STM JOURNALS

International Journal of
Advanced Robotics and
Automation Technology

Volume 1, Issue 2, 2023
July—December
DOI (Journal): 10.37591/IJARAT

https://journals.stmjournals.com/ijarat

Review IJARAT

Violent Event Recognition and Monitoring Using Deep Learning for Surveillance Videos

Vedangi Raut*, Rutvik Redkar

Abstract

The significance of real-time capabilities in human detection and tracking is discussed in the abstract of the paper. We talk about tracking, eye detection, and face detection. A thorough motion detection program for use in video monitoring and other applications is suggested by the study. The goal of the study is to further human tracking technology. Optical flow features and appearance-invariant features from a Darknet CNN model are integrated. Acquiring Knowledge of Complicated Sequences: In order to identify complicated activity sequences for ultimate violence detection, a Long Short-Term Memory (LSTM) network is utilized, which enables the system to identify long-term patterns. Thorough Evaluation: The approach outperforms current methods and provides a baseline for violence detection systems when tested in a variety of inside and outdoor surveillance scenarios. The study provides an overview of classical and deep learning-based forms of violence.

Keywords: Convolution neural network (CNN), long short-term memory, violence, models, datasets

INTRODUCTION

Artificial intelligence's Deep Learning field is rapidly becoming into a potent instrument for analysing and identifying violence in security footage. Deep learning technologies use strong computer processors and complex algorithms to analyse films, match them to archived footage, and identify questionable activity. With the use of this technology, violent acts that could have gone undetected could be tracked down and detected, thus improving public safety and averting criminal situations. This article will examine the operation of deep learning-based violence tracking and detection systems and go over some of the possible advantages and disadvantages of the technology.

A MOTION DETECTION SYSTEM IN PYTHON AND OPENCV

In front of a camera, in particular, the research presents a thorough motion detection within an area. Detecting undesired community or animal movements, tracking children's screen time, detecting unauthorized entry to yards, and taking screenshots while working remotely are just a few of the many useful uses for this software. The study suggests using a webcam to record the first frame of a motion detection system, which then computes the phase difference between that frame and future frames—known as Delta frames. Image processing methods including as contouring, dilation, shadow removal, and pixel power augmentation are applied to these Delta frames in order to improve their ability to detect larger objects. Real-time tracking and analysis of item movement is made possible by the acquisition of object entry and exit timestamps. The study demonstrates a comprehensive grasp of the topic by examining a range of relevant works and approaches, such as Gaussian Mixture Model

***Author for Correspondence**
Vedangi Raut
E-mail: vedangiraut1@gmail.com

Research Scholar, MCA, Thakur Institute of Management Studies, Career Development & Research (TIMSCDR), Mumbai, India

Received Date: March 11, 2024
Accepted Date: March 26, 2024
Published Date: April 25, 2024

Citation: Vedangi Raut, Rutvik Redkar. Violent Event Recognition and Monitoring Using Deep Learning for Surveillance Videos. International Journal of Advanced Robotics and Automation Technology. 2023; 1(2): 39–44p.

algorithms and background separation. Using graph plots, the suggested system, which was developed with the help of the OpenCV, Bokeh, Pandas, and Datetime libraries, efficiently visualizes object movement. The report also addresses the value of motion detection in the technologically advanced world of today, showcasing its uses in a range of contexts such as entrance lighting, automated sinks, security lighting, and more. In summary, the study offers a technically sophisticated and user-friendly motion detection system that advances both security and technology. It is accepted that there is room for more modification and extension depending on user needs.

AN INTELLIGENT SYSTEM FOR COMPLEX VIOLENCE PATTERN ANALYSIS AND DETECTION

The article begins with a compelling introduction highlighting the growing importance of artificial intelligence (AI)-driven intrusion detection techniques in the context of industrial security surveillance. It effectively highlights how these technologies are key to creating safer environments in environments as diverse as factories, banks and subways, enabling automated systems to proactively monitor and prevent violent activity. This introduction addresses the main need for security management. Focusing on the challenges and traditional methods, the following article describes the formidable obstacles to video-based violence detection. These include lighting changes, motion blur, complex background dynamics and perspective changes. The paper is careful to note that traditional methods usually rely on manually created features, which are inefficient processes and false predictions that lead to false alarms. The paper then presents the proposed method for detecting violent incidents in surveillance video streams. This method progresses through a carefully planned series of steps. It starts with pre-processing and object detection, which uses a mask-R convolutional neural network (CNN) to segment important objects such as individuals and vehicles to optimize computing resources. To deal with the variety and complexity of violence patterns, the paper presents a framework augmentation algorithm for data augmentation using geometric transformations. Feature extraction is central and includes a CNN-based optical flow estimation method for motion capture, complemented by appearance-invariant features from the Darknet CNN model. The method culminates in the use of a long-short-term memory (LSTM) network to identify complex sequences of actions, allowing the extraction of long-term patterns associated with violent activity. The report carefully evaluates the proposed method in several indoor and outdoor monitoring scenarios. This evaluation serves as a rigorous litmus test that establishes a solid basis for abuse detection methods and, in particular, shows that the proposed approach outperforms existing techniques. A large part of the article is devoted to related work, providing a comprehensive survey. and traditional and deep learning-based violence detection methods. The paper explores methods based on low-level features such as motion and appearance descriptors and deep learning techniques that include 3D CNNs, LSTM networks, and innovative hybrid approaches. The paper skilfully synthesized their contributions and highlighted the effectiveness of this method of violent detection for solving real-world surveillance problems. It emphasizes the sustainability of the method, its adaptability to different models and its possible application in intelligent industrial systems. In addition, the paper confirms the superior performance of the proposed method compared to state-of-the-art techniques, which reinforces its central role in the field of violence detection.

VEHICLE DETECTION AND TRACKING BASED ON VIDEO IMAGE PROCESSING IN INTELLIGENT TRANSPORTATION SYSTEM

The research paper under review explores the pivotal role of Intelligent Transportation Systems (ITS) and their application in vehicle detection and tracking through machine vision technology. The introduction effectively sets the stage by highlighting the real-time, sensitive, and accurate advantages of ITS, underlining its multifaceted significance in modern transportation management. The paper astutely acknowledges the evolving landscape of computer and image processing technologies, pinpointing machine vision detectors as pivotal. These detectors are poised to supplant traditional methods, emerging as a cornerstone in contemporary intelligent transportation systems.

The paper further delves into the research landscape both domestically and internationally, showcasing noteworthy endeavours in the field. One such study focuses on the implementation of

algorithms for vehicle queue length estimation, vehicle detection, counting, and speed estimation. Significantly, these algorithms are executed through hardware means, harnessing parallelization and pipeline technology for real- time image processing. Additionally, the integration of processors for advanced image analysis and communication with external systems showcases a holistic approach to intelligent transportation systems.

Another noteworthy contribution, as highlighted in the pa- per, underscores the importance of vehicle tracking in challenging conditions, particularly in dark environments. The methodology involving grayscale information and robust structural features for continuous vehicle tracking is a significant stride in the pursuit of real-time applications. The proposed method's promising results suggest its potential for practical implementation.

The paper also introduces an adaptive sliding window strategy to enhance vehicle detection accuracy while concurrently reducing processing time. The adaptability of the sliding window approach, as described, showcases a substantial improvement in detection accuracy and computational efficiency when compared to conventional methods, thus substantiating its applicability in real-world highway scenarios.

The paper proceeds to outline an innovative approach for vehicle detection and tracking, anchored in the mixed Gaussian background model. By segmenting detection targets through various techniques and using location and colour information, the method aims to refine vehicle tracking. This methodology exhibits promise for addressing the complexities of real-world traffic scenarios.

Finally, the paper touches upon the importance of Region of Interest (ROI) extraction in machine vision and image processing. The explanation of different methods such as frame difference and background difference techniques for ROI extraction adds depth to the discussion, underlining the practicality of these approaches for efficient video coding and bandwidth utilization.

In summary, the research paper provides a comprehensive overview of the critical intersection between Intelligent Transportation Systems and machine vision technology for vehicle detection and tracking. It effectively conveys the importance of this field, underscores key research contributions, and intro- duces innovative methodologies, making it a valuable resource for professionals and researchers in the realm of intelligent transportation systems and computer vision [1, 2].

DEEP-VIOLENCE: INDIVIDUAL PERSON VIOLENT ACTIVITY DETECTION IN VIDEO

Anuja Jana Naik1,2 M. T. Gopalakrishna1,2, Multimedia Tools and Applications (2021) [3] The text describes a research project that focused on enhancing video surveillance with advanced human activity detection, particularly the detection of violent acts. In today's security-conscious world, with ever-evolving security issues in mind, it is extremely important to monitor and alert authorities to potential threats. The research addresses the complexity of recognizing human activity, including individual and group activity, and dealing with anomalies in crowded, noisy or dynamic environments. It delves into different identification methods, both hierarchical and non-hierarchical, emphasizing the importance of location and time information in accurately identifying actions. The method proposed in the study combines deep neural networks, key point detection, and long-short-term memory (LSTM) networks. The process starts with identifying individuals in video frames, followed by extracting key information and body shape. This data is then fed into an LSTM classifier to predict violent acts by considering spatial features. The study uses a variety of datasets to evaluate the system's performance, including well-known datasets such as KTH and Weizmann, as well as a custom dataset designed specifically for single- person violent acts. Comparative analyses show that the proposed model outperforms existing approaches, especially for the KTH dataset. The training activity at different time steps shows the model's learning curve, while the confusion matrices provide insight into its ability to discriminate between activity classes, even those with similar pose characteristics but different temporal

characteristics [4]. In addition, the AUC-ROC curve analysis illustrates the better performance of the model in classifying positive classes and highlights its performance in detecting violent acts. In addition, the study extends the analysis to multi-class ROC curves, which highlight the model's ability to effectively discriminate between different job classes. Overall, this research is an important step towards improving video surveillance capabilities using deep learning and spatio-temporal analysis to detect and classify human activity, with a special focus on violence detection. The observations and methods described in this study provide valuable information to the field of human activity detection and lay the foundation for further development of video surveillance technology.

VIOLENCE DETECTION IN VIDEOS USING INTEREST FRAME EXTRACTION AND 3D CONVOLUTIONAL NEURAL NETWORK

Javad Mahmoodi1 Hossein Nezamabad pour 1 Dariush Abbasi-Moghadam, Multimedia Tools and Applications [5]. In recent years, video content analysis has been closely related to machine vision and image processing techniques. Its applications have expanded to include various fields such as monitoring prisons, schools, parks and other public spaces. Two main applications of video content analysis are violence detection and action detection. Although many studies and datasets have been devoted to action detection, violence detection has received less attention until more recent efforts, such as hockey fight and action movie datasets. The emergence of such datasets and the availability of deep learning techniques and efficient computer systems have transformed traditional methods into deep learning methods. Traditional violence detection methods are often based on hand-crafted features and can be time-consuming, especially when using techniques such as Spatio-Temporal Interest Points (STIP) or Motion SIFT (MoSIFT) to extract interest points. Deep learning methods, on the other hand, allow generalization without prior knowledge of the data. They can outperform manual methods when applied to previously unseen data. Typically, deep learning methods use 3D Convolutional Neural Networks (3D CNN) and Long Short Memory (LSTM) networks to extract spatiotemporal features from videos. One of the main challenges in violence detection is to select regions of interest for feature extraction that are robust to background changes and object appearance. Traditional methods often use techniques like STIP or MoSIFT to do this. However, using deep learning techniques, researchers have developed more efficient approaches. In this context, the paper proposes a new deep neural network to detect violence in videos [6, 7]. This method involves extracting interests that are crucial to capturing violent behaviour. It introduces a special 3D CNN designed to detect whether previous frames contain violent or non- violent content. In addition, the spatial attention module is used to focus on specific areas in the frames. The proposed approach is evaluated on three benchmark datasets: hockey fight, fight movies and violent streams to demonstrate its effectiveness. This paper proposes a new technique to extract objects of interest, improve the performance of 3D CNN extracting objects of interest, improve accuracy by using a spatial attention module and show that combining 3D CNN with extracting frames of interest can improve the detection accuracy of violence The paper is structured with a review of related work in Section 2, followed by a detailed explanation of the proposed method in Section 3, and experiments and observations in Sections 4 and 5 [8, 9].

A FULLY INTEGRATED VIOLENCE DETECTION SYSTEM USING CNN AND LSTM

The development of an automated violence detection system utilizing deep learning methodologies, particularly a combi- nation of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. The primary objective of this system is to enhance public safety by identifying violent activities in real-time surveillance videos and thereby mitigating security threats.

In the introductory section, the text underscores the growing incidence of violence and the consequential threats it poses to society. It highlights the limitations of relying solely on human surveillance and advocates for automated systems to alleviate these challenges.

The text proceeds to delve into the role of technology in violence detection. It emphasizes the effectiveness of deep learning techniques, particularly CNNs, in classifying video sequences, and

references the utilization of well-known datasets such as Movies and Hockey in the realm of violence detection research.

The concept of transfer learning is introduced, wherein pre- trained neural networks like Xception are re-purposed for the task at hand. This approach is justified as an effective means of dealing with the challenges associated with large datasets.

The text then outlines the model architecture proposed for violence detection, which encompasses the combination of CNNs for spatial feature extraction and LSTM networks for sequence prediction. The Xception model is elaborated upon as a feature extractor, and the LSTM component is described as handling temporal aspects of the video data.

Datasets play a pivotal role, and the text highlights the usage of benchmark datasets like Hockey and Movies, in addition to the UCF Crime dataset, which is tailored for violence detection. Specific categories within the UCF Crime dataset, such as fighting, assault, abuse, and arrest, are mentioned, contributing to the violence classification task.

The overall system architecture is elucidated, emphasizing its integration with scalable and robust cloud services. This architecture includes a mobile application for real-time surveillance monitoring and alerts to relevant authorities, with PostgreSQL being the chosen database for storing critical information.

The results and analysis section presents the model's ac- curacy on benchmark datasets and the UCF Crime dataset. The text notes that models trained on benchmark datasets may not perform effectively in real-world CCTV footage scenarios, demonstrating the importance of datasets that mirror real- world conditions [10].

Lastly, the text addresses the real-time processing capabilities of the proposed model, highlighting its suitability for applications that require quick violence detection in video streams. It details the computational performance, indicating that the use of the UCF Crime dataset results in superior real-time analysis outcomes. Overall, the text presents a comprehensive overview of the violence detection system's development, underscoring its potential to enhance public safety and law enforcement efforts.

CONCLUSION

In conclusion, open deep learning integration with surveillance systems provides a strong instrument to improve public space security. We can create a more secure society and safer communities by advancing research in this area and resolving outstanding issues.

REFERENCES

1. Naveen P, Sivakumar P (2021) Adaptive morphological and bilateral filtering with ensemble convolutional neural network for pose-invariant face recognition. J Ambient Intell Human Comput 12:10023–10033.
2. Sarthak Sharma, B Sudharsan, Saamaja Naraharisetti, Vimarsh Trehan, Kayalvizhi Jayavel, Vol. 11, ISSN: 2088-8708, DOI: 10.11591/ijece.v11i4.pp3374-3380
3. Anuja Jana Naik1,2 & M. T. Gopalakrishna1,2, Multimedia Tools and Ap- plications (2021) 80:18365–18380, https://doi.org/10.1007/s11042-021- 10682-w, Published online: 17 February 2021
4. Parveen, S., & Shah, J. (2021). A Motion Detection System in Python and OpenCV. 2021 Third International Conference on Intelligent Com- munication Technologies and Virtual Mobile Networks (ICICV).
5. Javad Mahmoodi1 & Hossein Nezamabadi-pour 1 & Dariush Abbasi- Moghadam1, Multimedia Tools and Applications; Dordrecht Vol. 81, Iss. 15, (Jun 2022): 20945-20961. DOI:10.1007/s11042-022-12532-9

6. Dong-yuan Ge1 • Xi-fan Yao2 • Wen-jiang Xiang3 • Yue- ping Chen1, S.I.: Machine Learning and Big Data Analytics for IoT Security and Privacy (SPIOT 2021), Published: 01 February 2022
7. Ullah FU, Obaidat MS, Muhammad K, Ullah A, Baik SW, Cuzzolin F, Rodrigues JJ, de Albuquerque VH. An intelligent system for complex violence pattern analysis and detection. International Journal of Intelligent Systems. 2022 Dec;37(12):10400-22.
8. Ullah FU, Ullah A, Muhammad K, Haq IU, Baik SW. Violence detection using spatiotemporal features with 3D convolutional neural network. Sensors. 2019 May 30;19(11):2472.
9. Jain A, Vishwakarma DK. State-of-the-arts violence detection using ConvNets. In2020 international conference on communication and signal processing (ICCSP) 2020 Jul 28 (pp. 0813-0817). IEEE.
10. Vilamala MR, Hiley L, Hicks Y, Preece A, Cerutti F. A pilot study on detecting violence in videos fusing proxy models. In2019 22th international conference on information fusion (FUSION) 2019 Jul 2 (pp. 1-8). IEEE.

