International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

SVM-GA Based A Novel Technique for the Detection of the Vehicle in an Optimized Overlapped Multi-Camera System

[1]Joshi Rakhi Madhukaro, [2]D. S. Rao

Submitted: 30/08/2023 **Revised**: 25/10/2023 **Accepted**: 01/11/2023

Abstract: The presence of CCTV cameras on highways helps enhance public safety and instills a sense of security among road users. It promotes responsible behavior, discourages reckless driving, and can aid in resolving disputes related to accidents or incidents. This research proposes an innovative and very trustworthy approach to identify vehicles on the road. This paper also examines vehicle tracking in the presence of uncalibrated Cameras with overlapping fields of view used to capture images, which are characteristics of traffic surveillance systems. To discover a greater accuracy of the desired field of vision in such a situation, it is crucial to create and construct a system to compute the road geometry and establish correspondence between the positions of the multiple cameras. We outline a special technique for automatically identifying traffic camera angle positions using a collection of dimensions and the GA-SVM machine learning algorithm followed by the novel vehicle detection architecture.

Keywords*: CCTV, FOV, GA-genetic algorithm, Multi-Camera setup, machine learning, SVM, vehicle detection, Video Surveillance, neural network.*

1. Introduction

Monitoring of visual traffic employing a multiplicity of sensors with intersecting fields of view is critical for the development of smart cities, interstates, and highways. To provide safer transport services, together with smarter transportation infrastructure both academics and industry are paying attention to new advances in computing and communication technologies.

Traffic surveillance makes use of multi-camera, multi-object tracking. Even the most straightforward of these arrangements needs more than one camera for two reasons: First, because of a single camera's constrained area of vision, it is unable to completely encompass the environment (FOV). Second, many cameras should be placed in strategic locations to give resilience against occlusion-like scenarios [1, 2, 3, 4].

Establishing connections between several viewpoints is important in order to benefit from more cameras. The main challenge of tracking with a single camera is the connection between frames over time [5]. The angle of the overlapping field of view is a quantitative metric for measuring the correlation between tracks of objects observed from various cameras.

Visual sensors are used in visual traffic monitoring with video analytics to offer information, such as traffic flow prediction or vehicle detection. Although it has recently gained popularity among computer vision experts, it is still a difficult process, especially when several cameras are involved.

Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Hyderabad-500075, Telangana, India
bhardwaj.rakhi78@ klh.edu.in ,dsrao@klh.edu.in

Multi-camera installations require a more complicated infrastructure, the capacity to handle more data at once, as well as greater processing capabilities than mono-camera traffic monitoring.

The examination of visual signals collected from several cameras while accounting for setups with overlapping fields of vision (FOVs), in which cameras may be placed at various angles and separated by considerable distances. A rise in the number of cameras poses technological and architectural obstacles, such as issues with camera positioning and calibration, object matching across numerous cameras, automated switching between cameras, and information integration.

In this paper, we propose a machine learning methodology to more accurately capture roads and highways by automatically detecting the position angle of several cameras with overlapping FOVs along with the accurate detection of the vehicles.

The paper is structured as shown below. The cutting-edge of multi-camera technology in traffic surveillance design methodologies is covered in Section 2. The methodology for identifying the optimal observation zone automatically is introduced by the suggested method in section 3. Section 4 describes the evaluation strategy and implementation details. Last but not least, Section 5 describes the conclusion remarks.

2. Literature Survey

One fundamental concept from [8], which is used in this study, is identifying the optimal observation zone using low-level characteristics from the video. This method built on models can give us a positive outcome by offering a collection of models for each car. In developing countries, however, roads are inadequately split into lanes for various types of mobility (such as autos, bicycles, motorbikes, pedestrians, and others). It seems difficult to make models. Due to the wide variety of vehicles, the issue is even more complicated [6, 7, and 8].

In order to manage and regulate traffic congestion, bottleneck conditions, and prevent mishaps, CCTV-closed-circuit TV cameras are being used more and more in traffic network administration [9, 10, 11, and 12]. They can be used with fixed or steady FOVs, broader FOVs, PTZ capabilities, or multi-camera monitoring systems.

The study of numerous camera tracking characteristics is a huge task in computer vision, with publications categorized by features, matching approach, and camera calibration.

A. Approaches for Feature Matching

In multi-camera object tracking, the object is matched simultaneously in multiple cameras' view scopes, with various camera characteristics visible in each camera. Geometric constraint and feature identification are the two possible techniques.

Establishing correspondence constraints by matching the color or other color or other attributes of monitored objects in each camera may be the easiest way to assure equal labeling. This has been accomplished with either a Bayesian network approach or the computational Kalman Filter approach. At both times, the authors use a variety of features within the same structure rather than being limited to a single kind of feature.

To understand the camera geometry and infer additional limits, they also use camera calibration data. Geometry-based modalities include epipolar geometry, homograph, and landmarks, while recognition-based modalities include perceived height and color. These are divided into two categories: geometry-based and recognition-based.

Researchers have attempted to make color feature correspondence more trustworthy, however it is unreliable when the disparity is high. Because separate cameras have varying intrinsic characteristics and photometric features, having security cameras positioned in opposite directions is considered an optimal approach. Deviations in illumination also cause a similar object to appear differently in different cameras.

B. Approaches based on 3D information and camera calibration

Camera calibration and the 3D surroundings model structure can be used to offer systematic labeling by establishing the precise position of each 3D object in the worldwide coordinate system and determining equivalence across objects. This technique is used in [13, 14], as every single camera has calibration and the globe is delimited by a known ground plane. View equivalence is produced by connecting views with similar projected 3D locations. [15, 16, 17] also employ ground plane homograph.

This strategy may be effective in controlled situations, but calibrated cameras and accurate environment maps are difficult to obtain in a general surveillance scenario. We contend that camera calibration is redundant because most of the pertinent data may be obtained by monitoring motion over time. Camera calibration assesses intrinsic and/or extrinsic parameters relating to the camera's internal features, such as focal length, skew, distortion, and image centers. Multi-camera calibration and camera arrangement with interspersed and not interspersed FOV are distinguished by FOV.

It is vital in a multi-camera approach to calculate real-time precise locations for all cameras, which will better monitor numerous objects simultaneously, Object reconstitution in 3D within a scene, and a combination of innovative viewpoints [18].

C. Approaches for Alignment

Alignment-based techniques concentrate on recreating the geometric change across the cameras and guarantee that the tracks of an identical object overlap in alignment with the anticipated transition. Recently, the authors of [19] approached the issue from the perspective of frame alignment, extending spatial picture alignment approaches to integrate time information [20], describing a different strategy that employs trajectory information for alignment. On the other hand, [21], [22], and [23] describe ways to establish time correspondences between non-overlapping FOVs. Cameras set at regular intervals along a corridor or on a motorway are typical uses. Correspondence is established using the MAP method.

Multi-camera tracking systems must deal with the object handover problem, which involves the diverse occurrences of the same object in multiple cameras. In recent technology, fusion frameworks and methods are used to identify whether there exist overlapping regions among cameras. After surveying related research papers and methodology, multiple camera systems in the overlapping FOV with deep learning techniques emerged with better solutions. In a multi-camera tracking system, many methodologies and algorithms have been utilized in camera calibration, object matching, and information fusion, but there is a scope for future study.

D. Vehicle Detection

Currently, ITSs, in addition to a few industrial and military systems, make extensive use of object detection and classification. For example, ITSs can carry out vehicle identification and classification for complete analysis of passing vehicles in order to accomplish efficient vehicle traffic management and oversight as well as urban planning

Due to their powerful representational capabilities, convolution neural networks (CNNs) form the foundation of many object identification models currently in use [27].

The CNN feature extraction's performance of visual recognition is comparable to how humans perceive things [28]. Convolution, pooling, fully The methods for detecting objects that are now in use can be split into two categories: hardware-based approaches and vision-based methods.

The hardware-based strategy involves integrating hardware components like temperature sensors, lidar, and other devices [26]. While technologies based on vision are becoming popular and linked, and other sorts of layers are common in CNNs, and each layer converts the 3D input volume into a 3D output volume of neuron activations [29]. Different CNN architectures have been invented up to this time. The first of them, the Region-based CNN (R-CNN), successfully applied DL for object identification and other computer vision applications, as well as automatic picture feature extraction. The success of RCNNs, whose cost has been greatly decreased by sharing convolutions among object proposals, has really been a driving force behind recent developments in object detection [30].

After it, Fast R-CNN [31], Faster R-CNN [30], Mask R-CNN [32], and Mesh R-CNN[33] were developed as iterations of the R-CNN model. These are all illustrations of two-stage object identification models that produce sparse candidate frames (i.e., region suggestions retrieved from a scene) in the first stage and then validate, categorize, and enhance candidate frames in the second stage to improve the scores and locations [28]. Two benefits of these models are their high precision and localization of objects, while their main disadvantages are a more involved training process and slower operational speed [28], especially when you consider how crucial real-time object detection is becoming to practical applications. You Only Look Once (YOLO) [34] and Single Shot Multi-Box Detector (SSD) are examples of single-stage object detection models that perform better in this regard by adopting a regression method for object detection immediately, leading to faster operating speed. However, because SSD does not take into account the interaction between different sizes, it is limited in its ability to recognize small objects, whereas YOLO is faster and easier to learn general traits from [35]. However, the graphic area cannot be handled completely by SSD or YOLO, leading to significant detection error and missing rate rates.

From the literature, it is clear that there is a need to work on the camera angle along with highly reliable vehicle detection. In the next chapter, the proposed method is presented followed by the result and discussion chapter.

3. Proposed Method

To determine the overlapping region of two cameras with given angles, we have considered the field of view (FOV) of each camera and the distance between them. The camera is positioned as shown in figure 1.

Now we have calculated the FOV of each camera. The FOV is the angle at which the camera can capture an image. It is usually given in degrees.

We can determine the overlapping region by finding the area where the two FOV areas intersect. This is the region where both cameras can capture an image. Note that the size and shape of the overlapping region will depend on the specific angles and distances involved. It is important to use accurate measurements and calculations to get an accurate result.

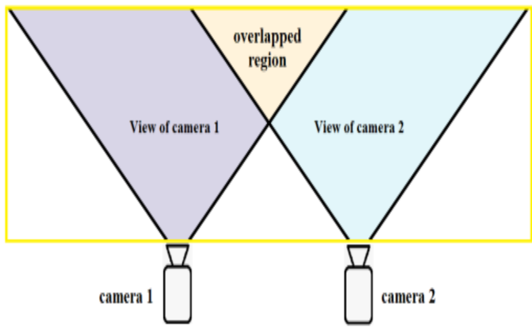

Fig 1: Viewing angle of two cameras with overlapping region

The size and shape of the overlapping region depend on several factors, including the camera's field of vision and the distance across them. A larger angle of view results in a larger overlapping region, while a shorter distance between cameras results in a smaller overlapping region. The shape of the overlapping region can vary from a simple rectangular or square shape to a more complex shape depending on the specific angles and distances involved.

The SVM classifier using genetic algorithm (GA) is used to determine the camera angle with the most overlapping region. The architecture of GA-SVM is illustrated in Figure 2.

The Support Vector Machine (SVM) is a well-known machine-learning technique for prediction [24]. The Genetic technique (GA) is an optimization technique inspired by natural selection. Combining SVM with GA can enhance the performance of the SVM algorithm and help find the optimal hyper parameters for the SVM model. The basic idea behind using GA with SVM is to search for the optimal set of hyper parameters that can maximize the SVM's performance on the given data. The main parameters of the SVM model, such as the kernel function, regularization parameter, and gamma, have a substantial impact on the model's performance

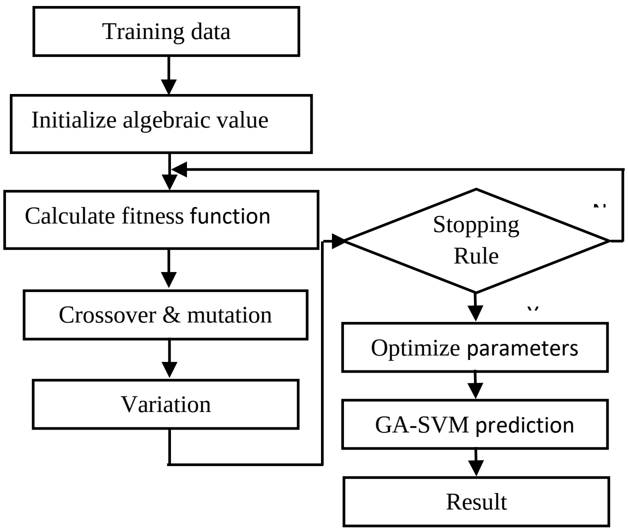

Fig 2: SVM with Genetic algorithm

In the GA-SVM approach, the SVM model's hyper parameters are considered as the genes in the GA. The GA generates a population of solutions; each represented by a set of hyper parameters, and evaluates their fitness by training the SVM model on a subset of the data. The accuracy of the SVM model on the remaining data is used to determine the fitness of each solution. The fitness function is calculated by using the following mathematics. We want to get the maximum overlapping of two camera outputs while maximizing the width of the margin. The fitness function which is also the objective function of the SVM becomes:

$$\max_{w,b} \frac{1}{2} \| w \|^2 + C\frac{1}{n}\sum_{i} \xi_i$$

$$\text{subject to } \begin{cases} y_i(x \cdot w + b) \geq (1 - \xi_i) & \text{for } i = 1, \dots, n \\ \xi_i \geq 0 & \text{for } i = 1, \dots, n \end{cases}$$

… Equation (1)

Where w is the decision boundary vector and c represents the cost of the slack. Boundaries produced by larger C will have fewer support vectors. SVM decreases its variance by increasing the number of support vectors because it is less dependent on any one observation. The model becomes more general when the variance is decreased. Therefore, lowering C will result in more support vectors and less over-fitting.

With Lagrange multipliers:

$\alpha_i \geq 0$ and $\mu_i \geq 0$ … Equation (2)

The primary Lagrangian function can be used to rephrase the constrained optimization problem as follows:

$$\min_{w,b,\xi} \max_{\alpha,\mu} \left[\frac{1}{2} \| w \|^2 + C\frac{1}{n}\sum_{i} \xi_i - \sum_{i} \alpha_i [y_i(x_i \cdot w + b) - (1 - \xi_i)] - \sum_{i} \mu_i \xi_i \right]$$

… Equation (3)

We can maximize over the multipliers subject to the relations derived before for w, b rather than minimizing over w, b, subject to restrictions. The dual Lagrangian formulation is what we refer to as:

$$\max_{\alpha} \left[\sum_{i} \alpha_i - \frac{1}{2}\sum_{i,i'} \alpha_i \alpha_{i'} y_i y_{i'} x_i \cdot x_{i'} \right]$$

$$\text{subject to } \begin{cases} 0 = \sum_{i} \alpha_i y_i \\ 0 \leq \alpha_i \leq C \text{ for } i = 1, \dots, n \end{cases}$$

… Equation (4)

With the use of Sequential Minimization Optimization, this quadratic programming issue is now rather simple.

The GA then employs selection, crossover, and mutation procedures to develop new solutions from the population's fittest individuals. The method is done multiple times until the ideal collection of hyper parameters is discovered.

The benefits of employing GA with SVM include the ability to avoid over fitting and increase model generalization. It can mitigate the time and effort needed to manually tune the hyper parameters of the SVM model. It is crucial to note, however, that the success of the GA-SVM technique is significantly dependent on the fitness function's quality, the size of the population, and the number of generations used. The result of the prediction of the angle for the maximum overlapping is presented in the next section.

The union of the footage from the two cameras (overlapped) region is then fed as an input. The image (a,b) is passed through the feature extraction process to get P_{ψ}^{i} (k,c,d). there is equation 5.

$$P_{\varphi}(k_0, c, d) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1N-1} .\, \text{image}(a, b)\varphi_{j=0}(a, b)$$

…Equation (5)

$$P^{i}_{\psi}(k, a, b) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M=1} . \sum_{y=0}^{N-1} .\, \text{image}(a, b)\psi^{i}j_{k,c,d}(a, b)\psi i$$

…Equation (6)

To enhance the performance of the network, we added residual blocks of sizes 1x64 and 2x128 to the Darknet-53's initial layers. In order to encourage feature reuse and provide the network the ability to learn more robust and discriminative features, the output from this stage is transmitted to a Dense Architecture, where every layer is connected to every other layer in a dense block. To enhance the network's information flow, the outputs from each layer are concatenated and transferred to the following dense block. Batch normalization and ReLU activation functions, which also prevent over fitting, improve the training process.

The front detection layer, or FDL, is in charge of locating and confirming the existence of objects in an input image. This layer, which is frequently a convolution layer, scans the input image at various scales and locations, looking for regions that potentially contain the objects of interest.

The front detection layer (FDL) uses small rectangles with specified sizes and aspect ratios called anchor boxes or prior boxes to identify objects in the input image. The FDL determines scores for each anchor box that represent the likelihood that an object will be found there, and bounding box offsets show the distance between the anchor box and the bounding box of the actual object. The architecture is made for small item detection, such as finding a gun in a video that looks little and far away. Up-sampling layers are used to increase the feature map's resolution and improve detection accuracy. This enables accurate detection of small items like firearms.

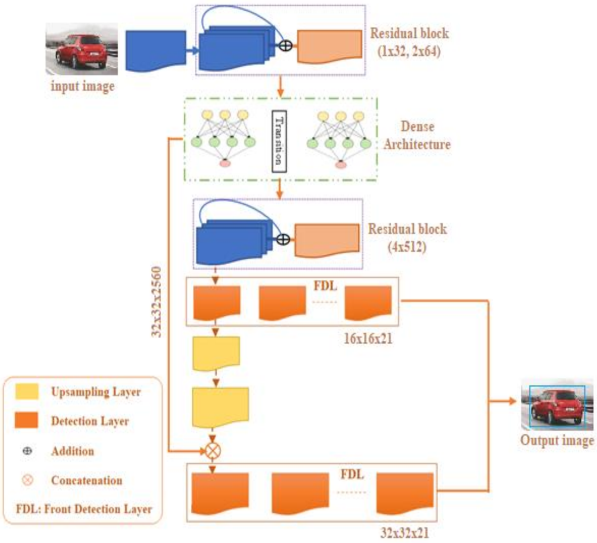

Fig 3: PSA for the detection of vehicle

In order to improve performance, we changed the size of the residual block at the top layer where our network operates in depth. A residual group contains different residual blocks, such as 1x, 2x, 4x, and 8x, after each convolution layer. Before each residual group, stride convolution with a stride of 2 is employed to down sample the spatial dimension of the feature maps. By encoding positional information necessary for object detection and preventing the loss of low-level features, stride convolution also made the down sampling less completely non-parametric than max-pooling. It made it easier to find smaller objects.

The network can learn more precise information from early feature maps and more significant semantic information from up sampled later-layer feature maps by up sampling and concatenating features with various scales.

We used 17760 photographs of different types and sizes of vehicles to train our PSA. The proposed algorithm is presented below and incorporates this architecture. The result of the PSA is presented in the next chapter.

4. Results and Discussion

The database is generated by calculating overlapping areas with every possible angle. This database is then given as input for the training GA-SVM. The result is presented in figure 4.

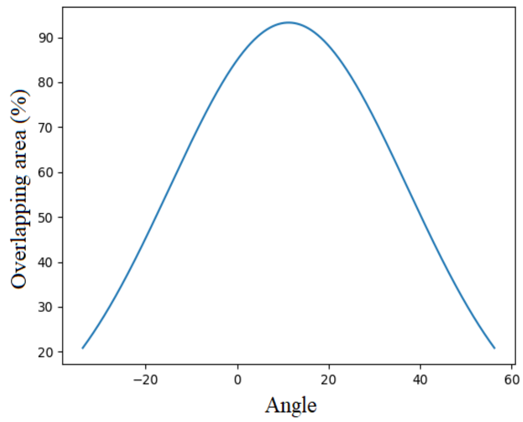

Fig 4: Result of GA-SVM for calculating best angle of camera for maximum overlapping

Here, the best angle of camera for maximum overlapping comes out as 14.36°. For this angle the overlapping is 93.27%. The result of the maximum overlapping is presented in figure 5.

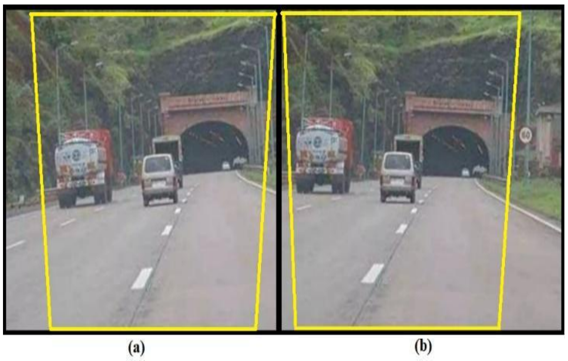

Fig 5: Result of overlapping part in both images for optimized angle calculated using proposed methodology

The overlapping of the two cameras is predicted by using other machine learning also. The overlapping along with the predicted angle is tabulated in table 1.

For object tracking purposes, overlapping footage from numerous cameras can be useful. The various viewpoints can be used by tracking algorithms to increase object localization, tracking accuracy, and resilience. Even when objects cross camera boundaries, object trajectories can be estimated more precisely by merging data from various camera viewpoints.

Table 1: optimized angle prediction along with percentage overlapping with various ML algorithms

Method	Overlapping percentage	Predicted Angle
Proposed SVM + GA	93.27%	14.36°
SVM + PSO	92.32%	14.08°
SVM	89.84%	15.89°
Decision Tree + GA	91.48%	15.94°
Decision Tree + PSO	91.04%	13.52°
Decision Tree	84.52%	17.94°

The proposed SVM+GA gives the maximum overlapping than another ML algorithm as illustrated in table 1. The result of the vehicle detection using PSA is presented in figure 6.

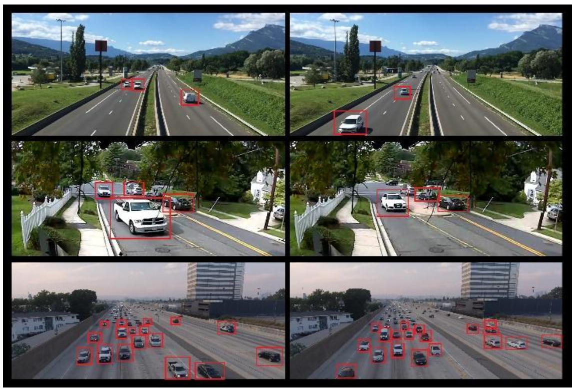

Fig 6: Result of vehicle detection using PSA

Confusion matrix is used to calculate the performance metrics. The performance metrics are calculated to check and compare PSA with other existing architectures after the feature extraction algorithm has been finalized. Equations 7, 8, 9, and 10 are used to calculate accuracy, precision, recall, and F1 score, respectively. The confusion matrix is first constructed to determine the performance metrics. Figure 7 shows a confusion matrix followed by formulae for performance parameters.

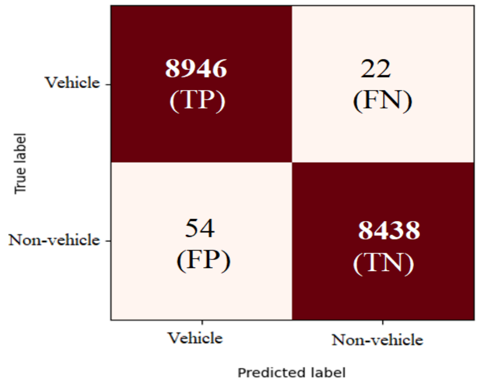

Fig 7: Confusion matrix of the PSA

$$Auracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{...equation (7)}$$

$$Precision = \frac{TP}{TP+FP} \quad \text{...equation (8)}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{...equation (9)}$$

$F1\ score = \frac{2TP}{2TP+FP+FN}$ …equation (10)

Figure 7's confusion matrix has been used to calculate performance metrics, which are reported in table 2.

Table 2: Performance parameters of vehicle detection using PSA

Performance Parameter	Value
Accuracy	99.56
Precision	99.5
Recall	99.5
F1 score	100

The performance parameters of the PSA is calculated and tabulated in table. The accuracy of the vehicle detection is compared with another ML algorithms also which is tabulated in table 3.

Table 3: Comparison of the PSA with other architectures

Architecture for Vehicle Detection	Accuracy (%)
PSA	99.56
SVM	78.32
CNN	87.92
RCNN	92.31
Yolo V3	95.82
Yolo V4	97.29

The table 3 illustrates that the PSA gives the maximum accuracy than that of the standard ML algorithms. The time complexity of the algorithm is calculated on different hardware platforms and tabulated in table 4.

Table 4: Time complexity of PSA, on different hardware platforms [25]

Different Hardware Platforms	Time taken to obtain a result (in seconds)
CPU, i3 processor, 8GB RAM	0.58
CPU, i5 processor, 8GB RAM	0.78
CPU, I7 processor, 8GB RAM	0.94
GPU, Nvidia K80	0.0038

After testing the PSA on different hardware platforms it comes to know that the higher-end CPU gives less time complexity. Whereas there is a drastic change in time complexity once this algorithm is tested on GPU.

5. Conclusion

In this research, firstly we presented a camera configuration that used the SVM classifier with the genetic algorithm (GA) to find the overlapping region of two cameras with given angles. We examined the field of view (FOV) of each camera and the distance between them; the size and form of the overlapping region will be determined by the precise angles and distances involved. To obtain an exact outcome, proper measurements and computations must be used. Combining SVM with GA can improve the efficiency of the SVM algorithm and aid in the discovery of appropriate hyper parameters for the SVM model. The benefits of employing GA with SVM include the ability to avoid over fitting and increase model generalization. It can also save time and effort when manually tuning the SVM model's hyper parameters.

Then the result of this overlapping is fed as an input to the proposed vehicle detection architecture. It shows the PSA gives 99.56% accuracy for the detection of the vehicles. The PSA is compared with other standard algorithms also and it comes out as the accuracy of the PSA is higher than that of the other algorithms. The time complexity is also calculated and it only shown that the detection of vehicle takes very less time in the real time scenario.

6. Declarations

- The authors have no relevant financial or non-financial interests to disclose.
- The authors have no conflicts of interest to declare that are relevant to the content of this article.
- All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.
- The authors have no financial or proprietary interests in any material discussed in this article.
- All authors contributed equally in this manuscript.

References

[1] L. Ciampi, C. Gennaro, F. Carrara, F. Falchi, C. Vairo, And G. Amato, “Multi-Camera Vehicle Counting Using

Edge-AI,” 2021, [Online]. Available: Http://Arxiv.Org/Abs/2106.02842.

[2] R. Cucchiara, “Multimedia Surveillance Systems,” VSSN 2005 - Proc. 3rd ACM Int. Work. Video Surveill. Sens. Networks, Co-Located With ACM Multimed. 2005, No. January 2005, Pp. 3–10, 2005, DOI: 10.1145/1099396.1099399.

[3] Devarajan, Z. Cheng, And R. J. Radke, “Calibrating Distributed Camera Networks,” Proc. IEEE, Vol. 96, No. 10, Pp. 1625–1639, 2008, DOI: 10.1109/JPROC.2008.928759.

[4] R. Dick And M. J. Brooks, "A Stochastic Approach To Tracking Objects Across Multiple Cameras," No. December 2004, 2015, DOI: 10.1007/978-3-540-30549-1.

[5] Y. Wang, K. Lu, And R. Zhai, ‘‘Challenge Of Multi-Camera Tracking,’’ In Proc. 7th Int. Congr. Image Signal Process., Dalian, China, Oct. 2014, Pp. 32–37, Doi: 10.1109/CISP.2014.7003745.

[6] A. Alshammari And D. B. Rawat, “Intelligent Multi-Camera Video Surveillance System For Smart City Applications,” 2019 IEEE 9th Annu. Comput. Commun. Work. Conf. CCWC 2019, Pp. 317–323, 2019, DOI: 10.1109/CCWC.2019.8666579.

[7] A. Koutsia Et Al., “Intelligent Traffic Monitoring And Surveillance With Multiple,” Pp. 125–132, 2008.

[8] P. Kumar, A. Mittal, And P. Kumar, "Study Of Robust And Intelligent Surveillance In Visible And Multimodal Framework," Vol. 31, No. 2007, Pp. 447–461, 2011.

[9] C. Micheloni, G. L. Foresti, And L. Snidaro, “Intelligent Distributed Surveillance Systems A Network Of Co-Operative Cameras For Visual Surveillance,” Image (Rochester, N.Y.), No. 20041147, Pp. 205–212, 2005, DOI: 10.1049/IP-Vis.

[10] Mohana And H. V. Ravish Aradhya, “Object Detection And Tracking Using Deep Learning And Artificial Intelligence For Video Surveillance Applications,” Int. J. Adv. Comput. Sci. Appl., Vol. 10, No. 12, Pp. 517–530, 2019.

[11] Y. Qian, L. Yu, W. Liu, And A. G. Hauptmann, “Electricity : An Efficient Multi-Camera Vehicle Tracking System For Intelligent City.”

[12] S. N. Raza, H. Raza Ur Rehman, S. G. Lee, And G. Sang Choi, "Artificial Intelligence-Based Camera Calibration," 2019 15th Int. Wirel. Commun. Mob. Comput. Conf. IWCMC 2019, No. May, Pp. 1564–1569, 2019, DOI: 10.1109/IWCMC.2019.8766666.

[13] A. R. Dick And M. J. Brooks, "A Stochastic Approach To Tracking Objects Across Multiple Cameras," No. December 2004, 2015, DOI: 10.1007/978-3-540-30549-1.

[14] A. S. Ladkat, S. S. Patankar And J. V. Kulkarni, "Modified Matched Filter Kernel For Classification Of Hard Exudate," 2016 International Conference On Inventive Computation Technologies (ICICT), Coimbatore, India, 2016, Pp. 1-6, Doi: 10.1109/INVENTIVE.2016.7830123.

[15] A. I. Abubakar Et Al., "Modified Neural Network Activation Function," 2014, DOI: 10.1109/ICAIET.2014.12.

[16] H. Hsu, T. Huang, G. Wang, J. Cai, Z. Lei, And J. Hwang, “Multi-Camera Tracking Of Vehicles Based On Deep Features Re-ID And Trajectory-Based Camera Link Models,” Pp. 416–424.

[17] Ajay S. Ladkat, Sunil L. Bangare, Vishal Jagota, Sumaya Sanober, Shehab Mohamed Beram, Kantilal Rane, Bhupesh Kumar Singh, "Deep Neural Network-Based Novel Mathematical Model For 3D Brain Tumor Segmentation", Computational Intelligence And Neuroscience, Vol. 2022, Article ID 4271711, 8 Pages, 2022. Https://Doi.Org/10.1155/2022/4271711

[18] A. Kholik And A. Harjoko, “Classification Of Traffic Vehicle Density Using Deep Learning,” Vol. 14, No. 1, 2020.

[19] A. Alshammari And D. B. Rawat, “Intelligent Multi-Camera Video Surveillance System For Smart City Applications,” 2019 IEEE 9th Annu. Comput. Commun. Work. Conf. CCWC 2019, Pp. 317–323, 2019, DOI: 10.1109/CCWC.2019.8666579.

[20] O. Elharrouss, N. Almaadeed, And S. Al-Maadeed, “A Review Of Video Surveillance Systems,” J. Vis. Commun. Image Represent., Vol. 77, No. April, P. 103116, 2021, DOI: 10.1016/J.Jvcir.2021.103116.

[21] T. D’Orazio And C. Guaragnella, “A Survey Of Automatic Event Detection In Multi-Camera Third Generation Surveillance Systems,” Int. J. Pattern Recognit. Artif. Intell., Vol. 29, No. 1, 2015, DOI: 10.1142/S0218001415550010.

[22] A. L. I. O. Ercan, A. E. L. Gamal, And L. J. Guibas, “Object Tracking In The Presence Of Occlusions Using Multiple Cameras : A Sensor Network Approach,” Vol. 9, No. 2, 2013.

[23] S. Khan And M. Shah, “Consistent Labeling Of Tracked Objects In Multiple Cameras With Overlapping Fields Of View,” Vol. 25, No. 10, Pp. 1355–1360, 2003.

[24] M. Shobana, V. R. Balasraswathi, R. Radhika, Ahmed Kareem Oleiwi, Sushovan Chaudhury, Ajay S. Ladkat, Mohd Naved, Abdul Wahab Rahmani, "Classification And Detection Of Mesothelioma Cancer Using Feature Selection-Enabled Machine Learning Technique", Biomed Research International, Vol. 2022, Article ID 9900668, 6 Pages, 2022. [Https://Doi.Org/10.1155/2022/9900668](https://doi.org/10.1155/2022/9900668)

[25] Ladkat, A. S., Date, A. A. And Inamdar, S. S. (2016). Development And Comparison Of Serial And Parallel Image Processing Algorithms. International Conference On Inventive Computation Technologies (ICICT), 2016, Pp. 1-4, Doi: 10.1109/INVENTIVE.2016.7824894.

[26] Adwait A. Borwankar, Ajay S. Ladkat, Manisha R. Mhetre. Thermal Transducers Analysis. National Conference On, Modeling, Optimization And Control, 4th – 6th March 2015, NCMOC – 2015.

[27] S. Woo, J. Park, J.-Y. Lee, And I.-S. Kweon, ‘‘CBAM: Convolutional Block Attention Module,’’ In Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, Pp. 3–19.

[28] M. Haris And A. Glowacz, ‘‘Road Object Detection: A Comparative Study Of Deep Learning-Based Algorithms,’’ Electronics, Vol. 10, No. 16, P. 1932, Aug. 2021.

[29] K. F. Hussain, M. Afifi, And G. Moussa, ‘‘A Comprehensive Study Of The Effect Of Spatial Resolution And Color Of Digital Images On Vehicle Classification,’’ IEEE Trans. Intell. Transp. Syst., Vol. 20, No. 3, Pp. 1181–1190, Mar. 2019.

[30] S. Ren, K. He, R. Girshick, And J. Sun, ‘‘Faster R-CNN: Towards Realtime Object Detection With Region Proposal Networks,’’ IEEE Trans. Pattern Anal. Mach. Intell., Vol. 39, No. 6, Pp. 1137–1149, Jun. 2017.

[31] R. Girshick, ‘‘Fast R-CNN,’’ In Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Santiago, Chile, Dec. 2015, Pp. 1440–1448.

[32] K. He, G. Gkioxari, P. Dollár, And R. Girshick, ‘‘Mask R-CNN,’’ In Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Venice, Italy, 2017, Pp. 2980–2988.

[33] G. Gkioxari, J. Malik, And J. Johnson, ‘‘Mesh R-CNN,’’ Presented At The IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Seoul, South Korea, 2019.

[34] J. Redmon, S. Divvala, R. Girshick, And A. Farhadi, ‘‘You Only Look Once: Unified, Real-Time Object Detection,’’ In Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, Pp. 779–788.

[35] W. Kong, J. Hong, M. Jia, J. Yao, W. Cong, H. Hu, And H. Zhang, ‘‘Yolov3-DPFIN: A Dual-Path Feature Fusion Neural Network For Robust Real-Time Sonar Target Detection,’’ IEEE Sensors J., Vol. 20, No. 7, Pp. 3745–3756, Apr. 2020.