Traffic Violations Detection Review based on Intelligent Surveillance Systems

Hasan Thabit Rashid¹, Prof. Dr. Israa Hadi Ali²

¹²IT College & University of Babylon
¹Education College for Women & University of Kufa

Received Date: 20 May 2021
Revised Date: 23 June 2021
Accepted Date: 08 July 2021

Abstract — Currently, IT develops our life superficially and quickly has become faster and more complicated. However, this paper offers a brief study of previous techniques for violation of vehicles on surveillance systems expressed by suitable processing methodologies to intelligent surveillance techniques (such as Wi-Fi sensors, image processing, machine learning, and object detection based on appearance and motion) and the use of different types of cameras networks (fixed, motorize, and PTZ) and area topologies (efficient FOVs). This study provides a quick look at various techniques that alert individuals and users to vehicle anomalous movements in the environments of traffic and intelligent systems.

Keywords — computer vision, intelligent surveillance systems, multi-camera networks, traffic control, video tracking.

I. INTRODUCTION

An intelligent traffic surveillance system has the target of deciding and assessing the infringement vehicles from the recorded video of any setting traffic map scenario. However, the sheer volume of recordings and the need for time, mean that much of the material is never observed or checked on. As a result, incidents and episodes are ignored, and suspicious behaviors are not taken note of in time to avoid wrongdoing from happening. The utilizing of intelligent video surveillance (IVS) systems is distributing quickly in a varied applications. Intelligent video or analytics is the method of analyzing video information with the objective of changing it into actionable information. Analytics-based frameworks utilize complex numerical calculations to analyze the video and extricate moving objects or other recognizable shapes whereas sifting out unimportant developments [1]. According to [2], multi-camera collaborative networks (MCCN) established methods have numerous diverse assumptions setting up such as: several fixed cameras with intersection field of views (FOV); several fixed cameras with non-intersection FOVs; and several Pan-Tilt-Zoom (PTZ) cameras, where cameras’ FOVs can be crossing or non-crossing at diverse periods, and however, both circumstances have to be taken into account. It is considering a big challenge when following the objects over moving cameras. In multi-camera setting up, the real-time framework will fuse tracking information from different cameras, and hence will be growing the advantages of IVS by allowing the client to characterize instructions for outline the specific full zone environment, and independent of single cameras to be used. The combination depends on all cameras being standardized to a site outline, whereas the single cameras generally stay unchanged.

According to [3], using user-defined instructions, IVS frameworks can naturally distinguish possible dangers or gather commercial cleverness statistics by identifying, following, and exploring objects in a scene. In general, as to [4], the design of observation methods is done based on gathering their fundamental highlights onto three fundamental models to be referred to as global taxonomy. The first is the behavioral model, which helps in processing, describing the manners of a surveillance environment problem as a set of tasks such as finding, locating, identification, following, and content treatment. The behavioral model can be utilized to determine the tasks which are essential for a specific situation. The second is the implementation model, which buildings the vital choices to implement the surveillance tasks defining by the first model. It could be a set of tools such as cameras type, unit connectivity, and unit fixture. The final is the actuation model, which is in charge of taking some protective procedures if the surveillance system notices some irregular conditions. The actuation model then may be considered as an important tool to reduce effort, cost, and time as in the following Figure1.
Video surveillance technology has been used to control traffic, detect an event such as accidents on highways, and monitor any crowded behavior in public spaces. Simple video surveillance system (VSS) components are camera and lighting, managing and recording unit, and output interface with display. Compatibility and quality of all these components should be guaranteed by the effective interaction of VSS components [2][5]. Video content analytics (VCA) is the ability to automatically evaluate video to notice abnormal actions, some articles for such applications are: ego-motion, Motion detection, shape recognition, object detection, recognition, and video object tracking [6]. Crowding on the traffics' roads are coming to be a very great problem in our time. Smart transportation systems aim to provide better techniques for managing traffic follows. It facilitates numerous users to be more informed, and safer to provide an informative and well-organized traffic management facility [3]. However, the subsequent data are kept in such a database that can be investigated with sets of rules that applied, for example: an object passing a virtual line in the video, or More than 6 vehicles waiting in a drive-through route. A lot of applications identify vehicle license plates, while others focus on defending important infrastructure through virtual lines that can activate alarms and alerts in the event of an intrusion, vehicle counting, recognition of traffic incidents, vehicle overtaking, vehicle occlusion treatment at traffic connections. However, rules can be programmed to help in control if the events detected in the video are normal or if they should be flagged as alerts to security staff or the police department [1, 5].

II. RELATED WORKS

A. Sensors Techniques

In [7], a system for auto-defining the traffic violation was offered to announce the vehicles that violate one or more of the traffic's rules and direct the vehicle pictures to the traffic'
managing center. Vehicle details are ready-made, and the bill is created. However, the system does not work depending on any statistical method, and the road lanes are not explored in any way. Capturing the vehicle pictures and violation details are performed at the same time; it is transmitted to the traffic managing center to be analyzed and performed, and the bill is generated based on vehicular named data networking (VNDN) summarized as in the following Figures 2.

![Pending tickets entries (PTE)](image1)
![Tickets received (TR)](image2)

**Fig2: Structure form of smart cop data [7]**

In [8], a system for editing road traffic violated tickets were offered based on the vehicle's engine speeds, and therefore the violation with the violator was recognized. The vehicle owner is imitated and informed the violated traffic rule, and at the same time, the ticket is created. This work uses the radio frequency identification (RFID) tools to deal with such violations. Speed control automatically governs the velocity of a vehicle's engine. It has the ability to keep the drivers' speed, and however, the work's objective is to govern the velocity along with the driver's intimation every time the vehicle overran the particular speed. Active RFID reference tags are reserved at ordered durations. However, the results of applying the system are showed as in Figure 3.

![Fig3: Violations representation graph based on the system [8]](image3)

**B. Video Tracking**

In [9], a statistical method is proposed to simplify discovering of several traffic violations on highway roads. The method, interprets lanes on the road as equations of lines using some image processing and video tracking procedures. A tracking procedure creates a path equation for each vehicle. A linear algebra is used for the method modeling and computing the violations. After that, many computed equations are resolved to fix and recognize the crossing paths with the road lanes as to following Figure 4. It is designed to detecting specific situations such as "over-speeding, carpooling, tailgating, dangerous lane change, single lane change, double lane change, and illegal lane change detection". The violations classification on recorded video is achieved according to the above wrong traffic rule. A single label is assigned to the related violated vehicle, and its path and location coordinates are kept if a specified event has been occurred. However, if the violated vehicle is overrun the traffic rule once more, then it is re-identified by the system; the time changes of such violations are computed and identified by discovering the crossing among the vehicle paths and the lane equations.

![Fig 4: Detect different driving trajectories in the highway [9]](image4)

**C. Image Processing and Machine Learning**

In [10], a suggested vehicle flow detection method to distinguish traffic anomalies. Data collection, feature extraction, and learning phases are used. The abnormalities include stopping and driving in a back way. Video recording is done using Closed Circuit TV traditional cameras focusing on the vehicle's front and rear to help in getting
robust results to the differences in the running environment situations. For tracking purposes, numerous successive frames are picked up for moving detection and its properties, such as edges and license platelets. The path of the traffic flow is achieved using a trained classifier model (K-nearest neighbor), which is adapted for unsupervised learning. However, the method is calculated and estimated on a highway road; the system is set up in Figure 5 below.

![Fig 5: The system set up](image)

Database is created with optical flow statistics cooperative based on the required metadata, which is the date, time, and license plate to be used for feature selecting and extracting. Open CV library is used with Lucas-Kanade pyramidal method where the positions with high information such as edges of the license plate and forward-facing grid are choosing for following. However, visual flow data is collected to generate an individual motion vector showing the path of flow as to the following example is given in Figure 6.

![Fig 6: Subsequent movements flow (arrow) and characteristic location (dot)](image)

The path of traffic over the highway is achieved throughout the learning process; the specific region is initiated to omit unimportant places on the scene. Unsupervised learning is chosen and done finally by a system designed based on learning data itself only and does not need any guiding such as tags allocated. Along the traffic road, path and the interesting regions are detected to avoid unimportant scene locations to enhance the performance. However, repeated places over the road surface by the vehicles are shown as following figure where the red placeless is considered only. See Figures 7 and 8.

![Fig 7: Successive vehicles' directions extracted from the road surface. Left: forward camera while right: reversed camera](image)

![Fig 8: The vehicles' motion vectors. Left: forwarding motions (training set size: 3360), right: reversing motions (training set size: 3510)](image)

**D. Object Detection**

In [11], wide-ranging of the vehicle-identifying methodologies and their implementations in intelligent transporting systems under variable situations are reviewed. Because of the instability of driving situations, there are several difficulties and challenges in determining and detecting the vehicles. However, it can be classified as "appearance-based methods and motion-based methods" as in the following Figure 9. Furthermore, different illumination changes, environmental conditions, and traffic situations are discussed. For more details, see the next sections. However, the authors are recommended that, for the future, to focus on strong vehicle detection methods for several situations.
E. Approaches based on Appearance

The vehicle's appearance, such as (cars, trunks, and etc.) differs in size, figure, and color. For detection purposes, segment the foreground and the background methods are employing prior knowledge and using various features such as the color, the symmetry, the horizontal or vertical edges, and the shadow according to the nature of the rectangular shape of targets. It is achieved based on object generation and object verification.

a) Object Generation

where the locations of the vehicles are considered and hypothesized in three-level processing.

i. Low-level Features

It is simple, effective, and very helpful in extracting vehicles' information. The color system, video images contain rich information provided by colors; the RGB model often operates in traffic surveillance cameras and also is used to identify optical characteristics of vehicles such as colored light, and the platelet numbers. However, (RGB) layers are more associated with each other, and the singular value of each pixel be determined by its corresponding brightness intensity as in [12]. Projected Shadow, it has happened when the illumination is changed, and however, it can be identified if the vehicle moves across the scene. Shadow is still an important challenge and has the responsibility for many problems such as distortions, and tracking loss. Many proposed algorithms to extract and remove the shadow in different situations, as in [11]. Figure Symmetry, the front or rear view of vehicles usually has some symmetry. Summity is discovered by defining an arithmetical process for concluding the symmetry axes or the center pixels to detect the vehicle location in frames' images. However, symmetry feature is used in rare due to the topology of installing video cameras nowadays with diverse angles and views. Image Edge, it is a basis of contour information and is computing in few processing steps as comparing with other features; it is widely used for this purpose even in a real-time mode. Sobel, Canny, or Prewitt operators are popular edge-based vehicle detection methods that we're applying to create gradient edge
plots and are often gathered with extra characteristics; for example, a multi-scale edge mixture was operated to detect objects like vehicles.

ii. Local Feature Detectors

It is a robust and more reliable than the local feature descriptors are designed to allow searching objects' blobs in the consequent images of the frames even when many objects and updating background are considered at the same time. Also, local feature descriptors are invariant to geometric and photometric alteration. Histogram of Gradients (HOG, the object blob recognition by picking up the structure of edge gradient that is more representative of the local figure; however, the various modification was proposed such as the mixture of two built vectors of HOG to get the features of the target, the pyramid features method (PHOG), horizontal edge (H-HOG), vertical edge (V-HOG), concentric rectangular (CR-HOG) detectors and the symmetry of HOG vectors. Haar-like descriptor. Its work based on the edge, line, and the point-surround features which are processed in simple, many modifications approach for representing the feature and collecting vehicle’s edges. Extra Feature Detectors, other appearance features also used to collecting contextual information such as Gabor features (GF), Speed-Up Robust Features (SURF), and the Scale-Invariant Feature Transform (SIFT) features. Blob Detector at multiple scales (interest regions) using SIFT to discard spurious regions whose magnitude is smaller than a predefined threshold. However, a high level of cleverness and work is needed to recognize blobs after the process of knowing the pixels' changes. Basically, the neighboring pixels that contain similar characteristics are candidating to form such blobs, and each blob should have borders that define each part of frame images to distinguish whether it is a vehicle or any other object. This is done by recognizing blobs' parameters such as figures, sizes, velocity, and many other metadata. Analyzing the metadata of the motion or static objects is requiring the complex and intelligent algorithms and applications to be classified [1, 11].

b) Object Verification

Tests are performed to verify whether it is vehicles or non-vehicles. Machine learning techniques are used to treat as a vehicle versus non-vehicle classes. Initially, collect from various vehicles sources divided as positive images and negative images (non-vehicle); then trained with classifiers such as Support Vector Machine (SVM), Adaboost, and Neural Networks (NN), which are popular for that purpose where the images of each frame are represented by single or numerous features' vectors in all training samples to make a decision border between the two classes.

NN can learn highly nonlinear decision boundaries in a time-consuming way of tuning the parameters. However, Deep learning technologies such as Convolutional Neural Networks (CNN) such as fast R-CNN, region proposal network (RPN), and fast vehicle proposal network (FVPN) have been successfully used in vision-based vehicle detection under various conditions as hybrid methods based on features and classifiers for vehicles detecting using the (front, rear, and side) views such as Gabor feature with backpropagation NN classifier, HOG with Adaboost, and many other methods. More details in [11].

F. Motion-based Approaches

a) Foreground and Background Detection

Methods to detect moving objects in a frame broadly classified based on non-statistical or non-parametric and statistical or parametric methods. The initial frame is regarded as the background, and the next frames are subtracted from the background in the non-statistical methods. Then the objects are regarded as the pixels having a value greater than a threshold. The background is updated or maintained in frames. However, they are suitable for real-time application since it is faster than statistical methods.

The PDF functions of the background pixels are estimated for the statistical methods. Then the probability is calculated in each video frame that a pixel is in the background. In contrast with the non-statistical approaches in the background modeling for outdoor scenes, statistical approaches perform better but more storage and computational time is required. One popular technique is background modeling based on Gaussian Mixture Model (GMM), which represented pixel statistics on the scene. The multimodal background modeling is highly helpful in removing repeated movements such as bright, branch-leaves, and a flagship flag [5].

b) Visual Flow

It is a typical method for vehicles detection over a high duration. Many proposed algorithms for moving vehicle detection, such as pyramidal optical flow estimating and extracting front vehicle features by refined image process, and then, the optical flow was applied. In addition to the above, optical flow estimation, the distances factors, and an accurate feature template were used to track on-road vehicles in low-resolution video sequences.

c) Extra Methods

Motion of dynamic textures over the image background is a prominent property, and hence, motion histogram was used in to split vehicles from the dynamic background. Another method was built according to the scene characteristics, and different parameters are used.

In [13], a vehicle tracking depending on features as (position, shape, light, scale) is proposed for online traffic by examining the vehicles' features based on computing their consistent matched structures, prominent features using an unsupervised manner. The tracking is performed by calculating the active pixels in the successive video frames. The method is compared with others that are based on SIFT and Kalman filter that gives enhanced outcomes by the
tracking under diverse circumstances such as lighting changes, scaling, and rotating in low time complexity and improved precision. The objective is to detect possible vehicle sections and then perform a segmentation process using a matching algorithm for zooming the vehicle to get its features, especially the particular vehicle and classify them amongst varying classes. However, segmented vehicles and their features pixels are matched and represented using white lines into two images in one frame while in case of changing the moving vehicle location in the next frame as in following Figure 10; matching features process will help to pinpoint the tracking of the vehicle even some metadata are changed such as location, figure, brightness, magnitude, and occlusion.

In [14], basic analytics applications (such as pixel-based motion detection) make decisions by determining the difference in the pixels' features. If a certain feature such as the color or figure is changed by changing their pixels properties, they can raise an alert or trigger other actions. Intelligent video applications (that is based on object tracking) process the images in a specific way, first they segment the object's blob in the sensor FOV, and then they track that object's blob over video frames or as it moves from one FOV to another. Tracking Objects using Hybrid localization (HY) Tracker (Particle-Filter + Mean-Shift (MS)), which is popular trackers that use the color histogram for detecting the particles to increase the accuracy in choosing the sampling as a good manner for dealing with maneuvering objects and minimizing the space dimensionality. MS alone cannot track fast-moving targets, while HY successfully tracks the object at different speeds.

III. ENVIRONMENT TOPOLOGY AND CROSSLINE DETECTION APPLICATIONS

In [15], smart cities have the problem of roadside unit (RSU) location proofing which is studied in order to secure vehicles even in modern urban-based on the ability of events' discovering among cars in traffic flow surveillance systems. The main purpose is to investigate the lowest amount of RSUs and to ensure signal broadcasting. Its conditions that for each traffic flow, the RSUs there should be available with an optimum number. However, three proposed algorithms are used to achieve that purpose by placing the edge of traffic flows with RSU based on optimal locations. Also, to raise the efficiency of such algorithms, real data-driven is used for tests. Vehicle flows, state-space graph, weather, and time will be used to model vehicle traffic flows as for future work. See Figure 11.

Moreover, according to [15, 16, and 17], many other different studies on urban or highway environments that are influenced by changing global or local ambient illuminations even in daylight, and varying weather circumstances (snow-storm, desert-storm, wet-snow, rain, cloudy, foggy, misty, etc.) as to the following figure. This will affect video acquisitions, different casing types of distortions such as noisy or saturation or excessive brightness change; histogram equalization and enhancement techniques are used widely to detect and handle the acquired images with such distortions as in [12].

Crossline, also known as virtual tripwire, perimeter guard, fence guard, and digital fence, are used to alert any rules' breaking or possible lines crossing. Scene planning for such a specific environment by considering a cross line or dangerous area to be alerted the system to start an actuation procedure that alerting the people in charge if objects were passing that line, entering, exiting a certain area. It is possible to set up multiple detection zones with different rules for allowing movement, zone crossing, sending alarms, and recording video. For example, anyone is allowed to move within the warning. Analytics for
business intelligence and operations includes object classification, object counting, people counting, and traffic management. Most analytics software generally goes through the following steps [1].

1. Detection step: analyzes all the video' pixels, compares each frame' pixels to a specific frame (background frame), and figures out any moving objects.
2. Segmentation step: extracts and allocates the moving objects with labeling marks based on their metadata such as color, size, figure, velocity, date, and time.
3. Classification step: sorts the segmented objects into different targets kinds, such as a person or car. Then, they are allocated a set of descriptors that describes them, for example, color, size, or direction. See Figure 12.

Data fusion procedure has the main target of gathering the metadata from all signal view cameras into a fused form that contains all related metadata streams on the universal view plan of such environment. The fusion sensor identifies all corresponding metadata that incoming at the same time of overall corresponding single view cameras. However, third-party time is used to fine-tuning the synchronization process where it's done based on metadata' time stamps while arriving data may altering and to substitute any delay, buffering for the input metadata is used by the fusion sensor with such dormancy and maintenance.

IV. MULTI-CAMERA COLLABORATIVE NETWORK AND SURVEILLANCE SYSTEMS

Many works and efforts have been applied to the MCCN concept divided into three aspects as following [2]. Several fixed cameras with intersection FOVs, a common concept that was showing devices based on several views of fixed cameras. Several fixed cameras with non-intersection FOVs, the cue of intersection FOVs limits the active area watched by cameras. Object tracking, such as vehicle tracking over many cameras with non-overlapping FOVs in such highway environments is allowed the cooperative FOVs of the surveillance system to be spreading over a wide range of kilometers. Several PTZ cameras, it represent a quite difficult to use a camera's FOVs where they can be intersecting or non-intersecting at varying times, continuously the backgrounds will be changing, and the FOVs. However, both states need to be addressed. The advantage of PTZ conception is surveillance a wide range of active watched area although its need a more computation complexity and going to be useful in a wide range of real-time systems applications.

For summarizing the main objectives to design any environment problem, one needs the following points: first, select an EFT for cameras’ FOV to obtain good surveillance results. Second: the use of the concept of MCCN is to obtain the best surveillance techniques, and as a result, avoiding the problems of Clutter (appearance similarity between object and background), occlusions (dynamic, scene, and appearance), and the illumination changes (sudden changes in light conditions and movement) [1, 2]. However, most challenges for designing surveillance systems are summarized as:

a. Tracking multiple targets via the cross-camera system by associating them in different individual cameras (even they defer in their parameters such as fixed and PTZ).

b. Considering strong algorithms that demonstrate all faced problems in the processing pipeline, such as clutter, occlusions, illumination changes, analyzing, detection, tracking, classification, and event detection.

c. Practically, need to design a unified format for fused video metadata from multi-type cameras to be used simply for deployment and maintenance.

d. Suitable technique for using both fixed and PTZ cameras in the term of the master-slave system.

e. One big challenge is to operate the system in real-time mode.

According to [1], the PTZ camera can operate with the static camera based on the Master-Slave system for zooming squared zone or specific area only to get the (front and rear) platelets images of each vehicle or to follow a particular object. It is an intelligence that automatically controls a PTZ camera to keep an object in the view. Zooming in the camera’s FOV can be controlled automatically to provide a good show of the vehicle. When one PTZ camera covers the area, then the difficulty is that it might be directed in the wrong way at the time of the specific action. However, this work needs more elaborate thought; hybrid approaches can preparing that issues.

Issue 1: the fixed cameras would detect the actions, while the PTZ cameras would track the objects.

Issue 2: If automatically PTZ camera only tracks one object at a time, and suddenly more than one object has appeared in the camera view, then the surveillance system will get a problem in recognition. However, two mechanisms
are there:

**A) Auto select and track:** in this case, the system looks to select and track the initial motion object till it misses that object. The system will then look at another object. This scenario is beneficial in low actions or activities situations such as carpark places and lobbies. It is operating well without the need for on-site security staff.

**B) Manual select and track:** in this case, the security man selects an object for tracking, and the system’s camera follows it. This scenario is guarantying to focus on the specific object correctly.

As a tradeoff for using (PTZ or fixed) cameras, PTZ cameras with a large visual zoom factor can give more information about the frame’s images and investigate a large zone by the use of PTZ’ conception abilities in totally. However, a full auto trip needs to be fixed to control the actions surveillance. For trading with fixed cameras, surveillance the videotapes without online watching; fixed cameras are usually much better in cost. There are several types of network cameras, such as the (fixed, dome fixed, PTZ, panoramic, covert, and thermal) cameras. PTZ cameras have many properties based on the purpose of use (image stabilization, present and guard tours, E-flip, Auto-flip, PTZ performance, and joystick control). However, some cameras type's examples and their properties as in following figure 22, and for more details, refer to [1].

According to [2], testing the whole surveillance system withers its mechanism is intersection, non-intersection and its fusion module is considered to be challenging task and however, video frames testing is difficult due to the variability of incoming videos. Good test results for such a system should have online and offline recording videos. However, the benefits of the longstanding examine the online outdoor environment will issue unpredicted situations. Offline videos help in testing a vast range of circumstances forcefully, making repairing difficulties simpler. Finally, many levels of processes are required to test the system, such as analytics on each single camera server, fusion metadata and incident recognition, control of camera follower, the whole system, and administrate the graphical user interface.

**V. CONCLUSIONS**

We present a quick description of diverse technologies for traffic control and vehicle violations following the traditional control approaches to intelligent surveillance systems. The use of techniques based on classifiers such as Haar-cascade or deep learning is better than background modeling techniques. Multi-camera network is better than using a single camera for surveillance systems. The study of area topology and detecting the efficient FOVs is more reliable and confident in design the system. This review introduces a fast introduction for the new academics and researchers who like to begin with such a field of study. However, the explanations of the review will put them in the right place as associating with the evolution of our world today.

**REFERENCES**


