Suspicious Action Detection and Recognition in Remote Areas Using AI and Machine Learning Techniques

Dhivya Karunya S[1]
Assistant Professor
S.E.A CET, Bengaluru
dhivyaskarunya@gmail.com

Krishna Kumar[2]
Professor
Gopalan College of Engineering and Management, Bengaluru

Abstract— The use of video monitoring to detect suspicious behaviours in public transportation zones is gaining popularity. For post-event analysis, such as forensics and riot investigations, automated offline video processing systems have been employed in general. However, there has been relatively little progress in the area of real-time event identification. We present a framework for processing raw video data received from a fixed colour camera set at a specific site and making real-time inferences about the observed activities in this research. First, using a real-time blob matching method, the proposed system gets 3-D object-level information by recognising and tracking individuals and baggage in the scene. Using object and interobject motion characteristics, behaviours and events are semantically detected based on the temporal aspects of these blobs. These supervised machine learning techniques are used to detect and track social distancing between one or more people's movements in public spaces, and these observations can be made using CCTV footage. To show the potential of this technique, a variety of sorts of behaviour that are significant to security in public transportation locations have been chosen. Abandoned and stolen items, fighting, fainting, and loitering are examples of these. The experimental findings reported here illustrate the approach's outstanding performance and minimal computing complexity using common public data sets.

Keywords: Abandoned luggage, behavior recognition, blob matching, fainting, fighting, loitering.

I. INTRODUCTION

To make their jobs easier, police and security personnel now rely on video surveillance equipment. Large public transit hubs, such as metro stations and airports, are particularly prone to this activity. However, these systems are still labor-intensive, and the individuals in charge of monitoring the video displays find it difficult to pay attention to accidents that occur at random [1], [2]. Although automated video surveillance systems exist, they have primarily been employed for offline video processing following an occurrence, most notably in riot investigations and forensics. These surveillance technologies are only marginally useful for real-time notifications at the moment. Furthermore, contrary to the media's and film industry's portrayals, research in this young yet promising sector has made little progress thus far.

An automated surveillance system's purpose is to alert monitoring employees to the occurrence of a user-defined suspicious behaviour as it occurs. The development of completely automated behaviour identification faces two problems. First, objects of interest in a scene, such as people and baggage, must be located, categorised, and monitored throughout time. Second, a consistent way of describing occurrences must be discovered. This is particularly problematic for complicated events with several alternative permutations, such as combat. They are undeniably tough to define in many instances.

The main drawback of machine learning is that it requires dependable standard data sets for training and testing. These are incredibly tough to come by, especially for unusual behaviour. When choosing classifier parameters and thresholds, this is a critical consideration. The semantic method, on the other hand, eliminates the necessity for training in favour of a more basic procedure based on human reasoning and logic. This, we believe, is a more reasonable and viable way. It, for example, eliminates the need to specify sophisticated learning parameters like decision-tree pruning thresholds, which are difficult to calibrate and require the assistance of specialists in the area. These are replaced by more obvious and meaningful parameters in the semantic approach. This study assumes that foreground blobs be retrieved using a standard background subtraction approach in each frame. These blobs are the semantic entities linked with the events described, and they represent the silhouettes of living (e.g., humans) and inanimate (e.g., baggage) items in the scene. In practise, however, we find that a single blob frequently represents many objects that are occluding or standing close to one other. Following the extraction of all blobs, inferences are made to segment, track, and categorise the objects they represent. Finally, the strange occurrences must be identified.

Object Tracking and Classification
We employ a Lab-based codebook background subtraction approach to separate the blobs of all foreground silhouettes

given an RGB video frame. Obviously, each blob does not necessarily represent a single semantic object, as is widely known. A group of them, for example, may obstruct each other in the picture and create a single blob from the camera's perspective. By matching these blobs in successive frames, objects representing semantic entities in the scene are located and tracked.

Object Modeling and Blob-to-Object Matching

This object tracking method is based on Tavakkoli et al[36] .'s work. A list of objects is updated at each frame by comparing blobs in the current frame to objects from the previous frame. This matching procedure isn't always one-to-one. To assure a valid update, cases of object splits, merges, one-to-one matches, creation, and deletion are all evaluated.

A concept of dependability is borrowed from [31] to reduce the confusion produced by the formation of misleading blobs by background subtraction. Because it is considered that items that do not remain long enough (about 1–3 s) after initially being recognised relate to noise or clutter, this idea mandates the inhibition and quick discarding of objects that do not last long enough (roughly 1–3 s).

II. PROPOSED WORK

Speech acquisition, feature extraction at each timeframe level, machine learning for each feature set, and information fusion to combine the data are the four primary components of the system. The system's main principle is depicted in Figure 1.

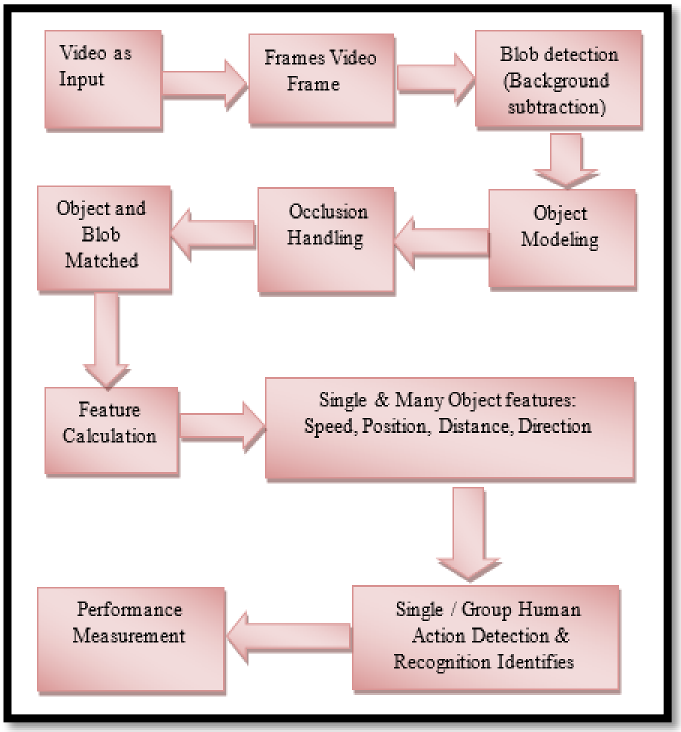

Video as Input
Frames Video Frame
Blob detection (Background subtraction)
Object Modeling
Occlusion Handling
Object and Blob Matched
Feature Calculation
Single & Many Object features: Speed, Position, Distance, Direction
Single / Group Human Action Detection & Recognition Identifies
Performance Measurement

Fig 1: System Architecture

Occlusion Handling

Because it affects the robustness of object tracking and coherence, occlusion handling is a crucial task. If occlusion is resolved wrongly, the inferences that follow will almost always result in a misunderstanding of the scene. We suggest, in agreement with [30] and [33], that locating the exact position of objects involved in occlusion inside a single blob requires an extensive search that is both computationally costly and wasteful. This is because blob-level localization offers enough spatial information to determine the object's position. As a result, we regard a blob's location to be the exact position of all of its constituent items. The question of which items are occluding which is largely neglected in this study, and we take the view that all merged objects form a pool (the blob) with no specific occluding/occluded connections.

In a nutshell, we turn the occlusion phenomena into a split/merge issue. Furthermore, we use the idea of potential occlusion [28], which allows an item that hasn't been clearly linked with any of the splitting blobs to be associated with all of the accompanying splitting blobs until resolution is achievable. [?] has a video that demonstrates this concept. Of course, this might result in erroneous temporary data characterising an object's location. Adaptive updating of the colour appearance model is disabled during occlusion to avoid the colour model of an object from becoming contaminated.

Evaluation

object tracking system was found to be quite reliable when using the aforementioned strategies. This is seen in a number of experiments conducted on publicly available data sets. Smooth tracking and occlusion management are demonstrated in instances [44] and [46]. The videos in Section VII further show that, despite not relying on learning approaches, our system produces reliable behaviour recognition.

It's worth mentioning that, despite our approach's robustness, problems like lost tracks and object misunderstanding are unavoidable. However, this technique was able to successfully track individuals and their baggage in the majority of tests conducted on a variety of standard data sets, even when three or four occluding objects were present.

III. FEATURE CACULATION:

Following the identification of the video's items of interest, their 3-D motion properties are computed, and a historical record is constructed. Objects are classed as either alive (people) or inanimate (things) based on this data. This categorization procedure is crucial since it is required for the definition of semantic behaviour. Many possible characteristics have been considered in the literature [14], [15], and [48]. Single-object information, such as location, and interobject features, such as alignment between two objects, can be separated.

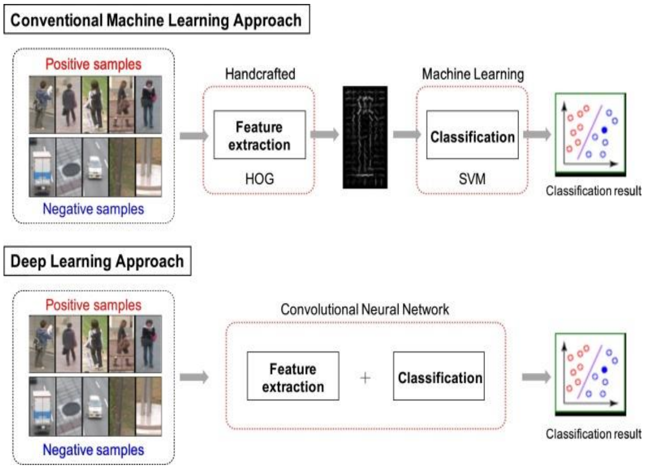


Fig2. The conventional Machine Learning and Deep Learning approaches for single or group human action detection and recognitions

These characteristics are assessed in real-world 3-D spatial coordinates, which may be derived using any typical camera calibration method from image (pixel) coordinates. By applying the transformation to the pixel positions of the feet, the position of an item is computed, from which practically all of its other attributes are derived. The lowest pixels of the 2-D blob to which the item belongs are simply referred to as these.

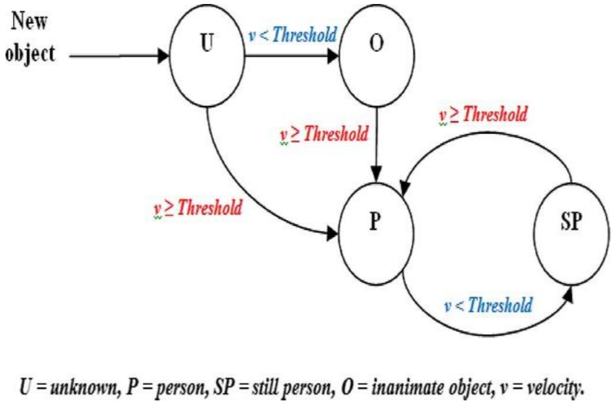


Fig. 3. Object classification state diagram.

A new item is classed as unknown when it appears in the scene for the first time. The velocity is utilised to assess whether it is a person or an inanimate item when its motion data are captured. Using this transition model, a stationary person will not be mistaken for luggage. The state diagram for the implemented method is shown in Fig. 3, which is essentially modified from [12] with a few minor changes. The system distinguishes between an inanimate item and a motionless human, a distinction that has significant implications for comprehending the scenario.

Abandoned and Stolen Objects

The identification of abandoned baggage has been a key topic in the literature to date. Detection has traditionally relied solely on background subtraction approaches, such as those described in [16] and [17], with no further types of reasoning such as object categorization or tracking. The difficulty is that such a method can't tell the difference between a stationary human and an abandoned object. Color, edges, shape completeness, and histogram contrast are used in other methods [51]. None of them have proven to be sufficiently resilient against noise and pose variations in our experience. Furthermore, the problem of locating the object's owner is still unsolved. This is critical for distinguishing between stolen and recovered luggage, for example. Using a semantic definition, this work overcomes the aforementioned flaws.

Meeting and Walking Together

Meeting and strolling together may be advantageous in some surveillance settings, despite the fact that it is not normally regarded suspicious. This would be especially true if face recognition were to be included as a feature. Individuals who meet with a suspect individual, for example, may need to be flagged for security reasons. Both events are semantically defined in Table III in terms of each person's speed, distance between them, and alignment.

III. EXPERIMENTAL RESULTS

A variety of challenges at several levels make evaluating behaviour recognition tests challenging [60]. First, most interesting tasks are complicated, which poses a problem in the test scenario when there is clutter. Another challenge is the scarcity of professional, tough, high-quality data sets for testing currently accessible. Furthermore, performance evaluation criteria such as a standard measure, hit-and-miss weighting, and ground truth construction are still up for debate. As a result of these difficulties, experimental results in different articles in the literature are inconsistent.

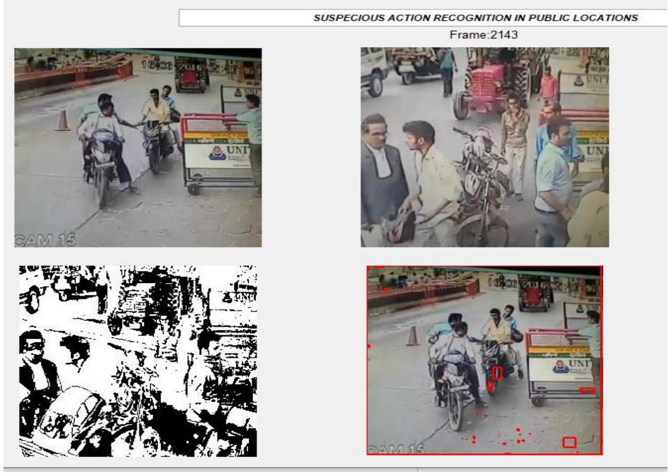


Fig 4. Illustrating the Real time fighting video in public area using image processing

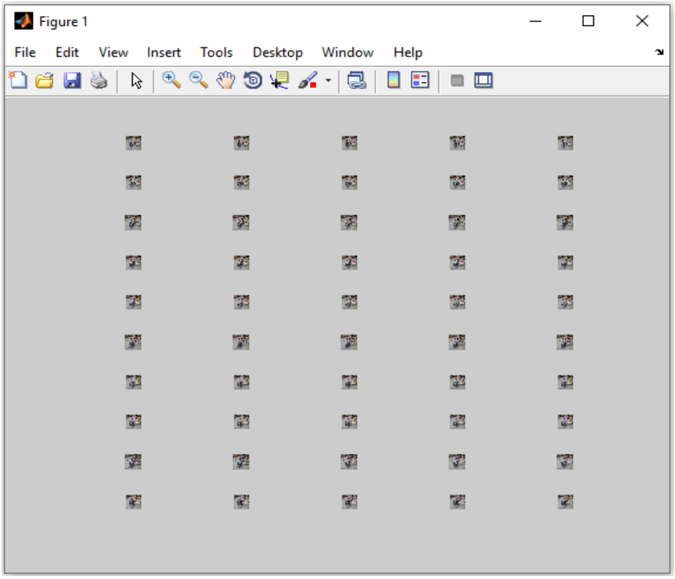
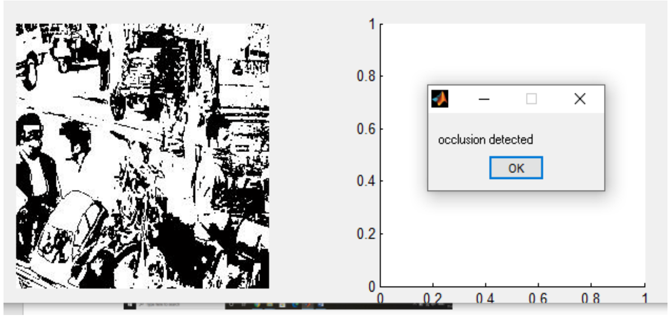


Fig 5. Illustrating the process of converting video into frames of Real time fighting video in public area using image processing



occlusion detected

Fig 5. Illustrating the blobs detection and Occlusion detection outputs of Real time fighting video in public area using image processing.

These supervised machine learning approaches are used to identify and monitor social separation between one or more people's movements in public spaces, and these observations may be made using CCTV footage. From this observation, the system will square mark as green for obeying social distance regulations, and if two or more people are close together, the system will square mark as red, indicating that the people moving in that location are breaching the social distance depicted in fig.5.

IV. CONCLUSION

Parameter tweaking, we note, may be compared to the problem of undertraining in machine learning since both indicate a lack of knowledge. The semantic method, on the other hand, has the benefit of allowing human reasoning to readily represent parameter values (e.g., speeds, distances, and angles). In contrast, collecting enough large and useful data sets for training machine learning algorithms is challenging. Learning, of course, necessitates fine-tuning of parameters like neural network size, connections, and learning parameters. Finally, present machine-learning techniques and systems based on semantics appear to be unable to generalise.

This work introduces and investigates a comprehensive semantics-based behaviour identification system that relies on object tracking. The first step in our method is to convert backdrop segmented objects into semantic entities in the scene. These items are monitored in two dimensions and categorised as alive (people) or inanimate (things) (objects). This technology provides real-time performance, flexibility, resilience against clutter and camera nonlinearities, ease of interaction with human operators, and the removal of the training required by machine-learning-based systems. Experiments were conducted on a variety of typical publically available data sets with varying crowd density, camera angle, and lighting conditions. The results of the experiments showed that the numerous actions of interest were successfully detected.

REFERENCES

1. Dr. H S Mohana and Mahanthesha U, “Human action Recognition using STIP Techniques”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-7, May 2020
2. J. F. Allen, “Maintaining knowledge about temporal intervals,” Commun. ACM, vol. 26, no. 11, pp. 832–843, Nov. 1983.
3. C. Fernandez, P. Baiget, X. Roca, and J. Gonzalez, “Interpretation of complex situations in a semantic-based surveillance framework,” Image Commun., vol. 23, no. 7, pp. 554–569, Aug. 2008.
4. J. Candamo, M. Shreve, D. B. Goldgof, D. B. Sapper, and R. Kasturi, “Understanding transit scenes: A survey on human behavior-recognition algorithms,” IEEE Trans. Intell. Transp. Syst., vol. 11, no. 1, pp. 206–224, Mar. 2010.
5. Y. Changjiang, R. Duraiswami, and L. Davis, “Fast multiple object tracking via a hierarchical particle filter,” in Proc. 10th IEEE ICCV, 2005, vol. 1, pp. 212–219.
6. A. Loza, W. Fanglin, Y. Jie, and L. Mihaylova, “Video object tracking with differential Structural SIMilarity index,” in Proc. IEEE ICASSP, 2011, pp. 1405–1408.
7. D. Comaniciu, V. Ramesh, and P. Meer, “Kernel-based object tracking,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 5, pp. 564–577, May 2003.
8. V. Papadourakis and A. Argyros, “Multiple objects tracking in the presence of long-term occlusions,” Comput. Vis. Image Underst., vol. 114, no. 7, pp. 835–846, Jul. 2010.
9. Mahanthesh U, Dr. H S Mohana “Identification of Human Facial Expression Signal Classification Using Spatial Temporal Algorithm” International Journal of Engineering Research in Electrical and Electronic Engineering (IJEREEE) Vol 2, Issue 5, May 2016
10. NikiEfthymiou, Petros Koutras, Panagiotis, Paraskevas, Filntisis, Gerasimos Potamianos, Petros Maragos “Multi-View Fusion for

Action Recognition in Child-Robot Interaction": 978-1-4799-7061-2/18/$31.00 ©2018 IEEE.

11. Nweke Henry Friday, Ghulam Mujtaba, Mohammed Ali Al-garadi, Uzoma Rita Alo, analysed "Deep Learning Fusion Conceptual Frameworks for Complex Human Activity Recognition Using Mobile and Wearable Sensors": 978-1-5386-1370-2/18/$31.00 ©2018 IEEE.
12. Van-Minh Khong, Thanh-Hai Tran, "Improving human action recognition with two-stream 3D convolutional neural network", 978-1-5386-4180-4/18/$31.00 ©2018 IEEE.
13. Nour El Din Elmadany , Student Member, IEEE, Yifeng He, Member, IEEE, and Ling Guan, Fellow, IEEE ,"Information Fusion for Human Action Recognition via Biset /Multiset Globality Locality Preserving Canonical Correlation Analysis" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 27, NO. 11, NOVEMBER 2018.
14. Pavithra S, Mahanthesh U, Stafford Michahial, Dr. M Shivakumar, "Human Motion Detection and Tracking for Real-Time Security System", International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 5, Issue 12, December 2016.
15. Lalitha. K, Deepika T V, Sowjanya M N, Stafford Michahial, "Human Identification Based On Iris Recognition Using Support Vector Machines", International Journal of Engineering Research in Electrical and Electronic Engineering (IJEREEE) Vol 2, Issue 5, May 2016
16. RoozbehJafari, Nasser Kehtarnavaz "A survey of depth and inertial sensor fusion for human action recognition", https://link.springer.com/article/10.1007/s11042-015-3177-1, 07/12/2018.
17. Rawya Al-Akam and Dietrich Paulus, "Local Feature Extraction from RGB and Depth Videos for Human Action Recognition", International Journal of Machine Learning and Computing, Vol. 8, No. 3, June 2018
18. V. D. Ambeth Kumar, V. D. Ashok Kumar, S. Malathi, K. Vengatesan and M. Ramakrishnan, "Facial Recognition System for Suspect Identification Using a Surveillance Camera", ISSN 1054-6618, Pattern Recognition and Image Analysis, 2018, Vol. 28, No. 3, pp. 410–420. © Pleiades Publishing, Ltd., 2018.

AUTHORS PROFILE

Dhivya Karunya S, is currently working as Assistant Professor in the department of Electronics and Communication Engineering and a research scholar at S.E.A CET, Bangalore under Visvesvaraya Technological University, Belagavi. She has done her bachelor's in Electronics and Communication Engineering at Sengunthar Engineering College Tamilnadu, affiliated to Anna University and her Masters in Digital Communication and Networking at S.E.A CET, Bangalore affiliated to Visvesvaraya Technological University, Belagavi. She has a rich experience of 13 years in Teaching. Her area of interest includes Signal Processing, Communication and Networking, Digital Electronics and Communication, Artificial Intelligence, Machine Learning. She is a member of ISTE.

Dr. Krishna Kumar is working as a professor at Gopalan College of Engineering and Management in the department of Electronics and Communication Engineering has done his Bachelor's in Electronics and Communication Engineering from S.J.C E Mysore, affiliated to University of Mysore and his Masters in Network communication and security from Dr M.G.R Educational and Research Institute Chennai and obtained doctorate from Techno global University, shillong. He has a rich experience of 22 years in teaching and research. He has worked at various capacities in affiliated universities. He has published more than 15 research papers. His area of interest includes Communication and networking, video processing, Machine learning, Artificial intelligence, Internet of Things, Data Science. He is also a member of ISTE.