Suspicious Counter Flow Detection For Real Time Video Surveillance

Abstract

The proposed system aims to detect suspicious counter flow in real-time video surveillance using a combination of YOLO (You Only Look Once) object detection, Hiding Time Kalman Filter (HTKF) and Probabilistic Data Association (PDA). First, YOLO is used to detect and track objects in the video feed. Then, HTKF is employed to estimate the hidden states of the objects, such as their positions, velocities, and directions. The HTKF model is able to handle occlusions and provide reliable estimates even when the object is temporarily hidden from view. Next, PDA is applied to associate the estimated object tracks with their corresponding identities. This helps to maintain object continuity and prevent false alarms from objects that enter the scene. Finally, the system uses a set of predefined rules to determine if a detected object is moving in the wrong direction, indicating suspicious counter flow. If such behavior is detected, an alert is triggered for further investigation. The proposed system offers a robust and efficient approach for detecting suspicious counter flow in real-time video surveillance systems, which can be used in a variety of applications such as traffic monitoring and security surveillance. In this article, we are proposing an novel and efficient approach to detect counter flow for humans in public or private places using tangent analysis of motion. The performance of the proposed method is 14 to 10 frames per second in dense regions (public traffic 35 to 60), 20 to 14 frames per second in moderate regions (public traffic 20 to 35) and more than 20 in sparse regions (public traffic < 35) in an Intel Core i5 cpu.

Introduction

Counterflow movement in an area refers to the movement of people who are walking or moving in the opposite direction to the majority of the crowd. This can occur in situations such as concerts, festivals, or religious gatherings, where there are large crowds of people moving in different directions. Counterflow movement in a crowded area can be difficult to navigate and may lead to congestion, confusion, and even accidents. In order to manage counterflow movement in a crowded area, organizers may employ a variety of strategies, such as setting up barriers or using signage to direct people towards the appropriate entrance or exit, or employing crowd control staff to monitor and manage the movement of people.

In addition, technology such as surveillance cameras and computer vision systems can be used to detect and analyze counterflow movement in the surveillance area, and provide real-time feedback to organizers and authorities to help manage and control the flow of people. These technologies can help improve safety and security in the surveillance areas and ensure that people can move efficiently and safely, even in situations where counterflow movement may occur. Deep learning based video surveillance is attracting the researchers to analyze the different notorious activities using automatic video surveillance systems. The cost of CCTV surveillance system has been reduced significantly as a result we can see video surveillance systems installed in various common public as well as private spaces like hospitals, multiplex, shopping malls, banks, schools and colleges etc. The CCTV surveillance system is actively playing an important role in the protection and management of public safety and preventing criminal activities [1]. Most of the existing CCTV surveillance is monitored by security personnels manually [2] which makes it cost inefficient. It also faces various difficulties in identifying and reporting suspicious activities for example security personnels are untrained and didn't have an exact idea of suspicious activities or proper training is not given to the concerned person. In order to design an efficient automatic suspicious counter flow detection for real-time video surveillance system, we need an efficient object detection model and an efficient spatiotemporal object tracking system. The convolutional neural network based models [3] are highly efficient in object detection as well as localization. YOLO (You Only Look Once) [4] is a very fast and accurate object detection and localization model, YOLO [4] divides an image into a number of grids of cells, it uses a single network to predict the coordinates of bounding boxes and object class probabilities. Faster-RCNN [5] is another deep convolutional neural network architecture that is highly accurate in object detection, this model can be divided into two components first the region proposal network [5] and other Faster-RCNN network [5]. Faster-RCNN is able to predict objects in complex images however the speed is the major limitation which restricts for real-time surveillance systems. Single shot multibox detector (SSD) [6] is another popular approach that uses the CNN architecture for object detection and localization. It uses a single network for object detection and localization. The SSD [6] predicts the class labels and the coordinates of bounding boxes of objects in an image by using a number of layers that produces a set of prediction maps at different scales which make it to handle different sizes of objects in same image and making it a good object detector in complex and cluttered scenes. The speed of SSD [6] is better than Faster-RCNN [5] but less than YOLO [4]. Once objects in an image are identified it is needed to be tracked over time so that the suspicious counterflow activity can be identified and security threats can be protected. To track objects over a sequence of time we need to keep track of object motion. There are a number of object tracking systems proposed in the literature. Correlation based visual tracking [7] have proven to be effective in visual tracking. Correlation based visual tracking [7] is a fast and efficient way to track objects in spatiotemporal data i.e. video. It is very efficient in handling illumination, rotation and deformation [7]. The Kalman Filter [8] is one of the most widely and commonly used estimation algorithms that is frequently used in fast and accurate object tracking. The Kalman filter uses this state representation to estimate the object's location and velocity at each time step, taking into account the measurement noise and other sources of uncertainty. One of the key advantages of Kalman filter-based object tracking is its ability to handle complex scenes, even in the presence of occlusions, camera motion, and other challenging conditions. The Kalman filter [8] is able to estimate the state of an object even in the presence of measurement noise and other sources of uncertainty, making it a robust and flexible solution for object tracking. Another advantage of Kalman filter-based [8] object tracking is its ability to track non-rigid objects, such as humans and animals. The Kalman filter can be used to estimate the motion of these objects even when they are deforming, providing a flexible and accurate solution for tracking these objects. Optical flow [9] is a method of tracking the movement of pixels in a video stream. Optical flow-based tracking algorithms use optical flow to estimate the motion of an object, allowing it to be tracked over time. Optical flow-based object tracking is a computer vision technique that uses the movement of pixels in a video stream to track the movement of objects. The idea behind optical flow is to estimate the movement of pixels between consecutive frames and use this information to estimate the motion of objects in the video. In this article, we have combined the YOLO [4] based object detection and Kalman filter [8] based object tracking along with our proposed mathematical model that can identify the suspicious counterflow of humans in the surveillance region.

Literature review

In this section we are going to review some recent techniques used in the detection of suspicious human activity. Brostow, G. J., & Cipolla, R. (2006) [23] have proposed a Bayesian approach for detecting independent motion in crowded scenes. The authors [23] addressed the problem of detecting abnormal behavior in crowds, such as sudden movements or changes in direction, that may indicate potential threats. The proposed approach uses a hierarchical Bayesian model to estimate the joint distribution of the velocity and direction of the crowd's motion. The model assumes that the motion of each individual in the crowd is independent, and the joint distribution is modeled as a mixture of Gaussian distributions. The model is unsupervised, meaning that it does not require any training data or prior knowledge about the behavior of the crowd. The paper presents experimental results on several datasets, including the PETS2009 dataset and the UCSD pedestrian dataset. The results show that the proposed approach can accurately detect abnormal behavior in crowded scenes, such as sudden movements or changes in direction. The approach also outperforms several baseline methods, including a simple threshold-based method and a method based on optical flow. Since the system uses the Multi-modal background modeling to track candidate objects which suffers from computational inefficiency [24] another limitation is its sensitivity to the selection of the number of Gaussian distributions used in the model. If too few Gaussians are used, the model may not capture the full complexity of the background, leading to inaccurate foreground detection [25]. On the other hand, if too many Gaussians are used, the model may become overly complex and may not generalize well to new scenes. Multi-modal background modeling can also be sensitive to changes in the environment, such as changes in illumination, weather, or scene geometry. In some cases, these changes may cause the model to become less accurate, leading to increased false positive or false negative detections [25].

Gnanavel, V. K., and A. Srinivasan [10] have proposed an approach to detect abnormal events in crowded places. The authors [10] have designed a multi-stage architecture to identify suspicious events in crowded areas, in the first stage the image is divided in small patches. In the second stage, the difference of gaussians (DoG) filters has been applied on patches of two sequences of images to extract edges i.e. two consecutive images of video frames are divided into disjoint spatial patches and then temporal difference of gaussian between the patches of consecutive frames are computer to find out edges. Then in the third stage, a multi-scale histogram of optical flow is computed along with an edge oriented histogram for each patch that we got from the difference of gaussians in the second stage. In the fourth stage, normalized cuts along with gaussian expectation maximization is used to make clusters of patches. In this stage, each cluster is assigned with a motion context. In the final stage, exploitation based KNN search has been performed to establish the difference between abnormal and normal activity at intervals of crowded scenes. The major drawback with this system is computational complexity and false zero crossing [14] which makes it unsuitable for real-time suspicious counterflow detector.

Wu, Ziyan, and Richard J. Radke [11] have proposed a method to identify suspicious human motion from low resolution cameras (320 x 240 pixels) in crowded places like shopping malls, railway stations, stadiums etc. Authors [11] have fast corner detectors to identify low level spatial features in a video frame and the low level spatial properties are tracked into temporal sequence using Kanade-Lucas-Tomasi (KLT) optical flow algorithm. The authors [11] have adapted the pyramid representation to keep track of large pixel motion by taking care of small integration windows. Optical flow based methods have a number of limitations like they are sensitive to occlusion, illuminations and texture variations at the same time they are also computationally intensive for long video sequences which is very obvious in real-time surveillance systems.

Fernández Rodríguez, Jose David, et al. [12] has proposed a method for automatic anomalous trajectory detection of objects in traffic video surveillance. Counterflow detection in traffic surveillance video is well known suspicious activity that violates traffic rules and a potential threat to law & order. The authors [12] have used a pre-trained deep convolutional network to identify vehicles in video frames. To track vehicles in temporal sequence, a linear sum assignment technique is used to identify similar objects in two consecutive frames. In order to compute vehicle motion direction, the euclidean distance between the centroids of similar vehicles along with direction is measured. In a few time steps the trajectory of the vehicle is determined and then finally behavior of trajectory is evaluated as per previous majority vehicle flow direction or some predefined traffic rule. The time complexity of linear sum assignment is $O(n^3)$ as well as it is sensitive to outliers [16].

Proposed Method

In this section, we are presenting our approach to identify the counter flow of humans in surveillance areas. In the proposed method, we first identify the persons in the video frame using YOLOV5 [13] one class person detection model then we will keep track of person movement using Kalman filter [8]. The use of YOLO comes with many advantages like i) Fast and efficient: YOLO processes images in real-time with high accuracy, making it ideal for applications that require quick detection, such as self-driving cars or surveillance systems, ii) End-to-end training: YOLO is an end-to-end system that can be trained on a single dataset, which makes it simpler to train and easier to use than other object detection algorithms that require multiple stages of training and fine-tuning, iii) Good generalization: YOLO performs well on a variety of object detection tasks and can detect multiple objects in an image simultaneously, making it versatile for a wide range of applications and iv) Low memory usage: YOLO has a smaller memory footprint than other object detection algorithms, which makes it more suitable for deployment on low-power devices such as edge devices and embedded systems [17][18]. Another major component in our proposed system is Kalman filter [8], the benefits of Kalman filter for real time object tracking are i) Accurate and robust: The Kalman filter is a powerful algorithm that can estimate the state of an object with high accuracy, even in the presence of noise, occlusions, and other disturbances, ii) Real-time tracking: The Kalman filter can track objects in real-time, making it suitable for applications that require fast and continuous tracking, such as video surveillance, robotics, and autonomous vehicles, iii) Predictive tracking: The Kalman filter can predict the future position and velocity of an object based on its previous states, allowing for smooth and natural tracking, even when the object is moving unpredictably, iv) Adaptive filtering: The Kalman filter can adapt to changes in the object's motion or environment, making it robust to dynamic and complex scenes, v) Efficient computation: The Kalman filter is computationally efficient, which allows for real-time tracking on resource-limited devices like edge and embedded devices [19] [20].

Tracking an object during occlusion using a Kalman filter can be challenging since the object is not directly visible to the sensor during the occlusion. One approach is to use an augmented state vector that includes an additional component that represents the object's occluded state. The filter predicts the object's state during the occlusion period using the dynamics model and previous observations. When an observation becomes available, the filter uses it to update its estimate of the object's state, which includes the occluded component.

Humans generally do not walk in a zigzag pattern during normal walking. When walking, humans tend to move forward in a relatively straight line, with minor variations in direction to avoid obstacles or adjust their path to reach their destination. However, there are some situations in which humans may walk in a zigzag pattern, such as when navigating a crowded area, trying to avoid obstacles or when walking on an uneven surface or in counterflow movement. Zigzag movement in a surveillance area can be an indication of potentially suspicious behavior. Zigzag movement may indicate that a person is attempting to avoid

detection or surveillance, or may be trying to approach a target in a non-linear fashion to avoid being seen. There are several factors that can help identify suspicious zigzag movement in a surveillance system, including: i) abnormal speed: a person moving in a zigzag pattern may do so at an unusual speed or pace, which can be an indication of suspicious behavior, ii) repeated patterns: if a person is observed moving in the same zigzag pattern repeatedly, this may be an indication of suspicious behavior or activity, iii) unusual changes in direction: zigzag movement may involve abrupt or unusual changes in direction, which can be an indication of suspicious behavior or evasion and iv) avoiding surveillance: if a person is observed moving in a zigzag pattern in an area where surveillance cameras are present, this may be an indication that they are attempting to avoid detection or surveillance.
In our proposed approach of suspicious counterflow detection we have handled following situations to detect suspicious counterflow:

1. ***Normal Movement in Counterflow Direction:*** Any normal movement in counterflow direction either having displacement or staying time more than the threshold is considered as suspicious counterflow movement.
2. ***Zigzag Movement:*** Assuming that humans generally do not walk in a zigzag pattern during normal walking, any zigzag movement having displacement in counterflow direction is considered as counterflow implies either displacement or staying time is more than the specified threshold.
3. ***Temporarily Hide during Movement:*** The Kalman filter is a powerful tool for tracking objects over time, even when the objects are temporarily hidden from view. To track an object during a hide using a Kalman filter, the filter needs to be modified to account for the fact that no observations are available during the hidden period. The approach we have used is a modified Kalman filter algorithm called the Hiding Time Kalman Filter (HTKF) [21], which models the time between observations as a random variable with a probability distribution. The HTKF estimates the hidden state of the object between observations by incorporating the hiding time distribution into the filter equations. When an observation becomes available, the HTKF updates its estimate of the object's state using the observed data and the estimated state during the hidden period.
4. ***Partial Occlusion during Movement:*** Partial occlusion is handled using a probabilistic data association (PDA) method [22], which uses a likelihood function to associate observations with potential tracks of the object. During the occlusion period, the filter maintains multiple potential tracks of the object and assigns probabilities to each track based on their consistency with previous observations. When a new observation becomes available, the filter updates the probabilities of the different tracks based on their likelihood, and selects the track with the highest probability as the current estimate of the object's state.

The threshold for displacement or staying time should be carefully chosen based on the specific context and environment. A high threshold may result in false positives, while a low threshold may result in false negatives. The threshold should be chosen based on a careful analysis of the expected flow patterns in the environment and the potential risks.

Algorithm (Counterflow detection)

1. Object detection using YOLO: The first step is to use YOLO (You Only Look Once) [13] object detection algorithm to identify the person in the scene.
2. Object tracking using Hiding Time Kalman Filter (HTKF) with probabilistic data association (PDA) method: Once the persons have been detected, the next step is to track their movement using a Kalman filter. The Kalman filter [8] is a mathematical algorithm that uses a series of measurements and predictions to estimate the state of a moving object. In the case of counter flow detection, the Kalman filter is used to track the movement of the objects and predict their future positions.
3. Direction estimation: After the objects have been tracked using the Kalman filter, the direction of their movement is estimated. This is done by comparing the predicted position of the object with its actual position. If the predicted position is ahead of the actual position, then the object is moving in the same direction as the traffic flow. If the predicted position is behind the actual position, then the object is moving in the opposite direction to the traffic flow.
4. Counter flow detection: Based on the direction estimation, the final step is to detect any objects that are moving in the opposite direction to the traffic flow. This can be done by setting a threshold value for the difference between the predicted and actual positions of the object. If the difference exceeds the threshold value, then the object is considered to be moving in the opposite direction to the traffic flow and is flagged as a potential counter flow.

The summary of our approach is whenever a new person is detected at any time instant a unique ID is given to the person and the centroid of each person is computed. At each time step the position of persons are updated using the Hiding Time Kalman Filter (HTKF) [8] [21] with probabilistic data association (PDA) method [22] and the number of steps since incarnation is also updated along with distance from initial position. If the distance traveled is more than the minimum allowed distance in the direction of flow and angle between the defined axis (typically a horizontal axis) and the direction flow from start position to current position is more than the allowed angle of flow then counter flow is detected.

Dataset Used

We have used a subset of Multiple Object Tracking (MOT) Benchmark [26] dataset having persons moving in one or two directions. It is a widely used benchmark dataset for evaluating the performance of multiple object tracking algorithms. The dataset includes several challenging video sequences with various scenarios, such as crowded scenes, occlusions, and changes in lighting conditions. The dataset includes annotations for each frame, providing ground truth information on the positions and identities of objects in the scene. The annotations include the bounding box coordinates of each object and a unique ID assigned to each object to track its movements across frames. Some selected youtube videos are also used to examine the performance of the proposed system.

Hardware and software resources required

In this section, experimental setup is explained.

1. CPU: Intel @ 2.00GHz
2. Operating System: Ubuntu 20.04.5 LTS

3. Memory: 16 GB
4. Python Version: 3.8.0
5. Libraries: Scikit-learn, Pytorch, Numpy, Pandas

Results and Discussion

The results of the proposed method to identify suspicious activity in public and private surveillance areas is good enough to use in real-time surveillance. The speed of YOLOV5 [13] one class model for person detection is 20 fps on CPU which is not only suited for real-time video surveillance but also highly cost efficient as it can run one CPU instead of GPU which is recommended by most of the video surveillance systems.The processing speed of the proposed system is constraint by the density of person moving in the surveillance area. The frame per second in dense areas having 35 to 60 persons is reported as 14 to 10, similarly frame per second in the medium occupancy area having 20 to 35 peoples is 20 to 40, however in low occupancy areas with less than 35 the frame rate is more than 20 on a conventional CPU Intel Core i5. The proposed method accurately identifies the suspicious counterflow for video surveillance because the backbone of the approach relies on benchmarked YOLOV5 [13] model for person detection, the use of HTKF with PDA has been shown to be effective for tracking moving objects in complex environments. The method is able to handle occlusion, partial visibility, and missed detections, which are common challenges in tracking. The method is also computationally efficient, making it suitable for real-time applications. However, it is important to note that the performance of the method can be sensitive to the quality of the detections. If the detections are noisy or inaccurate, the method may produce inaccurate tracking results. Therefore, it is important to carefully select and preprocess the detections to ensure high-quality input to the tracking algorithm and finally the proposed mathematical formula is very simple yet effective in the suspicious counterflow detection.

In future, we can extend this work on multi camera view where the images from different cameras are fused using image fusion technique to form a single image and then suspicious counterflow is detected in that environment setting.

References

1. Kim, Jin Su, Min-Gu Kim, and Sung Bum Pan. "A study on implementation of real-time intelligent video surveillance system based on embedded module." *EURASIP Journal on Image and Video Processing* 2021.1 (2021): 1-22.
2. Held, Cornelius, et al. "Intelligent video surveillance." *Computer* 45.3 (2012): 83-84.
3. Cai, Zhaowei, et al. "A unified multi-scale deep convolutional neural network for fast object detection." *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*. Springer International Publishing, 2016.
4. Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
5. Girshick, Ross. "Fast r-cnn." *Proceedings of the IEEE international conference on computer vision*. 2015.
6. Liu, Wei, et al. "Ssd: Single shot multibox detector." *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer International Publishing, 2016.
7. Danelljan, Martin, et al. "Convolutional features for correlation filter based visual tracking." *Proceedings of the IEEE international conference on computer vision workshops*. 2015.

8. Patel, Hitesh A., and Darshak G. Thakore. "Moving object tracking using kalman filter." *International Journal of Computer Science and Mobile Computing* 2.4 (2013): 326-332.
9. Bruhn, Andrés, Joachim Weickert, and Christoph Schnörr. "Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods." *International journal of computer vision* 61 (2005): 211-231.
10. Gnanavel, V. K., and A. Srinivasan. "Abnormal event detection in crowded video scenes." *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014: Volume 2*. Springer International Publishing, 2015.
11. Wu, Ziyan, and Richard J. Radke. "Improving counterflow detection in dense crowds with scene features." *Pattern Recognition Letters* 44 (2014): 152-160.
12. Fernández Rodríguez, Jose David, et al. "Anomalous trajectory detection for automated traffic video surveillance." (2022).
13. Jocher, G. (2020). YOLOv5 by Ultralytics (Version 7.0) [Computer software]. https://doi.org/10.5281/zenodo.3908559.
14. Sharifi, M., Fathy, M., & Mahmoudi, M. T. (2002, April). A classified and comparative study of edge detection algorithms. In *Proceedings. International conference on information technology: Coding and computing* (pp. 117-120). IEEE.
15. Fortun, D., Bouthemy, P., & Kervrann, C. (2015). Optical flow modeling and computation: A survey. *Computer Vision and Image Understanding*, *134*, 1-21.
16. Janson, B. N. (1991). Dynamic traffic assignment for urban road networks. *Transportation Research Part B: Methodological*, *25*(2-3), 143-161.
17. Du, J. (2018, April). Understanding of object detection based on CNN family and YOLO. In *Journal of Physics: Conference Series* (Vol. 1004, p. 012029). IOP Publishing.
18. Fang, Y., Guo, X., Chen, K., Zhou, Z., & Ye, Q. (2021). Accurate and Automated Detection of Surface Knots on Sawn Timbers Using YOLO-V5 Model. *BioResources*, *16*(3).
19. Swalaganata, G., & Affriyenni, Y. (2018, March). Moving object tracking using hybrid method. In *2018 International Conference on Information and Communications Technology (ICOIACT)* (pp. 607-611). IEEE.
20. Rameshbabu, K., Swarnadurga, J., Archana, G., & Menaka, K. (2012). Target tracking system using kalman filter. *International Journal of Advanced Engineering Research and Studies*, *2*, 90-94.
21. Jeong, J. M., Yoon, T. S., & Park, J. B. (2014, September). Kalman filter based multiple objects detection-tracking algorithm robust to occlusion. In *2014 Proceedings of the SICE Annual Conference (SICE)* (pp. 941-946). IEEE.
22. Gennari, G., & Hager, G. D. (2004, June). Probabilistic data association methods in visual tracking of groups. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* (Vol. 2, pp. II-II). IEEE.
23. Brostow, G. J., & Cipolla, R. (2006, June). Unsupervised bayesian detection of independent motion in crowds. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)* (Vol. 1, pp. 594-601). IEEE.
24. Trabelsi, R., Jabri, I., Smach, F., & Bouallegue, A. (2017). Efficient and fast multi-modal foreground-background segmentation using RGBD data. *Pattern Recognition Letters*, *97*, 13-20.
25. Langmann, B., Ghobadi, S. E., Hartmann, K., & Loffeld, O. (2010, September). Multi-modal background subtraction using gaussian mixture models. In *ISPRS Symposium on Photogrammetry Computer Vision and Image Analysis* (pp. 61-66).
26. Milan, A., Leal-Taixé, L., Reid, I., Roth, S., & Schindler, K. (2016). MOT16: A benchmark for multi-object tracking. *arXiv preprint arXiv:1603.00831*.