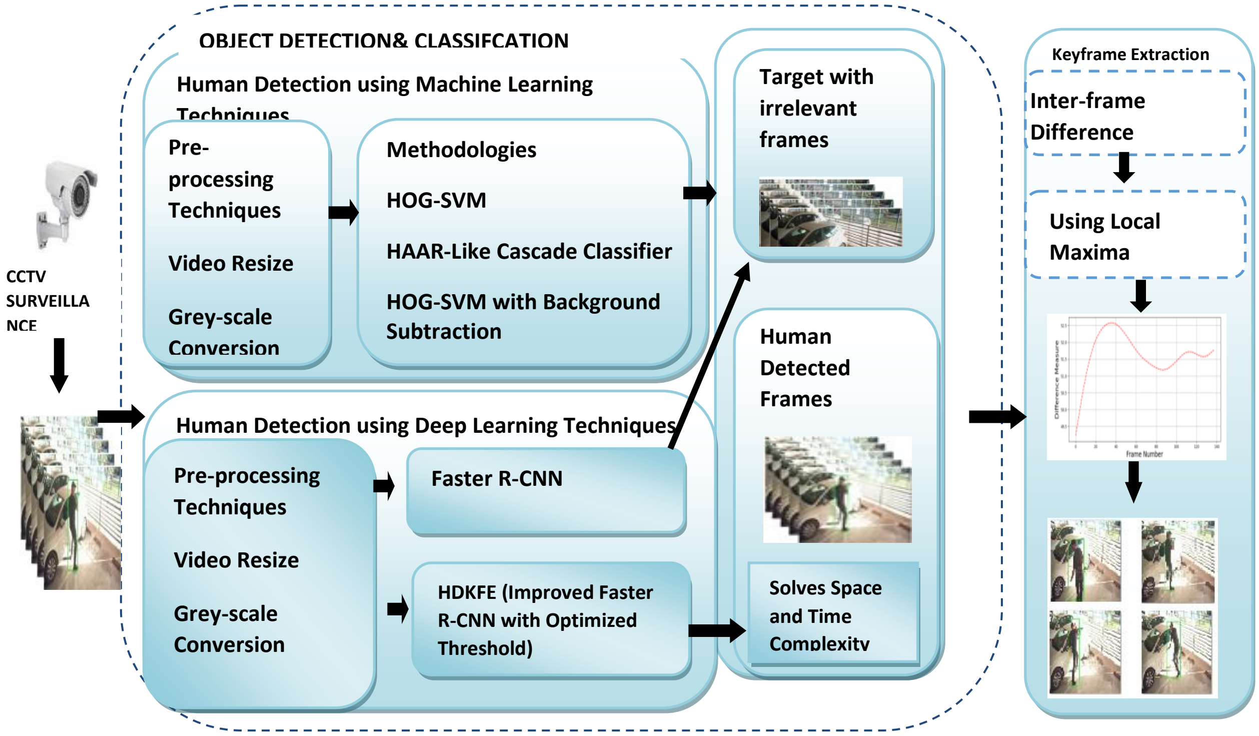

OBJECT DETECTION& CLASSIFCATION
Human Detection using Machine Learning Techniques
Pre-processing Techniques
Video Resize
Grey-scale Conversion
Methodologies
HOG-SVM
HAAR-Like Cascade Classifier
HOG-SVM with Background Subtraction
CCTV SURVEILLANCE
Human Detection using Deep Learning Techniques
Pre-processing Techniques
Video Resize
Grey-scale Conversion
Faster R-CNN
HDKFE (Improved Faster R-CNN with Optimized Threshold)
Target with irrelevant frames
Human Detected Frames
Solves Space and Time Complexity
Keyframe Extraction
Inter-frame Difference
Using Local Maxima
Difference Measure
Frame Number

Fig 1. Proposed Work Flow

3.2.3. FASTER R-CNN

This method comes under the Deep Learning category. Here the Faster R-CNN employs Region Proposal Network (RPN) to generate anchor boxes for probable human locations. Each box is then analyzed to verify the presence of human figures, enhancing human detection accuracy. The Convolutional layers refine the proposal, detecting human subjects by creating a feature map, which accurately identifies human presence with lower computational demand than typical sliding window methods.

3.2.4. HDKFE (An Improved Faster R-CNN)

The HDKFE[10] method introduces an optimized threshold mechanism within Faster R-CNN to ensure that only frames with a high probability of containing humans are marked. By fine-tuning the confidence threshold, the method minimizes false positives, ensuring that only frames with clear human detection are processed further.

3.2. Implementation Phase with the combination of Keyframe Extraction (KFE)

Keyframes are selected based on changes in pixel intensity across frames, capturing moments with the most distinct variations indicative of human movement. Canny Edge Detection identifies edges in each frame by calculating gradients. By focusing on edges, it can locate boundaries within human figures, even in complex backgrounds. When combined with local maxima identification, Canny Edge Detection enhances the system's ability to isolate meaningful frames that highlight human activity. The above methodologies are now processed with this Keyframe extractor to produce the human detected keyframes.

The metrics calibration of HOG-SVM with keyframe extractor[18] of machine learning results with 98.21% success rate, whereas the HAAR-like Cascade classifier results with 98.12% accuracy[19]. The first two categories fail due to lack of human detection in all frames. Hence the machine learning is further enhanced by using HOG-SVM with Background Subtraction resulting with best keyframes. This configuration doesn't support for the real time datasets so the deep learning methodologies are further processed. Thus, the Faster R-CNN network is created with a comprehensive framework for accurately identifying the required human detection in all resultant surveillance footages frames.

This step ensures that only essential frames representing key human activity are stored, reducing redundancy. Pixel intensity changes across consecutive frames are examined, allowing the system to identify frames with substantial content changes. This variation analysis ensures that redundant frames are minimized. Local maxima are selected based on peaks in frame intensity changes, indicating a significant shift in visual content. By evaluating these peaks, the system identifies frames that contain the highest relevance, reducing temporal complexity.

RESULTS

To evaluate the HDKFE method's performance, two primary datasets were utilized:

1. **Existing Datasets**: This dataset comprises pre-recorded surveillance footage from controlled environments with known human subjects in various postures and actions, totalling approximately 7,200 frames per sequence. Existing datasets such as Shutter-Stock (SS), PETS 2009, Mots20, Avg Town Centre (ATC), TUDS, and VIRAT offer diverse lighting and background conditions, enabling a comparison between HDKFE and other detection methods such as HOG-SVM, HAAR-like cascade classifier, and Faster R-CNN.
2. **Real-Time Datasets**: Real-time datasets were collected using a closed-circuit television (CCTV) system deployed on campus premises, recording natural human movements. This dataset, averaging around 7,100 frames per sequence, captures authentic scenes with dynamic lighting, moving crowds, and background complexities. Such conditions are crucial for testing HDKFE's robustness in real-world applications where environmental factors vary significantly.

Table 1. Real time Dataset using HDKFE method

Surveillance Video (Own dataset)	Actual video time (seconds)	Human detected video time (seconds)	Frames Obtained	Human detected Frames	Keyframes	Human Detected Keyframes	CR
Video 1	0.49	0.28	489	489	7	7	98.56
Video 2	0.53	0.29	643	292	8	8	98.75
Video 3	1.03	0.29	598	431	5	5	99.16
Video 4	0.54	0.28	639	424	6	6	99.06
Video 5	1.00	0.27	709	421	10	10	98.59
Video 6	1.03	0.40	956	625	8	8	99.16
Video 7	0.50	0.17	742	290	3	3	99.59
Video 8	1.00	0.06	886	131	2	2	99.77
Video 9	0.50	0.12	730	243	3	3	99.58
Video 10	0.50	0.14	747	260	3	3	99.60

Table 2. Existing Dataset using HDKFE method

Surveillance Video	**Actual video time (seconds)**	**Human detected video time (seconds)**	**Frames Obtained**	**Human detected Frames**	**Keyframes**	**Human Detected Keyframes**	**CR**
SS 1	0.19	0.13	434	272	6	6	99.07
SS 2	0.25	0.21	634	475	6	6	99.05
SS 3	0.54	0.12	1344	241	8	8	99.70
SS 4	0.20	0.12	360	245	8	8	99.16
SS 5	0.20	0.20	1185	920	7	7	99.41
ATC	2.59	0.14	451	362	8	8	98.66
Mots20	0.17	0.13	526	401	7	7	99.05
Mots20	0.30	0.24	901	867	9	9	99.11
PETS2009	1.53	0.20	796	740	9	9	98.86
PETS2009	1.02	0.14	437	250	3	3	99.31
TUDS	0.02	0.01	72	25	4	4	98.61

The HDKFE method performs the best accuracy as reported in the below table1 and table 2 with the existing and real time datasets. Thus, the proposed HDKFE methodology will be advantageous for crime investigations, which exclusively emphasizes the humans in each detected keyframes.

From the above tables the actual video time executed the human detected video time with comparatively less proving the time complexity as shown in table 3.

Table 3. Overview of Video Datasets Utilized in Keyframe Extraction

Dataset	**Total Videos**	**Total Frames Obtained**	**Average Actual Video Time (seconds)**	**Human Detected Video Time (seconds)**	**Average Keyframes**	**Average Human Detected Keyframes**	**Average Compression Ratio (CR)**
Real time Datasets	10	7,085	0.63	0.23	6.5	6.5	98.80
Exiting Datasets	10	6,658	1.00	0.16	7.1	7.1	99.10
Overall Average	20	**13,743**	0.82	0.20	6.8	6.8	98.95

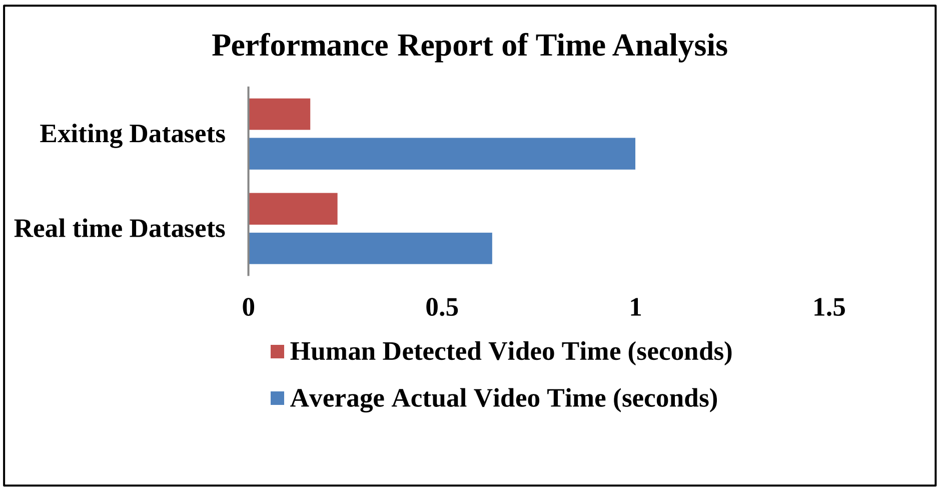


Fig 2. Graphical view of Comparative time analysis

In the proposed HDKFE methodology, a keyframe is extracted in the quickest time by the optimized human detected frames in videos and improving the precision rate of the system as determined in figure 2.

Table 4. Comparative Evaluation of Keyframe Extraction Methods on Existing Datasets with various methodologies

Method	**Frames Processed**	**Keyframes Extracted**	**Compression Ratio (CR)**
HOG-SVM	7249	101	98.6%
HAAR-like Cascade Classifier	7248	102	98.59%
HOG-SVM with BS	7248	99	98.63%
Faster R-CNN	7248	90	98.75%
HDKFE (Proposed)	**7140**	**75**	**98.94%**

Table 5. Comparative Evaluation of Keyframe Extraction Methods on Real-Time Datasets with various methodologies

Method	**Frames Processed**	**Keyframes Extracted**	**Compression Ratio (CR)**
HOG-SVM	7145	96	98.65%
HAAR-like Cascade Classifier	7140	94	98.68%
HOG-SVM with BS	7139	93	98.69%
Faster R-CNN	7139	88	98.76%
HDKFE (Proposed)	**7139**	**55**	**99.22%**

3.3. Performance Analysis

The HDKFE method not only satisfied with different datasets also proved with best accuracy with overall comparative with various methodologies as depicted in table 4 and table 5.

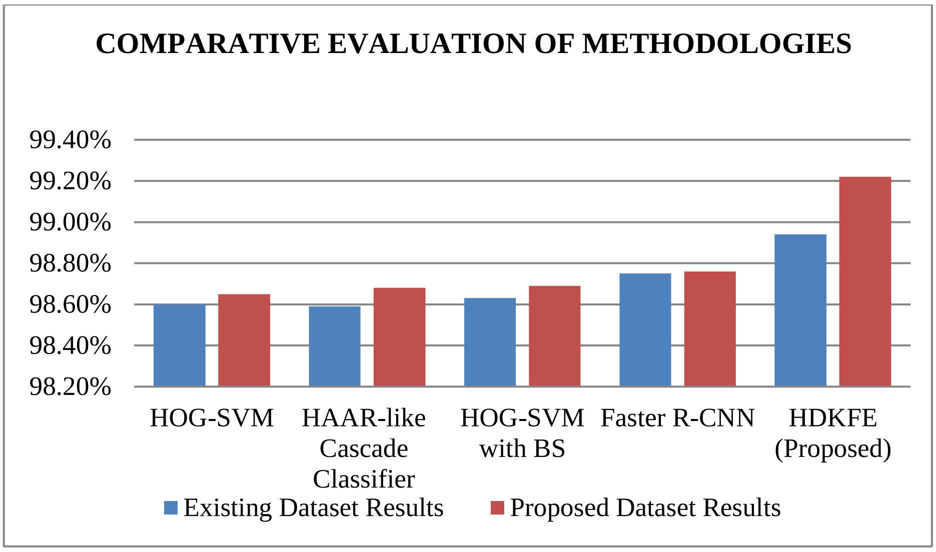


Fig 3. Comparative Analysis of Various Methodologies

As demonstrated in Table 4, HDKFE significantly outperforms other methods by reducing the number of extracted keyframes to 75 while maintaining the highest CR of 98.94%. This efficiency illustrates HDKFE's ability to minimize frame redundancy in controlled environments, thereby optimizing storage needs without losing essential data using existing datasets, whereas HOG-SVM and HAAR retain more frames, reflecting their limitations in differentiating critical content from redundant frames.

Similarly, in real-time datasets as demonstrated in table 5, HDKFE achieves the highest compression ratio of 99.22%with only 55 keyframes extracted. This reflects HDKFE's superior performance under realistic, dynamically changing conditions as shown in Figure 3 whereas HOG-SVM, HAAR, and other methodologies demonstrate lower compression ratios and higher frame counts, indicating higher storage demands due to redundant frame extraction.

3.4. Performance Metrics Analysis

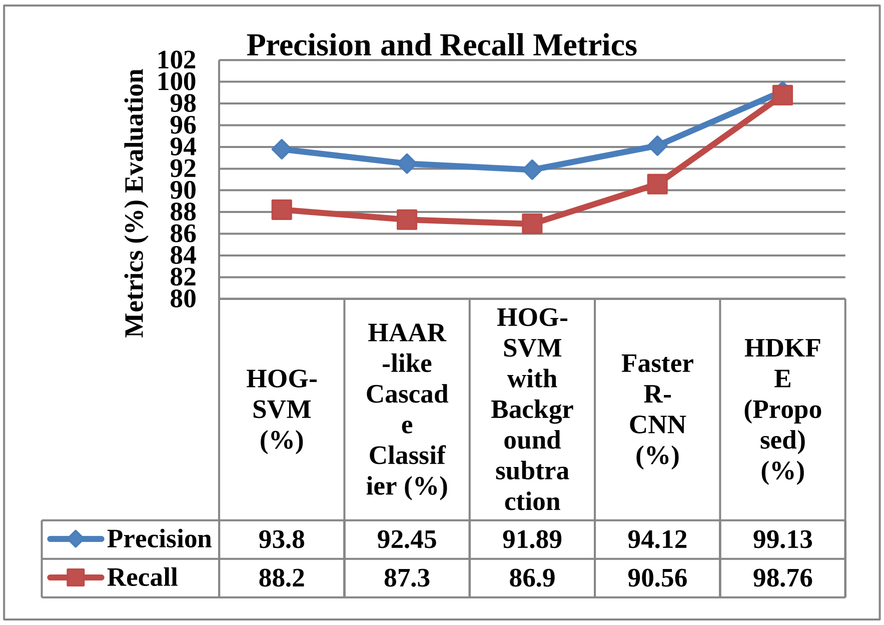
To further evaluate HDKFE's detection efficiency, we analyse its **precision, recall, and compression ratio (CR)** using metrics for Existing and real-time datasets. These values are represented in Figure 4 and Figure 5.**Precision and Recall** are calculated using true positives, false positives, and false negatives for human-detected frames. As determined by Eq. (1), precision denotes the percentage of positive instances that were accurately identified relative to the total number of instances predicted as positive.

$$Precision = \frac{TPH}{TPH+FPH} * 100\% \qquad (1)$$

Here, TPH denotes True Positive Human and FPH denotes False Positive Human. **Recall**, as defined in Eq.(2), quantifies the ratio of correctly detected true positive occurrences to the total number of actual positive instances.

$$Recall = \frac{TPH}{TPH+ FNH} * 100\% \qquad (2)$$

Here, FNH denotes False Negative Human.



	HOG-SVM (%)	HAAR-like Cascade Classifier (%)	HOG-SVM with Background subtraction	Faster R-CNN (%)	HDKFE (Proposed) (%)
Precision	93.8	92.45	91.89	94.12	99.13
Recall	88.2	87.3	86.9	90.56	98.76

Fig 4. Precision and Recall of HDKFE vs. Comparative Methods

Compression Ratio (CR) quantifies data reduction efficiency, maintaining video quality by representing core content in keyframes instead of the entire frame sequences. This metric quantifies the decrease in file size attained by predominantly representing the video content through keyframes as opposed to utilizing each frame. It showcases the efficacy of keyframe extraction in minimizing file size without compromising video quality, as represented by Eq. (3).

$$CR = 1 - \left\{\frac{HKF}{HF}\right\} * 100 \qquad (3)$$

By reaching a CR of 99.22% in real-time datasets, HDKFE effectively lowers storage demands without sacrificing important content. This capability is particularly beneficial in real-time surveillance systems, where data is continuously accumulated and storage efficiency is paramount.

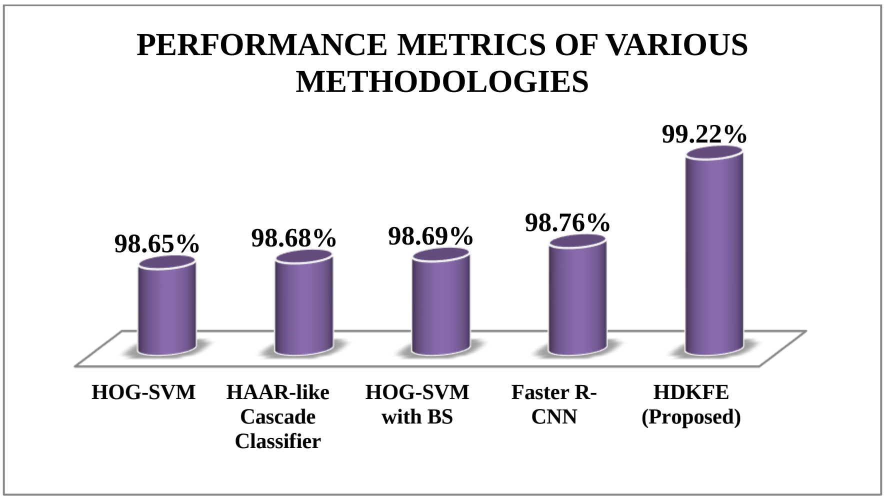
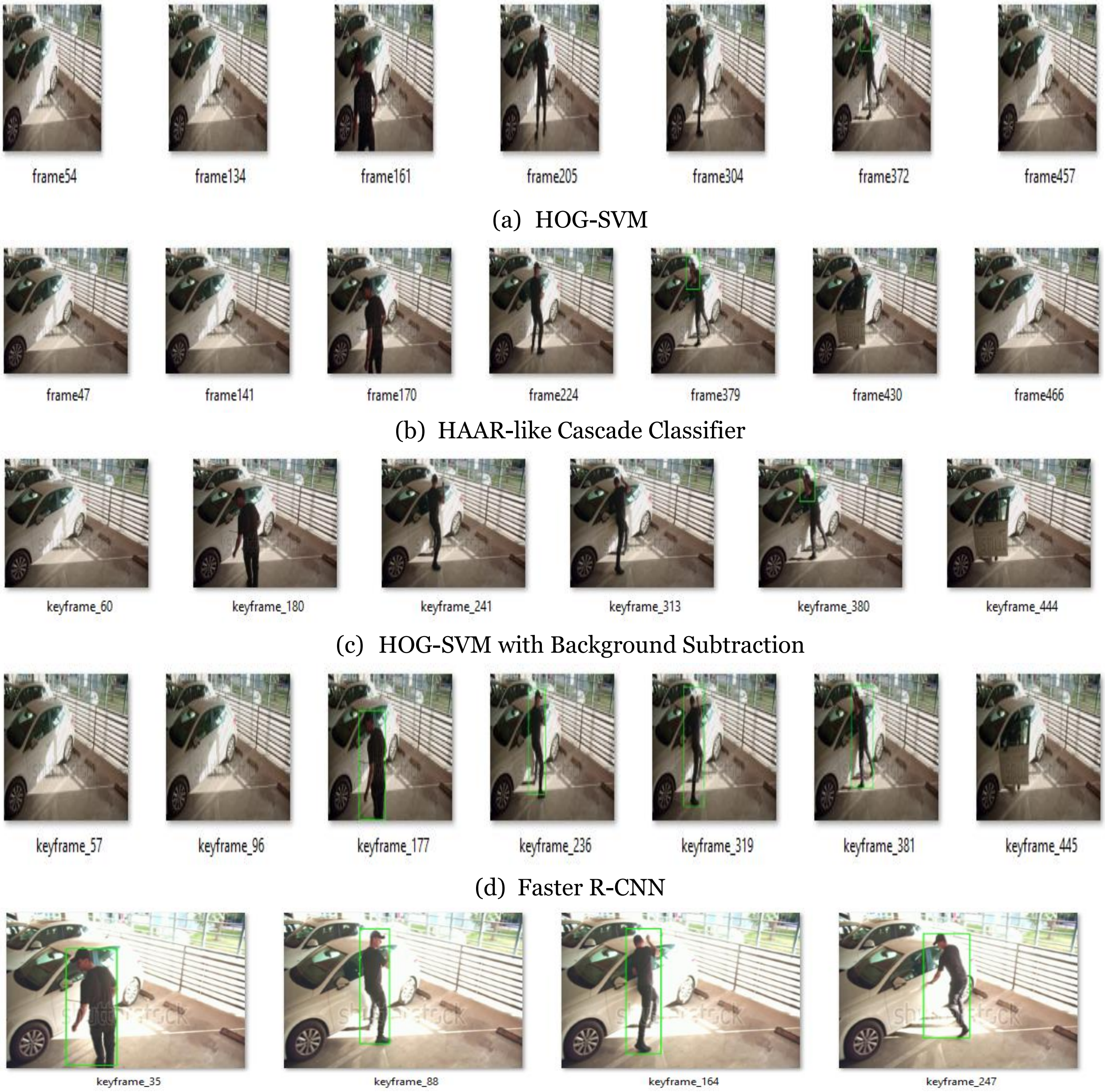


Fig 5. Compression Ratio Comparison of HDKFE and Other Methods

3.5. HDKFE's Key Advantages

Reduced Frame Redundancy: HDKFE's methodology achieves fewer extracted keyframes without compromising content critical for investigation, as evidenced by its consistently higher CR across both public and real-time datasets. By isolating only the most informative frames, HDKFE minimizes the storage burden on surveillance systems, which is crucial in long-duration recordings where storage and retrieval efficiency are key as represented in Figure 6.



(a) HOG-SVM

(b) HAAR-like Cascade Classifier

(c) HOG-SVM with Background Subtraction

(d) Faster R-CNN

(e) HDKFE

Fig 6. Resultant Keyframe extracted using Various Methodologies

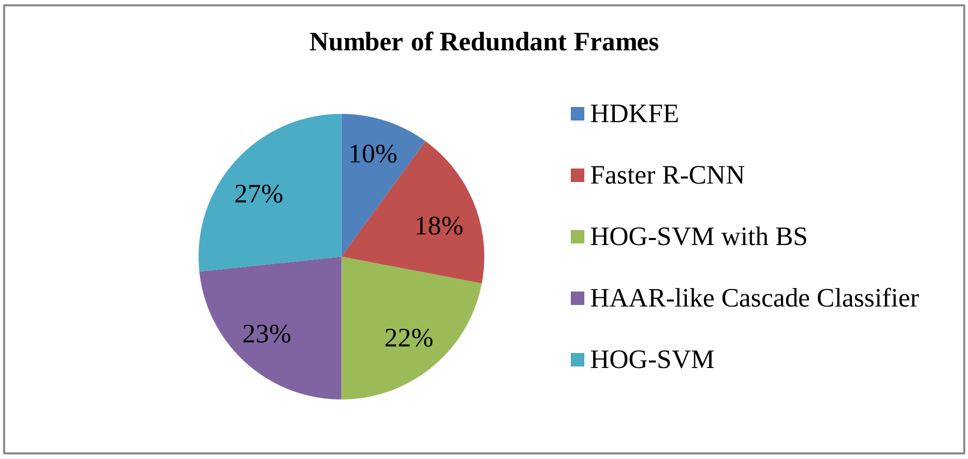


Fig 7. Frame Redundancy Analysis for HDKFE and Other Methods

HDKFE's precision of 99.13% and recall of 98.76% signify its accuracy in detecting human frames without falsely identifying non-human frames. Both HOG-SVM and HAAR exhibit lower precision and recall, showing limitations in accurately isolating relevant frames, particularly in scenes with complex backgrounds. Faster R-CNN performs better but is still outperformed by Human Detected Keyframe Extractor (HDKFE), indicating that HDKFE's threshold optimization allows for more refined frame selection as reported with its precision and recall graph as shown in Figure 7. The number of redundant frames identified across methods, with HDKFE showing the least redundancy, as indicated by fewer extracted keyframes across datasets. HDKFE's low redundancy rate is a result of its dual approach of using Faster R-CNN with optimized thresholding, reducing the presence of visually similar frames. This advantage is critical for surveillance footage, where prolonged recordings necessitate effective data compression without data loss. Thus, HDKFE extracts fewer yet more meaningful keyframes compared to other methods. This efficiency ensures that only essential human activity is captured, ideal for crime investigation applications where excess storage and processing costs can impede timely analysis

High Accuracy with 99.13% Detection Rate: The HDKFE method achieves an impressive **accuracy rate of 99.13%**. This high detection rate ensures that keyframes extracted contain relevant human activity, which is particularly beneficial for crime scene investigations were missing keyframes can lead to incomplete or inconclusive evidence.

Efficient Use of Canny Edge Detection: Canny Edge Detection is used in tandem with local maxima identification to refine the selection of keyframes. This process detects sharp changes in frame content, isolating frames with clear human outlines or movements. By using Canny Edge Detection, HDKFE reduces spatial complexity by filtering out frames with limited human activity, focusing instead on frames where human contours and actions are well-defined.

Robustness in Dynamic Environments: HDKFE's performance on real-time datasets, where background elements and lighting vary, demonstrates its robustness. This adaptability to natural environments supports HDKFE's application in varied surveillance contexts, making it a reliable tool for public safety agencies.

Effective Spatial Complexity Management: HDKFE's integration of Canny Edge Detection enhances spatial complexity management, as frames with minimal or no human presence are filtered out. This spatial efficiency is further demonstrated by HDKFE's reduced frame redundancy compared to other methods.

DISCUSSION

The Human-Detected Keyframe Extractor (HDKFE) method demonstrates significant advancements in keyframe extraction for surveillance footage, especially suited for applications in crime scene analysis. By integrating optimized Faster R-CNN with keyframe Extraction using Local Maxima and Canny Edge Detection, HDKFE effectively identifies frames with high human activity while minimizing redundant frames. This efficiency is reflected in its superior compression ratios (up to 99.22%) and high precision (99.13%) and recall (98.76%) values, significantly outperforming conventional methods like HOG-SVM, HAAR, and standard Faster R-CNN.HDKFE's unique design reduces spatial and temporal complexity, enabling faster processing and storage efficiency without sacrificing critical data. The robustness of HDKFE in both controlled and dynamic real-world environments underscores its suitability for continuous, high-volume surveillance applications, where accurate, space-efficient data representation is paramount. Overall, Human Detection Keyframe Extractor (HDKFE) stands as a valuable tool for surveillance systems, offering law enforcement and security agencies a reliable, resource-efficient means to process extensive footage while preserving essential information for effective crime investigation and evidence gathering. Future work may focus on enhancing HDKFE's real-time processing capabilities and exploring its integration with other real-time analytics to further optimize its performance in diverse security contexts.

REFERENCES

[1] S. Lakshmanan, D. Baskaran, and S. Kamalanathan, "Human Activity Recognition Through Images Using a Deep Learning Approach." 2024. doi: 10.21203/rs.3.rs-4443695/v1.

[2] K. Kardas *et al.*, "Video Retrieval by Extracting Key Frames in CBVR System," *Elektr. J. Electr. Eng.*, vol. 89, no. 3, pp. 343–361, 2017, doi: 10.1109/ICCSDET.2018.8821168.

[3] P. Saini, K. Kumar, S. Kashid, A. Saini, and A. Negi, *Video summarization using deep learning techniques: a detailed analysis and investigation*, vol. 56, no. 11. Springer Netherlands, 2023. doi: 10.1007/s10462-023-10444-0.

[4] O. Elharrouss, N. Almaadeed, and S. Al-Maadeed, "A review of video surveillance systems," *J. Vis. Commun. Image Represent.*, vol. 77, p. 103116, 2021, doi: https://doi.org/10.1016/j.jvcir.2021.103116.

[5] U. V. Navalgund and P. K. Priyadharshini, "Crime Intention Detection System Using Deep Learning," *2018 Int. Conf. Circuits Syst. Digit. Enterp. Technol. ICCSDET 2018*, pp. 1–6, 2018, doi: 10.1109/ICCSDET.2018.8821168.

[6] N. B. Rani, J. Lavanya, K. Sathwika, M. Likhitha, and N. Sowjanya, "People Counting System Based on Head Detection using Faster RCNN from Both Images and Videos," *Turkish J. Comput. Math. Educ.*, vol. 14, no. 03, pp. 947–954, 2023.

[7] G. V. Mohan and M. P. Arakeri, "Real Time Multi-Object Tracking based on Faster RCNN and Improved Deep Appearance Metric," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 12, pp. 887–894, 2021, doi: 10.14569/IJACSA.2021.01212107.

[8] A. Bam, P. Choudhary, and J. Bhoir, "Real-Time Human Detection and Tracking in Motion Environment," *Int. Res. J. Eng. Technol.*, pp. 3692–3695, 2021, [Online]. Available: www.irjet.net

[9] H. Wei and N. Kehtarnavaz, "Semi-Supervised Faster RCNN-Based Person Detection and Load Classification for Far Field Video Surveillance," *Mach. Learn. Knowl. Extr.*, vol. 1, no. 3, pp. 756–767, 2019, doi: 10.3390/make1030044.

[10] D. Rajeshwari and C. V. Priscilla, "An optimized real-time human detected keyframe extraction algorithm (HDKFE) based on faster R-CNN," vol. 15, pp. 2644–2650, 2024, doi: 10.58414/SCIENTIFICTEMPER.2024.15.3.32.

[11] B. V Patel, "Content Based Video Retrieval Systems," *Int. J. UbiComp*, vol. 3, no. 2, pp. 13–30, 2012, doi: 10.5121/iju.2012.3202.

[12] U. Chandrakant Patkar *et al.*, "International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Object Detection using Machine Learning and Deep Learning," *Orig. Res. Pap. Int. J. Intell. Syst. Appl. Eng. IJISAE*, vol. 2024, no. 1s, p. 466, 2023, [Online]. Available: www.ijisae.org

[13]J. Sujanaa and S. Palanivel, "Hog-Based Emotion Recognition Using One-Dimensional Convolutional Neural Network," *ICTACT J Image Video Process*, vol. 11, no. 2, pp. 2310–2315, 2020, doi: 10.21917/ijivp.2020.0328.

[14]I. Oztel, "Human Detection System using Different Depths of the Resnet-50 in Faster R-CNN," *4th Int. Symp. Multidiscip. Stud. Innov. Technol. ISMSIT 2020 - Proc.*, no. 1, 2020, doi: 10.1109/ISMSIT50672.2020.9255109.

[15]R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014, doi: 10.1109/CVPR.2014.81.

[16]A. Mushan and P. P. Vidap, "Video Summarization using Keyframe Extraction Methods," *Int. J. Recent Technol. Eng.*, vol. 9, no. 2, pp. 1030–1032, 2020, doi: 10.35940/ijrte.b4043.079220.

[17]R. D and V. P. C, "An Enhanced Spatio-Temporal Human Detected Keyframe Extraction," *Int. J. Electr. Comput. Eng. Syst.*, vol. 14, no. 9, pp. 985–992, 2023, doi: 10.32985/ijeces.14.9.3.

[18]C. Victoria Priscilla and D. Rajeshwari, "Video Keyframe Extraction Based on Human Motion Detection," in *Inventive Systems and Control*, V. Suma, Z. Baig, S. Kolandapalayam Shanmugam, and P. Lorenz, Eds., Singapore: Springer Nature Singapore, 2022, pp. 427–441.

[19]C. V. Priscilla and D. Rajeshwari, "Performance Analysis of Spatio-temporal Human Detected Keyframe Extraction," *Remit. Rev.*, vol. 7, no. 1, pp. 159–170, 2022, doi: 10.47059/rr.v7i1.2404.