IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, January 2024

Computer Vision Application: Vehicle Counting and Classification System from Real Time Videos

Akshay Narendra Thakare, Palash Kailas Kamble,
Kaushik Prabhakar Patil, Prof. Triveni Rahangdale
Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, India

Abstract*: Traffic analysis is a problem that city planners have been dealing with for years. Smarter methods are developed to analyze traffic and speed up the process. Traffic analysis can record the number of vehicles and vehicle classes in an area at a given time. People have been developing such mechanisms for decades now, but most of them involve using sensors to calculate the direction of moving vehicles and identify vehicles to track vehicle numbers. Although this system has matured over time and is very effective, they are not budget-friendly. The problem is that such systems require periodic maintenance and calibration. Therefore, this project aims to calculate and classify the vehicle based on vision. The system involves capturing frames from video to detect and count vehicles using Gaussian Mixture Model (GMM) background subtraction, then classifying vehicles by comparing contour areas with predicted values. A significant contribution of the paper is the comparison of two classification methods. Classification is done using Contour Comparison (CC) and Bag of Features (BoF) methods.*

Keywords: Vehicle counting, Traffic analysis, Contour Comparison.

I. INTRODUCTION

Today, countries and governments need a safe and affordable system to automate vehicles and control vehicle theft. Increasing traffic on roads and highways, increasing congestion, and problems with existing vehicle detectors have led to the development of new vehicle detection technologies. Computer vision systems are the most common choice, but several problems must be overcome to successfully perform classification. Real-time detection and tracking of objects or vehicles moving on different roads by intelligent vision systems is important for many fields of research and technology applications.

Extracting useful information such as traffic density, object speed, driver behavior and traffic patterns from these camera systems becomes essential. Manual analysis is no more. Developing intelligent systems that can extract traffic congestion and vehicle classification data from traffic control systems is essential for traffic management. Otherwise, the monitoring system is also important in driver assistance applications, as the vision system allows the detection and classification of the vehicles involved in the photographed incident.

A figure is a visual representation of something. The term has several uses in information technology. A photo is an image that is created or copied and stored electronically. An image can be interpreted as a vector graphic or a raster graphic. Digital image processing involves the manipulation of digital images using a digital computer. This is part of the signal and system, but pay special attention to the image. DIP focuses on developing computer systems that can perform image processing. The input of the system is a digital image and the system processes the image using an efficient algorithm with the image as output. It makes it possible to apply more extensive algorithms to the input image and can avoid problems such as noise and signal distortion during processing. images can be divided into the following three types.

A binary image consists of pixels that can be one of two colors, usually black and white. Binary images are called two levels or two levels. This means that each pixel is stored as an integer, ie 0 or 1. Gray is an intermediate color between white and white. It is a neutral color or achromatic color, which literally means "colorless" color because it can be composed of white and white. It was the color of cloudy skies, dust and lead. A color (digital) image is a digital image that contains color information for each pixel. This process is environmentally friendly as it does not require chemical processing. Digital imaging is often used to document and record historical, scientific, and personal events

This paper describes a vision-based system for detecting, tracking and classifying moving vehicles. Four different potential traffic groups can be defined, but the proposed software is flexible and the number of groups that can be classified.

II. LITERATURE SURVEY

Alpatov et al. the problem of road condition analysis for traffic control and safety is considered. The following image processing algorithms are proposed: vehicle detection and counting algorithms, road sign detection algorithms. Algorithms are designed to process images taken from stationary cameras. The developed vehicle detection and counting algorithms are also tested on embedded smart camera platforms.

Singing etc. proposed a vision-based vehicle identification system and counting system. This study published a high-level highway dataset with a total of 57,290 annotations on a total of 11,129 images. Compared to existing public databases, the proposed database contains small objects with annotations to provide a complete database for vehicle detection based on deep learning.

Neupane et al. created a training database of about 30,000 samples from seven existing vehicle camera classes. To solve P2, fine-tuning is based on this trained and applied transfer exercise in a modern YOLO (You Only Look Once) network. For P3, this work proposes a multi-vehicle tracking algorithm that quickly calculates each direction, classifies, and obtains the vehicle velocity.

Lin et al. It introduces a traffic monitoring system based on virtual detection zones, Gaussian mixture model (GMM) and YOLO to improve vehicle counting and classification efficiency. GMM and virtual detection bands are used to count vehicles and YOLO is used to classify vehicles. Additionally, vehicle distance and time are used to estimate vehicle speed. In this study, Montevideo Audio and Video Data (MAVD), GARM Road-Traffic Monitoring Dataset (GRAM-RTM) and our collection dataset are used to test the proposed method.

Chauhan et al. It uses an advanced Convolutional Neural Network (CNN) object detection model and trains various vehicle classes using data from Delhi roads. This work achieved 75% MAP in the 80-20 train test using 5562 video frames from four different locations. This work evaluates the latency, energy and hardware costs of deploying research based on our CNN model, as the growing region lacks strong network connectivity for continuous video streaming from the road to the cloud server.

Arinaldi et al. presented a traffic video analysis system based on computer vision techniques. The system is designed to automatically collect important statistical information for policy makers and regulators. These statistics include vehicle counts, vehicle type classifications, and vehicle speed estimates from video and vehicle usage monitoring. System detection like and vehicle classification in vehicle video. This work carried out two models for this purpose, the first is MoG + SVM network and the second is based on Faster RCNN, a new popular deep learning architecture for object detection in images.

Goma et al. presented an efficient real-time approach for detecting and counting moving vehicles based on YOLOv2 and showing point motion analysis. This work is based on the detection of synchronous vehicle features and tracking to achieve accurate counting results. The proposed strategy works in two phases; the first is to identify vehicles and the second is to count moving vehicles. For initial object detection, this work uses the fastest learning object detection algorithm YOLOv2 before filtering K-groups and KLT tracker. An efficient approach is then introduced using the temporal information of the inter-frame detection and tracking features to label and accurately calculate each vehicle with its respective trajectory.

Oltean et al. It proposes a real-time vehicle counting approach using a small YOLO for tracking. This program works on Ubuntu with GPU processing and the next step is to test it on low budget devices like the Jetson Nano. Test results show that this approach achieves high accuracy in real-time speed (33.5 FPS) in real traffic video.

Pico et al. proposed to implement a low-cost system for vehicle identification and classification using an ARM-based platform (ODROID XU-4) installed with the Ubuntu operating system. The algorithm used is based on an open source library (Intel OpenCV) and is implemented in the Python programming language. Experiments prove that the efficiency of the implemented algorithm is 95.35%, but it can be improved by increasing the training samples.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, January 2024

Tituana et al. review various previous works developed in this field and identify the technological methods and tools used in those works; In addition, this study also highlights trends in this field. The most relevant articles are reviewed and the results are summarized in tables and figures

III. PROPOSED SYSTEM

This system can be used to detect, recognize and track vehicles in video images, and then classify the detected vehicles into three different classes based on their size. The proposed system is based on three modules, namely background learning, foreground extraction, and vehicle classification as described in background subtraction, a classic approach to capture background images or in other words, moving object detection.

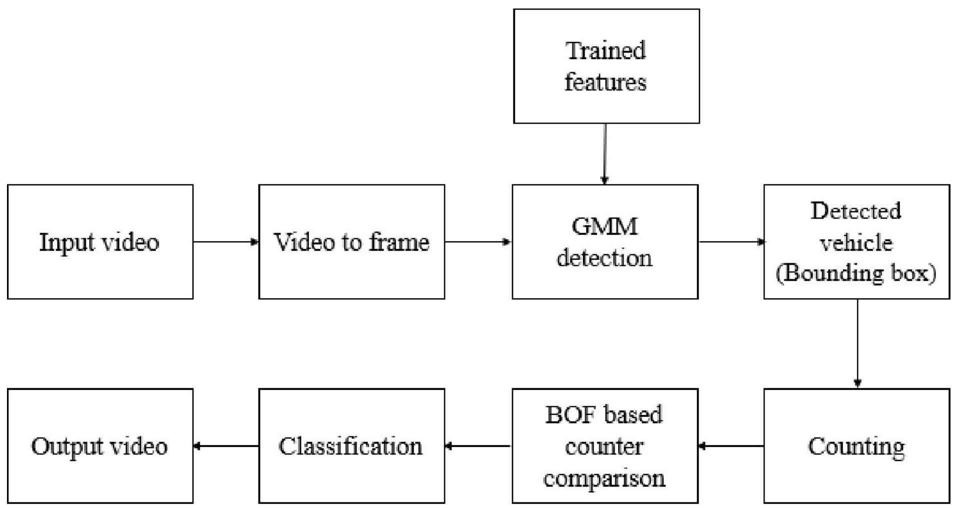


Fig.1: Block diagram of proposed system.

Gaussian Mixture Modelling (GMM)

In its simplest form, GMM is a type of clustering algorithm. As the name suggests, each group is modeled according to a different Gaussian distribution. This flexible and probabilistic approach to data modeling means that we have soft assignments instead of hard assignments to groups such as k-order. That is, each data point can be generated by one of the distributions and associated probabilities. In fact, each distribution has a specific "responsibility" for generating specific data points.

How can we evaluate the appearance of this model? Well, one thing we can do is introduce hidden variables Data (gamma) for each data point. This assumes that each data point is generated using some information about the latent variable. In other words, it tells you that the Gaussian generated a certain data point. But in practice we do not see these hidden variables, so we have to estimate them. How to do this? Well, luckily for us, we already have an algorithm that works in this kind of situation, the Expectation Maximization (EM) algorithm, and we'll talk about it next.

The EM algorithm: The EM algorithm consists of two steps, the E-step or Expectation step and the M-step or Maximization step. Suppose we have some latent variables (unobserved and defined by the vector Z below) and our data point X. Our goal is to maximize the marginal probability of X given our parameters (defined by the vector). In fact, we can find the marginal distribution as the union of X and Z and find the sum of all Z's (rule of probability).

$$ln\, p(X|\Theta) = ln\left\{ \Sigma_z p(X, Z|\Theta) \right\}$$

The above equations often produce complex functions that are difficult to scale. What we can do in this case is to use Jensens Inequality to construct a lower bound function that is easier to optimize. If we optimize this by minimizing the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, January 2024

KL difference (gap) between the two distributions, we can approximate the original function. This process is described. I have also provided a link to a video showing the derivation of the KL divergence for those who want a more rigorous mathematical explanation.

In fact, we only need to do two steps to evaluate our model. In the first step (E-step), we want to estimate the posterior distribution of our latent variable γ in terms of conditional weight (π), our term (μ) and the Gaussian mean covariance (Σ). Then we can enter the second step (M-step) and use it

Increase the likelihood associated with our parameter parameter θ. This process is repeated until the algorithm converges (the loss function remains unchanged).

Background Learning Module

The first module in this system is to learn how the background differs from the foreground. Also, since the proposed system works on the video feed, this module extracts frames from it and learns about the background. In a traffic scene captured by a static camera mounted on the roadside, moving objects can be considered foreground and static objects can be considered background. An image processing algorithm is used to learn about the background using the method described above.

Vehicle Detection and Counting

The third and final module in the proposed system is classification. After applying the pre-extraction module, the corresponding contours are obtained, these contour features are like centroids. Aspect ratio, area, size and stiffness are extracted and used for vehicle classification. This module consists of three steps, namely background reduction, image enhancement and foreground extraction. The background is removed to reveal the foreground. This is usually done by setting the static pixels of the static object to binary 0. After background subtraction, noise filtering, dilation, and erosion are used to obtain the correct contours of the background object. The output of this module is first class.

Area of interest: In the first frame of the video, I draw a line near the image and define the ROI. The goal is to recognize ROI in later frames, but recognize that ROI is not the primary vehicle. This is only part of the vehicle and can be deformed, rotate, translate, or even completely disappear from the frame.

Vehicle detection: A proactive strategy for selecting a search window for vehicle detection using image context, a GMM framework is proposed for vehicle detection with sequential movement with top-down attention. Consistently achieved satisfactory performance in identifying vehicles clear link box. proposed a systematic search strategy for detecting visual vehicles in the image, where the detection model proposed a deep RL framework to select the appropriate action to capture the vehicle in the image.

Vehicle Count: This module counts detected vehicles and the result of this count will be updated frequently based on vehicle detection, the result will be output for streaming video using OpenCV

Bag of Features Model

Model Visual Features (BOF) is one of the most important concepts in computer vision. We use avisual vocabulary model to classify image content. It is used to build a high volume tracking system (non-specific, precise). When we classify textures using textures, we instead use a visual vocabulary model. As the name suggests, the concept of "visual bag of words" is actually derived from the "bag of words" model used in information retrieval (eg, text-based search engines) and text analysis.

The general idea in Word Bag is to present a "document" (ie a web page, a Word file, etc.) as a collection of important keywords, completely ignoring the order in which the words appear. Documents that share the same keyword are considered to belong to each other regardless of the order of the keywords. Also, because we completely ignore the order of words in a document, we call this representation "bag of words" instead of "list of words" or "list of words": Treating a document as a "bag of words" allows us to efficiently analyze and compare documents. because we don't need to store information about the order or location of words - we count how many times a word appears in a document, and then use the frequency number for each word as a way to rate the document. In computer vision, we can use the same concept - only now instead of working with keywords, our "words" are now layers of images and related feature vectors:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, January 2024

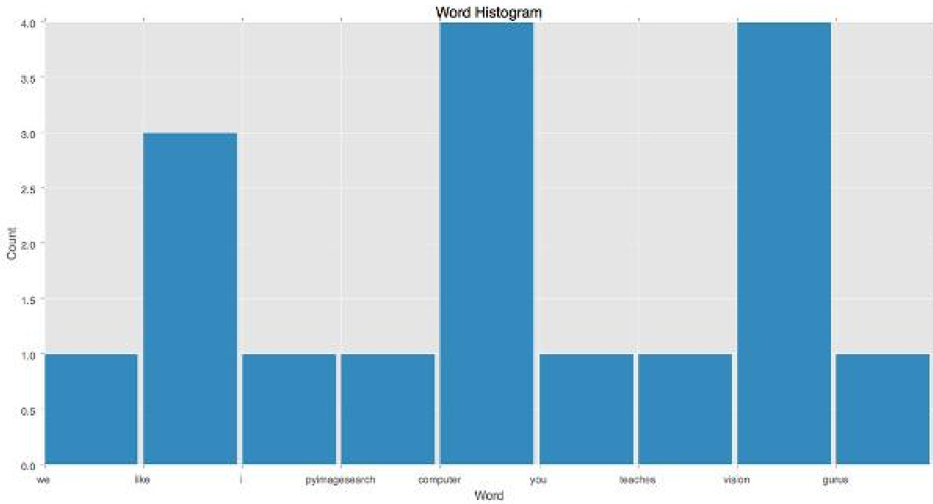


Fig. 2: An example of taking a blob of text and converting it into a word histogram.

Given a dictionary of possible visual words, we can then count the number of times each visual word appears and visualize it as a histogram. This histogram is a veritable bag of visual words. Building visual vocabulary can be divided into three steps

Step#1: Feature Extraction

The first step in building a visual bag of words is to extract descriptors and features from each image in our database.

Feature extraction can be done in several ways: identify key points and extract SIFT features from key regions of our image; using loops in regular intervals (for example, solid button detectors) and derivatives of other forms of local invariant descriptors; or we can extract the average RGB value from a random image region.

The point here is that for each input image we get several feature vectors:

Step#2: Dictionary/Vocabulary Construction

After extracting feature vectors from each image in our database, we need to build a vocabulary of possible visual words.

Word formation is usually done through a k-means clustering algorithm that summarizes the feature vectors obtained from step 1.

The centers of the resulting clusters (eg, centroids) are considered as visual word dictionaries.

Step#3: Vector Quantization

Given an arbitrary image (whether from our original database or not), we can identify and abstract the image using a bag of image words using this process:

Extract the feature vector as in step 1 above.

For each extracted feature vector, count its nearest neighbors in the dictionary created in step 2 - this is usually done using Euclidean distance.

Take the set of nearest neighbor labels and construct a histogram of length k (the number of clusters formed by k-verbs), where the i value in the histogram is the frequency of the i-visual word. When modeling an object by distributing prototype vectors, this process is usually called vector quantization.

Classification

One of the interesting features of our network is its simplicity: the classifier is only replaced by a masking layer, without any prior or convolutional structure. However, it should be prepared with a large amount of training data: vehicles of different sizes should appear almost everywhere.

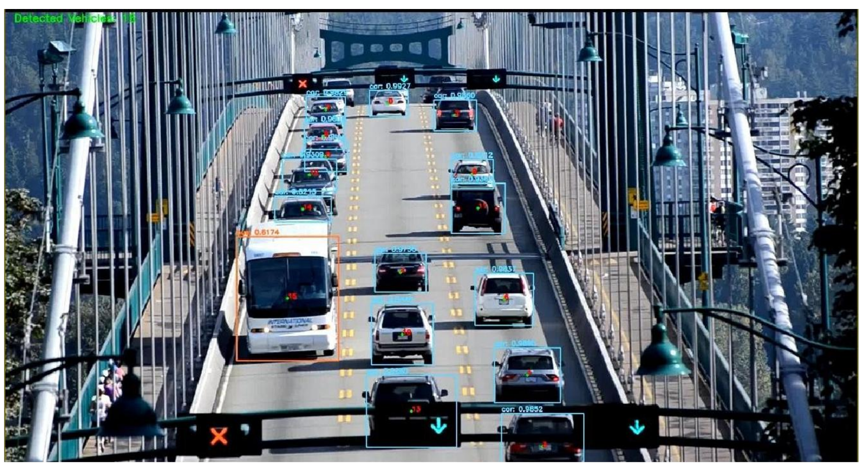
Visual tracking solves the problem of finding a target in a new frame from the current position. The proposed tracker dynamically tracks the target with sequence movement controlled by GMM. GMM predicts the motion to run the target from its position in the previous frame. The intersection box moves with the movement predicted from the previous state, and the next movement continues to be predicted from the moved state. We solve the vehicle tracking problem by repeating this process in a series of tests. GMM excels in RL as well as SL. Online adaptation takes place during real tracking.

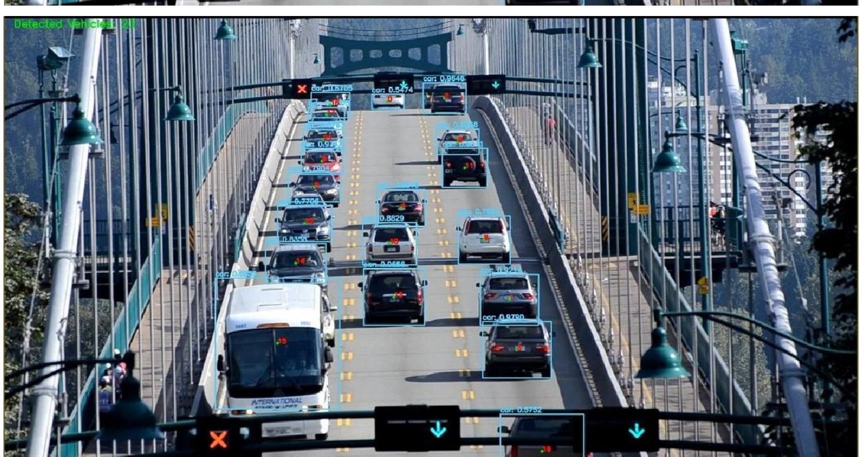
A GMM is designed to generate motion to find the location and size of the target vehicle in a new frame. The GMM algorithm learns a policy that selects the most optimal action to follow from the current situation. In GMM, a policy system is developed in which the input is a truncated image layer in the previous state and the output is the probability distribution of actions such as translation and scale change. The process of choosing this course of action requires a bit of research step more than the sliding window or candidate sampling approach. Furthermore, since our method can localize the target by selecting the motion, post-processing such as box regression is not required

Advantages of Proposed System

- Identify high-moving vehicles in video sequences.
- Vehicle tracking is unlocked.
- Identify the type of vehicle.
- Calculate the amount of traffic through video.

IV. RESULTS AND DISCUSSION





IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, January 2024

V. CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

The proposed solution is implemented in python using OpenCV bindings. Camera images from multiple sources are processed. A simple interface was developed for users to select regions of interest for analysis, and image processing techniques were used to count the number of vehicles and classify vehicles using machine learning algorithms. From the experiment, it can be seen that the CC method outperforms the BoF and SVM methods in all results and provides classification results closer to the ground truth value.

5.2 Future Scope

One of the limitations of the system is that it is not effective in detecting vehicle obstacles, which affects the accuracy of counting and classification. This problem can be solved by introducing secondary feature classification, such as color-based classification. Another limitation of the current system is the need for human supervision to determine areas of interest. To calculate the vehicle, the user must determine an imaginary line that intersects the center of the contour, so the accuracy depends on the judgment of the human observer. In addition, the camera angle affects the system, so camera calibration techniques can be used to determine the path to see the road better and improve efficiency. This system is not able to detect vehicles at night because it requires the visibility of objects in front for the extraction of contour features, as well as classification features using SIFT features. The system can also be optimized for greater accuracy using more sophisticated image segmentation and artificial intelligence processes.

REFERENCES

[1] Alpatov, Boris & Babayan, Pavel & Ershov, Maksim. (2018). Vehicle detection and counting system for real-time traffic surveillance. 1-4. 10.1109/MECO.2018.8406017.

[2] Song, H., Liang, H., Li, H. et al. Vision-based vehicle detection and counting system using deep learning in highway scenes. Eur. Transp. Res. Rev. 11, 51 (2019). https://doi.org/10.1186/s12544-019-0390-4.

[3] Neupane, Bipul et al. “Real-Time Vehicle Classification and Tracking Using a Transfer Learning-Improved Deep Learning Network.” Sensors (Basel, Switzerland) vol. 22,10 3813. 18 May. 2022, doi:10.3390/s22103813.

[4] C. J Lin, Shiou-Yun Jeng, Hong-Wei Lioa, "A Real-Time Vehicle Counting, Speed Estimation, and Classification System Based on Virtual Detection Zone and YOLO", Mathematical Problems in Engineering, vol. 2021, Article ID 1577614, 10 pages, 2021. https://doi.org/10.1155/2021/1577614.

[5] M. S. Chauhan, A. Singh, M. Khemka, A. Prateek, and Rijurekha Sen. 2019. Embedded CNN based vehicle classification and counting in non-laned road traffic. In Proceedings of the Tenth International Conference on Information and Communication Technologies and Development (ICTD '19). Association for Computing Machinery, New York, NY, USA, Article 5, 1–11. https://doi.org/10.1145/3287098.3287118.

[6] A. Arinaldi, J. A. Pradana, A. A. Gurusinga, “Detection and classification of vehicles for traffic video analytics”, Procedia Computer Science, Volume 144, 2018, Pages 259-268, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.10.527.

[7] Gomaa, A., Minematsu, T., Abdelwahab, M.M. et al. Faster CNN-based vehicle detection and counting strategy for fixed camera scenes. Multimed Tools Appl 81, 25443–25471 (2022). https://doi.org/10.1007/s11042-022-12370-9.

[8] G. Oltean, C. Florea, R. Orghidan and V. Oltean, "Towards Real Time Vehicle Counting using YOLO-Tiny and Fast Motion Estimation," 2019 IEEE 25th International Symposium for Design and Technology in Electronic Packaging (SIITME), 2019, pp. 240-243, doi: 10.1109/SIITME47687.2019.8990708.

[9] L. C. Pico and D. S. Benítez, "A Low-Cost Real-Time Embedded Vehicle Counting and Classification System for Traffic Management Applications," 2018 IEEE Colombian Conference on Communications and Computing (COLCOM), 2018, pp. 1-6, doi: 10.1109/ColComCon.2018.8466734.

[10] D. E. V. Tituana, S. G. Yoo and R. O. Andrade, "Vehicle Counting using Computer Vision: A Survey," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), 2022, pp. 1-8, doi: 10.1109/I2CT54291.2022.9824432.

[11] A. Khan, A., Sabeenian, R.S., Janani, A.S., Akash, P. (2022). Vehicle Classification and Counting from Surveillance Camera Using Computer Vision. In: Suma, V., Baig, Z., K. Shanmugam, S., Lorenz, P. (eds) Inventive Systems and

Control. Lecture Notes in Networks and Systems, vol 436. Springer, Singapore. https://doi.org/10.1007/978-981-19-1012-8_31.

[12] W. Balid, H. Tafish and H. H. Refai, "Intelligent Vehicle Counting and Classification Sensor for Real-Time Traffic Surveillance," in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 6, pp. 1784-1794, June 2018, doi: 10.1109/TITS.2017.2741507.

[13] N. Jahan, S. Islam and M. F. A. Foysal, "Real-Time Vehicle Classification Using CNN," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225623.

[14] M. A. Butt, A. M. Khattak, S. Shafique, B. Hayat, S. Abid, Ki-Il Kim, M. W. Ayub, A. Sajid, A. Adnan, "Convolutional Neural Network Based Vehicle Classification in Adverse Illuminous Conditions for Intelligent Transportation Systems", Complexity, vol. 2021, Article ID 6644861, 11 pages, 2021. https://doi.org/10.1155/2021/6644861.

[15] P. Gonzalez, Raul & Nuño-Maganda, Marco Aurelio. (2014). Computer vision based real-time vehicle tracking and classification system. Midwest Symposium on Circuits and Systems. 679-682. 10.1109/MWSCAS.2014.6908506.