Alexandria Engineering Journal (2022) **61**, 9637–9647

HOSTED BY

ELSEVIER

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

A novel real-time multiple objects detection and tracking framework for different challenges

Nuha H. Abdulghafoor, Hadeel N. Abdullah *

Department of Electrical Engineering, University of Technology – Iraq, Baghdad, Iraq

Received 9 January 2021; revised 25 July 2021; accepted 28 February 2022

KEYWORDS

Multiple object detection;
Multiple object tracking;
Classification;
Deep-learning;
PCP

Abstract Recently, there was a lot of researches on real-time detection and tracking algorithms, as the frequent use of surveillance cameras and the expansion of its applications, especially in security and surveillance. However, many challenges have emerged that hinder monitoring systems' work, whether in the detection or tracking stage. We propose a robust new algorithm to detect and track objects from natural scenes captured with real-time cameras to achieve this. This work aims to create a detection and tracking algorithm that is responsive to actual and fundamental changes. This algorithm is characterized by the detection of multiple moving creatures, limited resources, and different challenges. This algorithm combines principal component analysis and deep learning networks to make the most of these two approaches' advantages to achieve an intelligent detection and tracking system that works in real-time. It is done adaptively between the two approaches to enhance performance compared to the existing detection and tracking algorithms. The experimental results showed the new algorithm's effectiveness and efficiency by comparing it with other detection and tracking systems and obtaining good detection and classification accuracy.

1. Introduction

In recent decades, surveillance, detection, and locating systems have expanded in many modern applications. Examples of these applications are security and traffic surveillance systems, medical applications, automated driving systems, etc. Numerous research and articles have also appeared using different approaches and methods to reach an effective system that works in real-time. With the proliferation of surveillance systems and the increase in CCTV cameras, it has become difficult for many people in control centers to work with the same efficiency throughout the day and control many cameras simultaneously. Consequently, the need for effective detection and tracking systems was required that works in real-time. It can

* Corresponding author.
E-mail addresses: nuha.h.abdulghafoor@uotechnology.edu.iq (N.H. Abdulghafoor), 30002@uotechnology.edu.iq, hadeel.n.abdullah@uotechnology.edu.iq (H.N. Abdullah).

☆ We confirm that the submission represents original work that has not been published previously and is not currently being considered or submitted to another journal until a decision has been made. Also, we confirm that each author has seen and approved the contents of the submitted manuscript.

Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2022.02.068
1110-0168

be used in the early detection of alien organisms, violent crime, and other applications [1].

The general block diagram of Object detection and tracking system, as shown in Fig. 1. The basic principle is feature extraction from the image or video frame in real-time or offline applications. The basic principle is to detect moving objects from the difference between the current frame and the reference frame, which is often called a "background-image" or "background pattern" [2]. There are many methods. The essential methods are background subtraction [3] and Principal Component Pursuit (PCP) algorithms [4,5]. Object modeling represents the object of interest, as it is done by extracting the features that uniquely define an object. After that, these features of an object are used to track that object. The feature can be defined as an image style that distinguishes a particular object from its surroundings, and then these features are transformed into specifications desired by the appearance features. Recently, some algorithms that use the principal component analysis (PCA) approach have shown clear superiority and superfast video processing speed [6]. As well as the development of algorithms, object discovery is based on a deep learning approach. Each of the above approaches has good advantages, and at the same time, it may fail at times in specific challenges. In this article, a new proposed algorithm to overcome these challenges.

The importance of big data analytics and computational intelligence techniques applied to the resulting data lies in the presence of several scattered connected devices, where the peripheral devices that provide embedded information processing capabilities and the possibility of using appropriate computational intelligence techniques for effective processing and analysis of big data [7] It is also possible to develop modern applications for designing intelligent cities in the real world through the use of these powerful and intelligent tools and techniques [8].

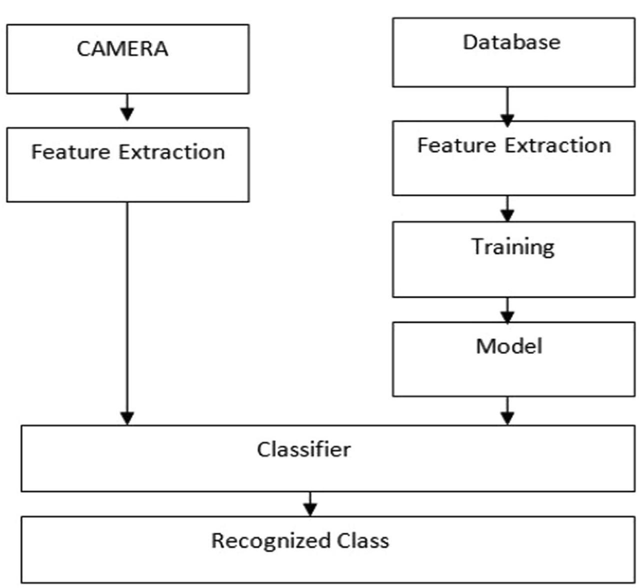

Fig. 1 The basic principles of object detection, classification, and tracking.

Due to the increasing number of surveillance cameras used worldwide, especially in critical cities, they will require constant monitoring by the operator. The cost increases with the increase in the number of operators and cameras. In addition to analyzing and storing these video clips, massive computers and high-quality servers must manage these cameras. Therefore, there was a need to build an intelligent detection and tracking system that works automatically. In addition, videos captured by cameras during real-time face various challenges, including the low resolution of the captured image, the presence of different noises, the appearance of shadows of moving objects, the different size and speed of objects, and the ability of the algorithm to distinguish them regardless of their size or speed. It may lead to a high computational cost [9], and it requires effective hardware and tools and large computational memory, which is sometimes challenging to provide. The main contribution is a new proposed algorithm to detect and track objects, where the algorithm consists of two stages as follow:

- The first stage is the detection stage, which also consists of two necessary steps. The pre-processing detection step is based on the PCP approach, which is done by capturing the video frame, removing noise, and detecting moving objects within the video frame by separating the object (objects detected) from the background (background scene). It is done with an adaptive hybrid framework between the two approaches to detect moving objects better. From the advanced processing step in the discovery stage, after identifying sites or areas of interest within the framework, the detected objects are identified and categorized using a deep learning approach.
- The second stage is tracking the detected objects and determining the identifier of each detected object, whereby the objects are tracked by linking the data of the object between the current frame and the previous frame adaptively between the appearance, location, and speed of the object. Some of the challenges in the process are also addressed, such as changing the speed of an object from the frame to frame, reducing switches for changing the object ID that is tracked during a video scene, and other challenges.
- It uses a dataset that includes multiple video clips, challenges, and actual video clips captured from a camcorder by implementing the proposed algorithm feasible on the integrated Nvidia intelligent TX2 Jetson platform.

It also develops information retrieval systems for software tasks. It can be investigated to what extent a detection domain research system can be developed to support activities related to their work [10]. The proposed system can be used in troubleshooting applications using deep learning. The approach was tested on data acquired on computer-based systems equipped with local and remote sensors [11]. The paper was arranged and organized in several paragraphs: A brief introduction about detection and tracking systems and the proposed algorithm. The second section will be similar or related works. The third section will be the intellectual background and basic principles of the proposed algorithm. After that, the proposed algorithm will be reviewed in detail. The fourth section is devoted to the experiment results and the tests conducted for the proposed algorithm and its comparison with the existing algorithm's results to evaluate its performance, effectiveness, and efficiency. As for the final section, we will present conclusions and discussions about the proposed algorithm.

2. Related work

The monitoring and surveillance systems generally consist of several steps: object detection and recognition and object tracking. It may also include distinguishing and analyzing the behavior of objects. Therefore, in the article [the stage of discovering things is vital for identifying areas of interest (ROI) [12], such as moving objects. After that, the objects are classified into predefined categories, such as humans and cars, representing part of recognizing the objects, and then these objects are tracked. The strange events are identified in the analysis of the behavior of these objects. Therefore, detection and tracking steps are essential for building a comprehensive and efficient monitoring system, as each stage depends on the previous stage and analyses its results.

Significant challenges affect the change in lighting and the camera angle, which leads to noise that affects the quality of the video captured in real-time [13]. Therefore, within specific environments and conditions, the proposed algorithms often exhibit failures when applied to real-time scenes, as in the system proposed earlier in his article. Most of the proposed systems and algorithms for detection and tracking technologies are designed to operate daily. Most critical events occur during the night, which calls for establishing and developing monitoring systems in different circumstances and environments. In the article [14], he proposed a system for pedestrians' detection and classification without other objects such as cars, for example. The proposed system shows a hybrid method between the background subtraction and bypass neural networks [3]. The foreground representing people is separated from the background representing the scene by using the background subtraction technique. Then comes the role of CNN in identifying and classifying people. The proposed system faced restrictions, including detecting people and without tracking them and analyzing their behavior.

Shaoqing et al. [15] proposed a detection algorithm that uses a deep learning network called (Fast R-CNN), characterized by accurate detection despite the high and slow computational cost and some limitations in weights. After that, Bochinski et al. [16] proposed a tracking algorithm based on the intersection over union principle, characterized by simplicity of work despite the presence of failures in tracking multiple objects. Then, Tijtgat et al. [17] proposed an algorithm designed for drones to detect and track objects in real-time using a computing platform embedded with a low-power camera. The proposed algorithm combines a deep learning network (Faster RCNN) with a single object tracking Kernelized Correlation Filters (KCF) approach. This algorithm shows accurate slow detection and fast-tracking despite the computational cost of detection and failure to track fast objects sometimes. Chandan et al. [18] represented moving objects as a set of points with spatial and temporal features using the Gabor 3D filter by analyzing each pixel in the sequential video and then linking them together using the Minimum Spanning Tree approach. Woke et al. [19] improved the previous algorithm by adding motion and appearance information to improve tracking accuracy by using a deep learning approach to detect objects and then using a Kalman filter with Hangareen's approach to tracking detected objects accurately.

Mahalingam and Subramoniam [20] proposed an algorithm split into three stages; the Foreground segmentation stage uses the Mixture of Adaptive Gaussian model, tracking step by using the blob detection and evaluation stage, which includes the classification according to the feature extraction. Blanco-Filgueira et al. [21] implemented a visual tracing of multiple objects based on deep real-time learning using the NVIDIA Jetson TX2 device. The results highlighted the algorithm's effectiveness under real challenging scenarios in different environmental conditions, such as low light and high contrast in the tracking phase and not consider in the detection phase.

We proposed an algorithm to detect and track objects. It integrates the Fast Principal Component Pursuit (FPCP) algorithm for detection with a Kalman Filter to implement a tracking system [2]. Then we evaluate its performance under its configurations and different suggestions. The detection and tracking system above has been improved by adding modern technologies in the detection phase, such as bypass the neuronal network and different methods in the tracking phase [5]. Then, we proposed a new algorithm for detecting multiple objects online or in real-time [6]. Simultaneously, it can reduce or remove different noise types that accompany the video sequence capture. The first proposed algorithm adopts the principle of separation between the low-rank model and the scattered one with noise removal, depending on the argument Lagrange multiplier (ALM) based on the Alternating Direction Method of Multipliers (ADMM) to solve the optimization reduction problem. The second proposed algorithm then adopts separating the background from the foreground and removing the noise by depending on spatial and spectral analysis using the spatial spread method. The work was enhanced by adding the Total Variation Regularization (TV) standard to both algorithms to give spatial smoothness and spatial softness to extract objects and preserve edge.

Recently, we proposed an efficient algorithm [22] to detect and track multiple objects and address some of the challenges that prevent good results and robust performance—training and testing different Deep learning Networks models of the detection stage. The proposed algorithm [22] was also implemented on the NVidia Jetson TX2 Platform. Calculation of essential detection and tracking performance factors such as the training and learning efficiency and the impact of the training sample sizes. Then, we utilized Deep Stream - Software Development Kit (DS-SDK) technology to improve the performance of object detection and tracking algorithm to make it able to process more than one capturing video simultaneously. This algorithm's hardware implementation on Nvidia Jetson TX2 Platform for multiple objects (such as cars and pedestrians) in real-time monitoring system applications [23]. Then, we extract useful information with high speed and accuracy.

A recent application is a short and long-range wireless communication essential for many remote monitoring applications such as industrial automation, digital health, and smart cities. Remote monitoring is a component of complex systems that can be improved using techniques based on machine learning. An intelligent parking system based on wireless technologies, computer vision and artificial intelligence is presented in this article [24]. It also deals with the problem of spatiotemporal inference in complex dynamical systems. It is based on the memory prediction framework, and deep neural networks (DNN) detect errors in automatic checks [25].

3. Basic principles (methodologies*)*

Modern detection techniques stemming from the in-depth learning approaches and Convolution Neural Networks have made significant computer vision development in general and detect and track algorithms [12] in particular. One of the advantages of using these technologies is detecting, distinguishing, and classifying the target. However, these technologies still suffer from some challenges, as they still need a training phase on a new data set, which requires a considerable calculation cost. It also suffers from the abundance of composition, which is the most problematic that hinders the application of video clips captured from real-time cameras, as they may be different in the accuracy and quality of the captured image, as well as the presence of noise or not, or changing the angle of view of the camera and other challenges. The type of camera also affects the detection quality and accuracy. Deep learning techniques may fail to detect objects of small size as well as background-looking objects. Bypass neural networks are also trained to detect one type without the other, such as detecting people without cars. Thus, reducing the cost of training and its computational resources is crucial in building a detection and tracking system for reading.

3.1. Principal components pursuit

In General, the essential task of the PCA algorithm is dimension reduction of data matrices to enhance the cost computation. Recently, it was used in various image processing applications to separate image matrix representation into two main types of Principal components (PC), which are low-rank (L) and sparse components (S), as shown in Fig. 2. The various proposed optimization algorithms, such as PCP, used to obtain the best solving for the background modeling problem demonstrate as follow Eq (1):

$$L, S \arg mn L_* + \lambda S_1 \; s.t. \; Y = L + S \tag{1}$$

where Y, L, and S $\in R^{(rxcxd)xnF}$, Y denote the captured video frames matrix, L is a low-rank matrix representing the background, S is a sparse matrix representing the foreground. $||.||_x$ is the matrix's nuclear norm, and $||.||_1$ is the ℓ_1 norm matrix. Argument Lagrange multiplier (ALM) - based solution involves Singular Value Decomposition (SVD) to determine the low-rank component. The Modified Fast PCP algorithm has been proposed [4] to solve the above problem. The final constrained problem formulated as:.

Fig. 2 The video components decomposition examples: (a) original frame, (b) background frame, and (c) foreground frame.

$$L_{k+1} = arg_L min||L + S_k - Y||_F \text{s.t.rank(L)} \approx \text{t} \tag{2}$$

$$S_{k+1} = arg_S min||L_{k+1} + S - Y||_F + \lambda||S||_1 \tag{3}$$

Fig. 2 represents the main results of this algorithm, where the background modeling task separates the pixels of the background and foreground components.

3.2. Deep-learning networks

The rapid development of the in-depth learning approach has expanded its use in almost all research areas. Good networks are designed and built with a more robust structure and performance [12], several networks have been developed for object detection and classification [9]. The most famous is the Region of Convolution Neural Network (R-CNN), as shown in Fig. 3, which marks the beginning of applying these networks to detect objects. After that, its performance was developed and improved for other tasks, such as discovering specific parts with certain features, such as identifying the face and pedestrians and classifying and distinguishing the object.

The multi-object tracking method is a good and valuable model for overcoming some challenges, such as frequent blockage, which is practically implemented using a Kalman filter and linking the data transferred to the correlation scale that measures the interference of the surrounding box of the object through the sequential frames that represent the video clip. This approach has achieved outstanding success and performance when a large data set is in place. It was used a lot not to discover people [19].

4. Proposed method

The proposed system's structure or framework consists of several stages or steps to reach the article's primary goal: detecting and tracking multiple objects online or in real-time applications. Most tracking methods for multiple objects may suffer from the appearance or increase in identification codes' substitution. The data relevance parameter is valid or accurate only when the status rating scale is enormous. Therefore, these routes suffer from the frequent interruption of tracking paths when blockages appear during the video clip. Therefore, the need arose to replace the data correlation scale with an adaptive scale that would be more accurate, combining it with the information resulting from movement and appearance simultaneously. Deep learning was applied to achieve this, where convolutional neural networks were trained on a data set to distinguish and track objects [18]. A large data set was used for training to increase the robustness of the proposed algorithm, provided that the simplicity of the system, its efficiency, and its applicability in live or real-time applications are preserved [26].

Some experimental tests, calculation of cost matrix, and other analyses were performed that evaluate the work performance of any proposed algorithm and compare it with existing algorithms and the possibility of being a good competitor for them. To this end, a set of different video data was used, some of which represent a live broadcast that reflects an actual reality for accurate monitoring, in addition to a real video clip captured on a platform.

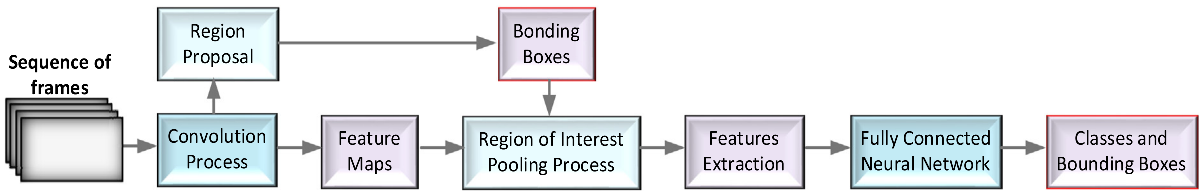

Fig. 3 The general block diagram deep learning based on object detection.

4.1. Object detection stage

In this article, a new object detection algorithm is presented, as shown in Fig. 4. Some of the challenges that hinder the creation of excellent and efficient detector modeling are presented. Some modern detection techniques based on deep learning were also covered. These detectors were developed and improved their performance by integrating them with the essential components' analysis algorithms as a pre-processing step, extracting some features, and sometimes removing noise. First, a single-stage detection technique is relatively fast and is often used in real-time applications [27]. The second type of detector is a two-stage detector [15], characterized by high accuracy at the expense of processing speed. The experimental results show the difference in accuracy between the two techniques. The director frames show the effectiveness of each in different scenes.

4.2. Object tracking stage

Fig. 5 represents the proposed tracking stage, which includes several steps. The object's location is defined by eight dimensions, four of which are related to distance: the coordinates of spin and yard, height and aspect ratio, and the velocity of those dimensions relative to the image's innovations. The Kalman filter (KF) used in the tracking is a linear velocity and linear tracer model. The association coordinates of the bounding box (BB) [28] take direct measurements of the object's condition. We count the number of frames associated with the prosperous pairing state for each path. This counter can be increased during the KF's prediction stage whenever the correlation or path is related to the measurements.

The trajectory life is calculated for each object within the scene, and the maximum trajectory age is determined. When discovering a new object within the sequence frames, a new

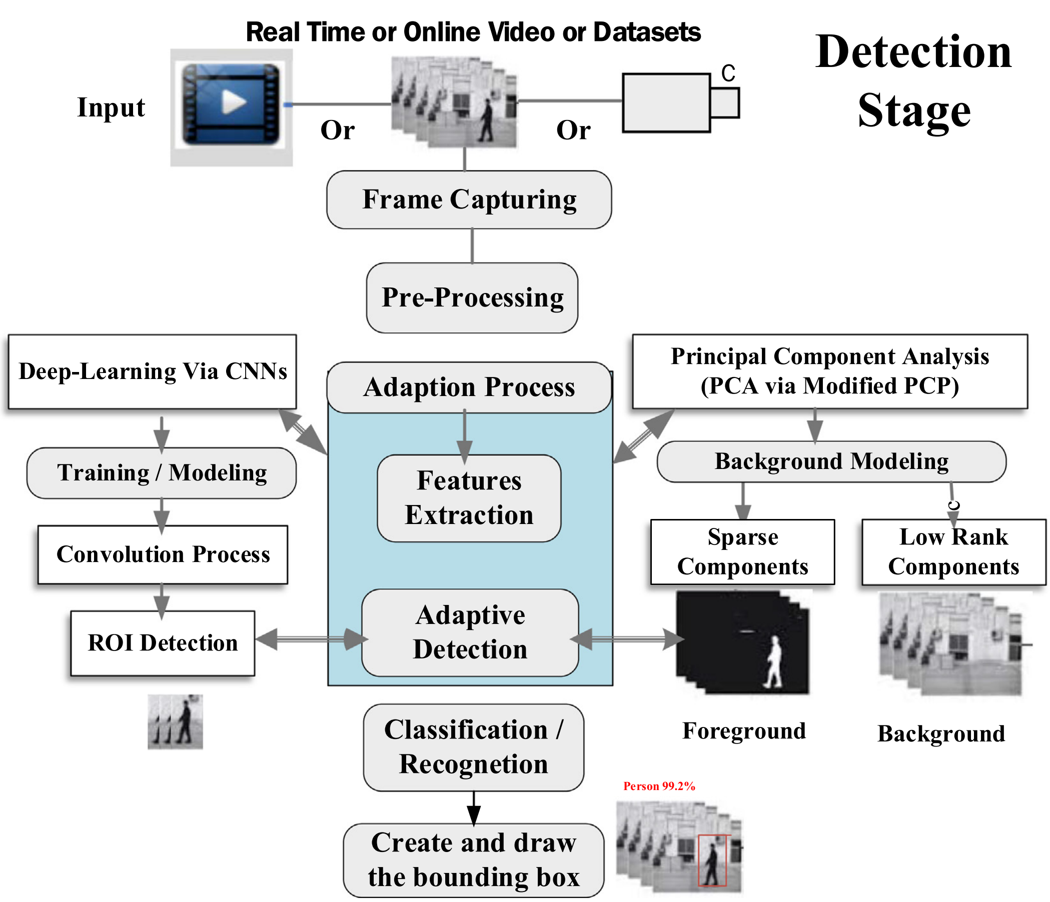

Fig. 4 The object detection and classification stage of the proposed algorithm.

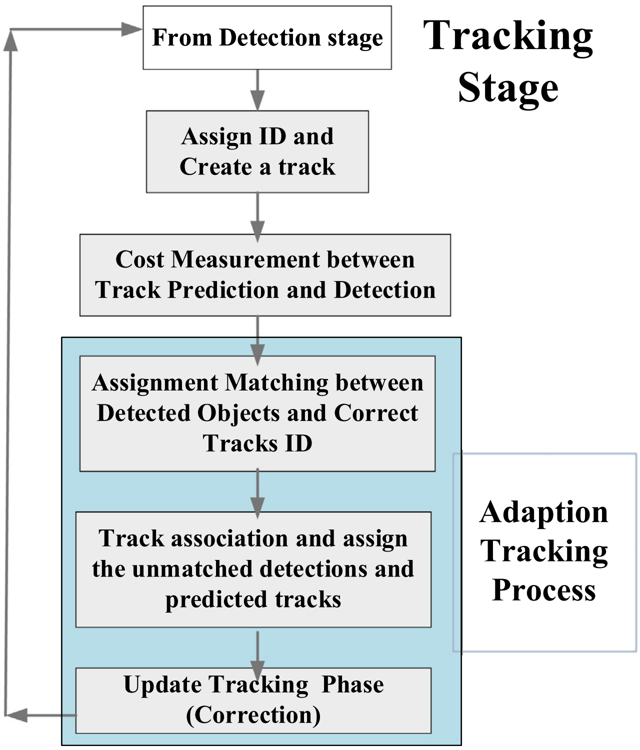

Fig. 5 The block diagram of the proposed tracking stage.

path is added to it after making sure that it cannot be linked to an existing trajectory. The new trajectories are considered temporary during the first frames from their appearance until a successful data association occurs at any time step. Each path exceeding the specified age limit assumes that the object has left the scene and is removed from the set of paths. To solve an assignment problem, we use an adaptive method to solve the correlation of data between predicted states of KF with new measurements. Both motion and appearance information is used using the scale of the Hungarian algorithm. Thus, the square distance is calculated between the values of the Kalman filter's predicted states and the recently arrived measurements' values.

$$dis^1(i,j) = \left(ds_j - y_i\right)^T S_i^{-1}(ds_j - y_i) \tag{4}$$

where i is index represents the tracing paths while (S and y) represent the projection of the path distribution within the measurement area. As for it, j indicates the discovery index, and ds represents the bounding box projection. Calculate the condition estimation uncertainty's fair distance using the number of standard deviations where the detection is out of the way. The impossible correlations are also eliminated by setting the confidence value of the squared distance with a high percentage, as shown below:

$$b^1_{i,j} = 1\left[d^1(i,j) \le t^1\right] \tag{5}$$

In real-time applications, the camera movement is unpredictable, which leads to rapid crawling operations within the video frame. Consequently, the square distance is a somewhat imprecise measure in light of the appearance of blockages.

An additional measure is added that depends on the appearance, which is the appearance descriptor (p_i with $\|p_j\| = 1$.) for each detection of a bounding box. The second metric is used to calculate the smallest distance between the I of tracks and j of detection in appearance area space, which describes as follow:.

$$dis^2(i,j) = min\left\{1 - p_j^T p_k^i | p_k^i \in \mathcal{R}_i\right\} \tag{6}$$

Then, we calculate the acceptance percentage between the tracks and detection as follow:

$$b^2_{i,j} = 1\left[d^2(i,j) \le t^2\right] \tag{7}$$

The Appearance descriptor scale provides valuable information for recovering identifying identities after blockages or slow movement appears. This scale was applied to a training data set, and multiple deep learning networks (the trained) were used to calculate the bounding box's adhesive appearance. The two scales complement each other, and the squared distance measurement provides more information on the objects' location. The two measures are combined using a weighted sum depend on the trade-off factor (λ), as shown below:

$$c_{i,j} = \lambda d^1(i,j) + (1-\lambda) d^2(i,j) \tag{8}$$

Then the acceptance ratio between detection and tracking is calculated using the following relationship:

$$b_{i,j} = b^1_{i,j} + b^2_{i,j} \tag{9}$$

The trained network architecture is composed of two convolutional layers as shown in Fig. 6, then a higher collector layer is followed by six remaining layers, followed by a dense layer to calculate the feature map. Finally, the batch layer obtains the normalization features with extreme accuracy to be compatible with the appearance scale.

5. Experimental tests, results, and discussions

The proposed system's performance was tested and evaluated on broad and multiple datasets that include multiple objects to measure its performance efficiency. From these data are the highway dataset as well as data based on the standard Change Datasets 2014(CD) [26] and MOT16 [29], among others. This data contains different scenes, such as clips captured from a moving camera, surveillance systems with different environments and events, etc. Then, PCP and pre-trained networks were relied upon to enhance object detection performance, as shown in Fig. 7. The first column represents the original captured frame. The second column represents the front part. It shows the results its efficiency for different Datasets challenges. The detection and tracking algorithms, mostly, and camera surveillance systems face different challenges, as in

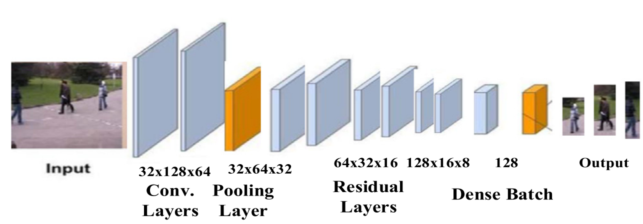

Fig. 6 The pre-trained CNN framework.

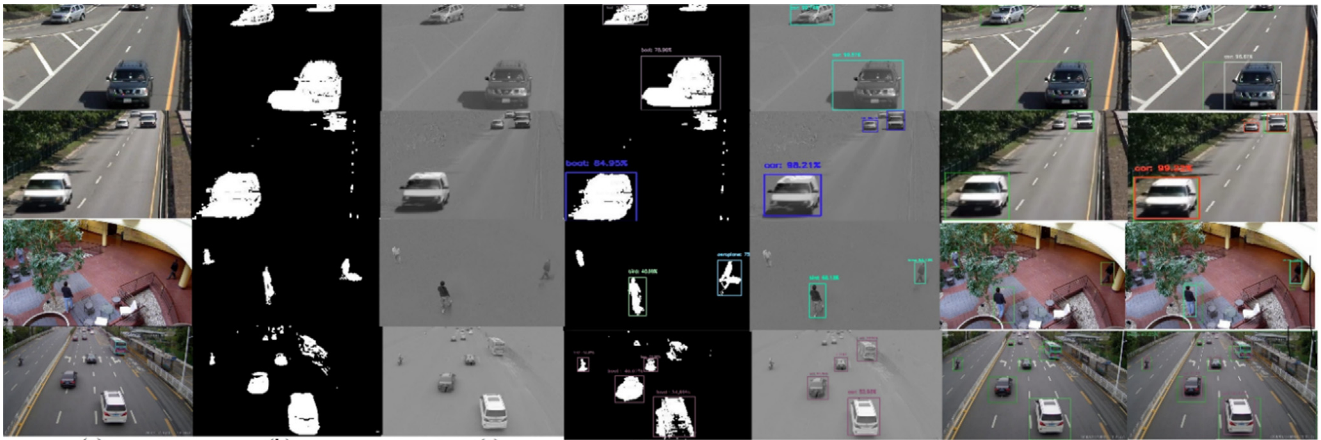

Fig. 7 The experimental results of the proposed algorithm: (a) original frame, (b) sparse frame, (c) outlier frame, (d), and (e) detected objects in sparse and outlier frame respectively, (f) detected object depend on modified PCP, and (g) detected objects frame by DL.

Fig. 8, for example, the difference in the size of the objects and their distance or proximity within the video scene, distortion or cutting of the captured image, camera movement, weather conditions such as fog or rain, etc. To demonstrate the performance efficiency of the proposed algorithm, we captured video clips using a linked camera. We mounted them on the NVidia Jetson TX2 Developer Kit [30,31] to obtain a different absolute control data set. The results showed good effectiveness and performance of the proposed algorithm in improving detection and tracking systems during real-time.

The results obtained from the proposed system were also compared with existing excellent and efficient performance algorithms, as shown in Fig. 9. There are several metrics for measuring object performance algorithms. Tracking a single object is somewhat simple while measuring algorithms' performance efficiency for multiple organisms is more complicated. It needs to apply a delicate design to create different sets of correspondence or paths for each object; we test the proposed algorithm's efficiency in different circumstances and challenges, it was tested on various datasets that include different environments, whether internal or external scenes, as shown in Tables 1 and 2. It shows that the proposed algorithm can be processed the different challenges from multiple video stream sources such as offline, online broadcasting, real-time monitoring source, and others. Then extract helpful information with high smoothness, speed, and accuracy where good results were obtained compared to previous existing algorithms, as shown in Fig. 9. Various methods and measures have been proposed in recent decades, and not a single specific method has been agreed upon. Several measures have emerged that have received researchers' approval and attention in performance measurement [1,32]. The proposed system was

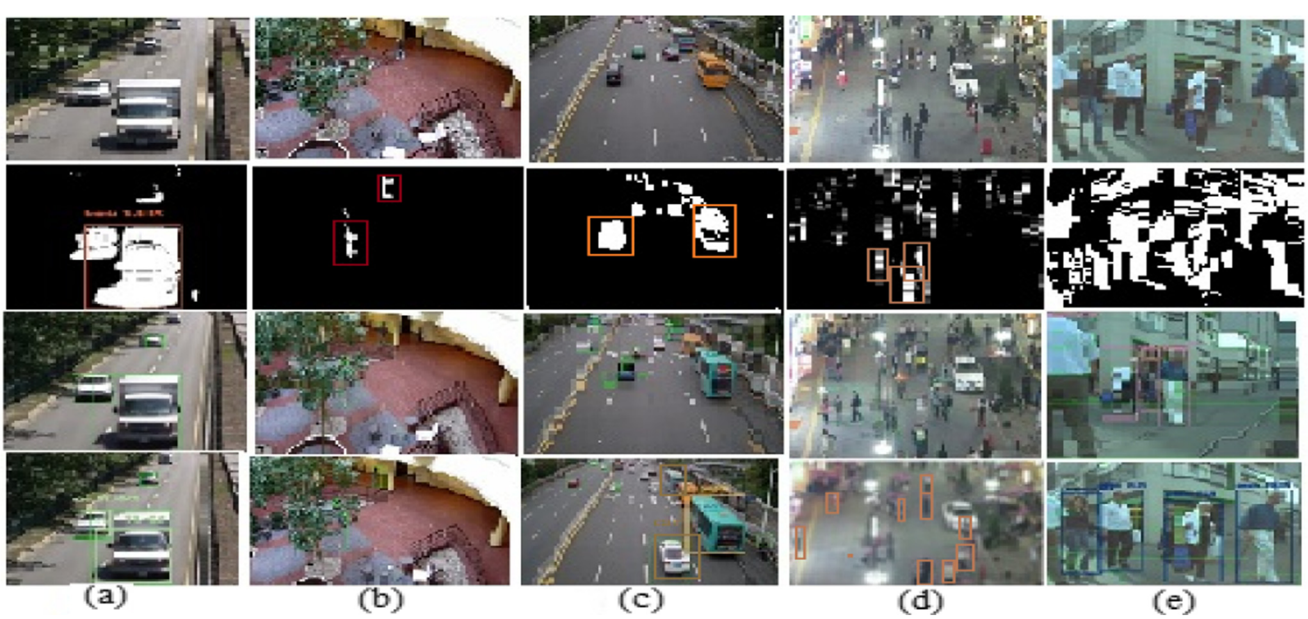

Fig. 8 The experimental results for different challenges: (a) multi-objects with different sizes, (b) hidden object, (c) the speed variation of objects, (d) the low confidence object detection, and (e) the occlusion.

Fig. 9 The experimental results of tracking stage comparison between the proposed algorithm and exciting algorithms: (a) ground truth, (b) KCF tracker [17], (c) IOU tracker [16], (d) DSORT tracker [19], (e) our tracker [22], and (f) the proposed algorithm.

Table 1 The evaluation performance matrices for the several different trackers by using MOT16 challenges dataset.

Method	Type	MOTA	MOTP	MT	ML	FM	Rcll	Prn	Runtime
KCF [17]	Batch	48.80	75.66	120	289	1116	52.52	94.22	10 fps
IOU [16]	Online	50.22	75.54	150	238	2302	59.12	90.56	8 fps
DeepSORT [19]	Online	61.44	79.07	249	138	2008	68.92	90.72	15 fps
Our Tracker1 [22]	Online/ Real	65.57	78.40	250	175	3112	73.51	93.75	13 fps
Our Proposed	Online/ Real	67.7	80.0	257	180	2580	75.9	94.87	12 fps

Table 2 The evaluation performance matrices for the several different trackers using CD2014 Challenges Dataset.

Method	Type	MOTA	MOTP	MT	ML	FM	Rcll	Prn	Runtime
KCF [17]	Batch	51.43	76.83	472	240	4233	62.66	84.67	12fps
IOU [16]	Online	44.00	76.13	312	284	5754	54.57	84.78	7fps
DeepSORT [19]	Online	54.8	79.78	383	321	4673	55.38	97.55	14 fps
Our Tracker1 [22]	Online/Real	60.95	78.00	312	189	7207	66.36	88.93	13 fps
Our Proposed	Online/Real	63.11	78.62	455	123	5874	72.82	91.56	12 fps

implemented in two platforms: first in Computer type MSI that contains an Nvidia GeForce RTX 2060 graphics card and second in an embedded artificial intelligence platform of the Nvidia Jetson TX2 type as shown in Fig. 10.

It shows the results and values for some of the proposed tracking algorithm's performance measures and its efficiency for different Datasets challenges, compared with some other tracking algorithms by calculating the different metrics [29] such as follow:.

- Multiple Objects Tracking Accurate (MOTA) and Multiple Objects Tracking Precision (MOTP) (standards: provide comparability to prove the efficiency of the proposed algorithm, including the identification of correspondence and movements, as well as false alarms. Where these events and findings are used in building data structures and analyzing them, as well as using these derived values to find other new measures.

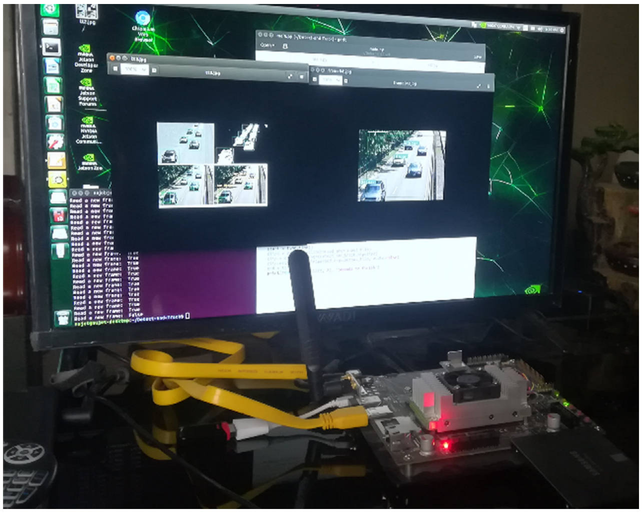

Fig. 10 The hardware implementation and experimental results using NVidia Jetson TX2.

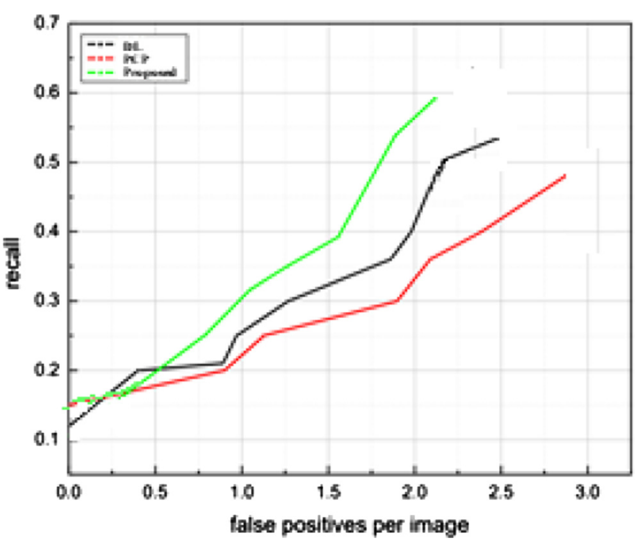

Fig. 11 The comparative accuracy analysis.

- Mostly Lost (ML): A percentage of ground truth paths can be traced to 20% of their assumed lives. Mostly Tracked (MT) is a percentage of ground truth tracks with the same address with a value of 80% as their minimum estimated life.
- Fragmentation (FM): The number of track interruptions that occurred as a result of a missing detection.
- The Precision (PRN) criterion was also calculated, which is the ratio of detected objects to the sum of positive and negative detected objects.
- The Recall (Roll) criterion represents the ratio of the number of positive detected objects to the original organisms' sum within the framework.
- It measured the clarity and accuracy of the trajectory, found the renewing paths for all objects, and guessed or predicted paths resulting from hypotheses. The measure of tracking the target to determine the tracer's efficiency and each hypothetical outcome means the actual object was present (True Positive TP), or a false alarm, i.e., a false positive. At the same time, the target object may be surrounded by multiple outputs.

Some measures of identification of matching were calculated, False Positive (FP) and False Negative (FN), which represent the number of false-positive and negative matches between the original pixels classifications with the results, respectively. The comparative result of the three approaches (DL, PCP, and proposed algorithm) are showed in Fig. 11. It draws the recall against the number of false-positive and shows the excellent performance of the proposed algorithm.

Table 3 shows the performance evaluation of the proposed method with the back-end modeling method using the quest and core components and the deep learning approach based on the (SSD MobileNet) network. The evaluation was done using a different data set (Gray and Color Scale). This comparison aims to evaluate the performance of the proposed method in terms of accuracy, processing time, and the advantage of using the hybrid framework over the use of the two approaches separately.

6. Conclusion

We introduced a new algorithm to detect and track objects by integrating the algorithm seeking essential components with deep learning networks adaptively at the detection stage. Both approaches may succeed in facing some challenges that the other may fail to overcome. The performance of the Kilman filter has

Table 3 Performance evaluation and comparison of the proposed algorithm.

Method	Frame type	Accuracy	Average F-measure	Average Processing time per frame (s)
Principal Component Analysis (PCP)	Gray Scale	0.70	0.65	1.4
	Color Scale	0.64	0.70	2.3
Deep Learning (DL)	Gray Scale	0.80	0.76	1.2
	Color Scale	0.83	0.80	2.6
PCP-DL hybrid framework	Gray Scale	0.89	0.84	1.7
	Color Scale	0.92	0.88	2.9

also been improved in the tracking stage by adding visual tracking to reduce fragmentation and false negative detections and improve the accuracy of the tracking paths of objects during the video sequence. Also, the challenges include the emergence of noise, change in speed, the scene of the object, blockage, and other challenges that appear during natural application scenes. The new algorithm has been tested, many simulations and training, and practical implementation on the Nvidia platform. The results showed the efficiency and reasonable accuracy of the proposed dialogues compared to the previous methods. Future proposals may include using new methods of object detection and tracking by testing more automatic identification of features and characteristics of the proposed algorithms and comparing their performance with the proposed algorithm. Expand the range of object detection and tracking in crowded scenes or extreme contrast in lighting and online scenes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B. Karasulu, S. Korukoglu, Performance Evaluation Software: Moving Object Detection and Tracking in Videos, Springer Science & Business Media, 2013.

[2] H.N. Abdullah, N.H. Abdulghafoor, Automatic objects detection, and tracking using FPCP, blob analysis, and Kalman filter, Eng. Technol. J. 38 (2) (2020) 246–254.

[3] C. Kim, J. Lee, T. Han, Y.-M. Kim, A hybrid framework combining background subtraction and deep neural networks for rapid person detection, J. Big Data. 5 (2018) 22.

[4] P. Rodriguez, B. Wohlberg, Incremental principal component pursuit for video background modeling, J. Math. Imag. Vis. 55 (1) (2016) 1–18.

[5] H.N. Abdullah, N.H. Abdulghafoor, Objects detection and tracking using fast principle component purist and Kalman filter, Int. J. Electr. Comput. Eng. 10 (2) (2020) 1317, https://doi.org/10.11591/ijece:v10i210.11591/ijece:v10i2.pp1317-1326.

[6] N.H. Abdulghafoor, H.N. Abdullah, Real-time object detection with simultaneous denoising using low-rank and total variation models, in: 2020 Int. Congr. Human-Computer Interact. Optim. Robot. Appl., IEEE, 2020. <https://doi.org/10.1109/HORA49412.2020.9152863>.

[7] R. Iqbal, F. Doctor, B. More, S. Mahmud, U. Yousuf, Big data analytics: computational intelligence techniques and application areas, Technol. Forecast. Soc. Change. 153 (2020), https://doi.org/10.1016/j.techfore.2018.03.024.

[8] R. Iqbal, F. Doctor, B. More, S. Mahmud, U. Yousuf, Big data analytics and computational intelligence for cyber-physical systems: recent trends and state of the art applications, Futur. Gener. Comput. Syst. 105 (2020) 766–778, https://doi.org/10.1016/j.future.2017.10.021.

[9] Z.-Q. Zhao, P. Zheng, S.-T. Xu, X. Wu, Object detection with deep learning: a review, IEEE Trans. Neural Networks Learn. Syst. 30 (11) (2019) 3212–3232.

[10] A. Grzywaczewski, R. Iqbal, Task-specific information retrieval systems for software engineers, J. Comput. Syst. Sci. 78 (4) (2012) 1204–1218, https://doi.org/10.1016/j.jcss.2011.10.009.

[11] R. Iqbal, T. Maniak, F. Doctor, C. Karyotis, Fault detection and isolation in industrial processes using deep learning approaches, IEEE Trans. Ind. Informatics. 15 (5) (2019) 3077–3084, https://doi.org/10.1109/TII.2019.2902274.

[12] D.J. Hemanth, V.V. Estrela (Eds.), Deep Learning for Image Processing Applications, IOS Press, 2017.

[13] R.S. Sidhu, M. Sharad, Smart surveillance system for detecting interpersonal crime, in: 2016 Int. Conf. Commun. Signal Process., IEEE, 2016, pp. 2003–2007.

[14] R. Arroyo, J.J. Yebes, L.M. Bergasa, I.G. Daza, J. Almazán, Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls, Expert Syst. Appl. 42 (21) (2015) 7991–8005, https://doi.org/10.1016/j.eswa.2015.06.016.

[15] S. Ren, K. He, R. Girshick, J. Sun, Faster r-CNN: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2017) 1137–1149.

[16] E. Bochinski, V. Eiselein, T. Sikora, High-Speed tracking-by-detection without using image information, in: 2017 14th IEEE Int. Conf. Adv. Video Signal-Based Surveill, IEEE, 2017. <https://doi.org/10.1109/AVSS.2017.8078516>.

[17] N. Tijtgat, W. Van Ranst, B. Volckaert, T. Goedeme, F. De Turck, Embedded Real-Time Object Detection for a UAV Warning System, in: 2017 IEEE Int. Conf. Comput. Vis. Work., IEEE, 2017. <https://doi.org/10.1109/ICCVW.2017.247>.

[18] G. Chandan, A. Jain, H. Jain, Mohana, Real-time object detection and tracking using deep learning and OpenCV, in: 2018 Int. Conf. Inven. Res. Comput. Appl., IEEE, 2018. <https://doi.org/10.1109/ICIRCA.2018.8597266>.

[19] N. Wojke, A. Bewley, D. Paulus, Simple online and real-time tracking with a deep association metric: 2017 IEEE Int. Conf. Image Process, IEEE, 2017, pp. 3645–3649.

[20] T. Mahalingam, M. Subramoniam, A robust single and multiple moving object detection, tracking and classification, ACI 17 (1) (2021) 2–18.

[21] B. Blanco-Filgueira, D. Garcia-Lesta, M. Fernandez-Sanjurjo, V.M. Brea, P. Lopez, Deep learning-based multiple object visual tracking on embedded system for iot and mobile edge computing applications, IEEE Int. Things J. 6 (3) (2019) 5423–5431.

[22] N.H. Abdulghafoor, H.N. Abdullah, Enhancement performance of multiple objects detection and tracking for real-time and online applications, Int. J. Intell. Eng. Syst. 13 (2020) 533–545.

[23] N.H. Abdulghafoor, H.N. Abdullah, Real-time moving objects detection and tracking using deep-stream technology, J. Eng Sci. Technol. 16 (2021).

[24] R. Iqbal, T. Maniak, C. Karyotis, Intelligent remote monitoring of parking spaces using licensed and unlicensed wireless technologies, IEEE Netw. 33 (4) (2019) 23–29, https://doi.org/10.1109/MNET.2019.1800459.

[25] T. Maniak, R. Iqbal, F. Doctor, Hierarchical Spatial-Temporal State Machine for (2020) 1–10.

[26] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, P. Ishwar, CDnet 2014: An expanded change detection benchmark dataset, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Work, 2014, pp. 387–394.

[27] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single Shot Multibox Detector, Springer, Eur. Conf. Comput. Vis., 2016, pp. 21–37.

[28] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, S. Savarese, Generalized intersection over union: a metric and a loss for bounding box regression, Proc. IEEE Conf. Comput. Vis Pattern Recogn. (2019) 658–666.

[29] A. Milan, L. Leal-Taixé, I. Reid, S. Roth, K. Schindler, MOT16: A benchmark for multi-object tracking, ArXiv Prepr. ArXiv1603.00831, 2016.

[30] L.P. Bordignon, A. von Wangenheim, Benchmarking Deep Learning Models on Jetson TX2, (n.d.).

[31] S. Mittal, A Survey on optimized implementation of deep learning models on the NVIDIA Jetson platform, J. Syst. Archit. 97 (2019) 428–442, https://doi.org/10.1016/j.sysarc.2019.01.011.

[32] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, K. Murphy, Speed/accuracy trade-offs for modern convolutional object detectors, in: 2017 IEEE Conf. Comput. Vis. Pattern Recognit, IEEE, 2017. <https://doi.org/10.1109/CVPR.2017.351>.