VEMON: Vision-based Vehicle Counter for Traffic Monitoring

Jonathan Cempron and Joel Ilao*
Computer Technology Department College of Computer Studies
De La Salle University
**Corresponding Author: joel.ilao@dlsu.edu.ph*

Abstract: VEMON: Vision-based Vehicle Counter for Traffic Monitoring is a traffic monitoring system that counts vehicles according to classification, estimates vehicular speed, and estimates ambient air quality using a traffic camera video stream or file as input. The following are the pre-defined vehicle classifications that are used: sedan, Sports Utility Vehicle (SUV), truck, van, bus, motorcycle, and jeepney. The system's vehicular classifications and counts were shown to agree with human annotations through correlational analysis.

Key Words: Traffic Monitoring, Computer Vision, Image Processing

1. INTRODUCTION

Road traffic congestion in the Philippines, particularly in its urban and economic centers continues to be a challenging problem for the local and national government, with various solutions being actively adopted and explored to mitigate its effects. An important component of these solutions are digital traffic monitoring systems that provide road- or highway- status reports indicating vehicular density and congestion levels (Aguirre, Alcantara & Trinidad, 2013). Traffic monitoring technology makes use of real-time video feeds from Internet Protocol (IP) cameras or Closed Circuit Television (CCTV) systems placed in areas of interest (Ambardekar et al., 2007; Chiu et al. 2010; Chen et al., 2012; Kiratiratanpruk & Siddhichai, 2009; Conci et al., 2011), which are continuously processed to generate traffic data at regular intervals.

The Metropolitan Manila Development Authority (MMDA) has installed a network of traffic surveillance cameras, whose feeds are monitored by personnel located at a centralized facility. This service is complemented by the Metro Manila Traffic Navigation (TNAV) through a partnership between MMDA and TV5 (MMDA, 2011). TNAV is a mobile and web-based system that provides regular traffic information for major roads in Metro Manila: EDSA, Quezon Ave., and Ortigas.

Traffic monitoring systems and other related projects have also been developed at De La Salle University, and particularly from the Center for Automation Research of the College of Computer Studies. The Archer's Eye broadcasted real-time video streams from IP cameras installed at strategic locations in DLSU (2011). Various web-accessible traffic monitoring systems have also been developed, such as TraVis (Aguirre, Alcantara & Trinidad, 2013) and VeMon (Batingal et al., 2016). To expand the systems' functionality and improve their performance, a CNN-based vehicle classification system was developed (Bautista et al., 2016), and occlusion handling through the use of color information was explored (Castillo & Tejada, 2017). A speed estimation module for tracked vehicles, with simplified camera calibration procedures, was also developed (Ibarrientos, 2018). Ambient air quality estimation based on vehicular traffic and meteorological conditions is currently being developed (Malana, 2018). These recent systems capitalize on findings and recommendations from earlier works (Dang et al., 2006; Lai et al., 2010).

In this paper, we present the latest implementation of the traffic monitoring system named "Vision-based Vehicle Counter for Traffic Monitoring" (VEMON). It takes as an input a traffic video stream or a traffic video file. The minimum video quality specifications to guarantee acceptable performance are 640x480 pixel resolution and a frame rate of at least 6 frames-per-second. The ambient lighting conditions should be bright enough not to obscure discernable visual features of vehicles; hence, the system is only guaranteed to perform reasonably on videos recorded during the day time, i.e. from 6 AM to 6 PM. Presently, the system can produce the following reports: vehicle count per classification, speed estimate, and air quality estimate. The pre-defined vehicle classifications are the following: sedan, SUV, truck, van, bus, motorcycle, and jeepney. We organize the rest of the paper as follows: Section 2 illustrates the VEMON network architecture while Section 3 describes the system's implementation. Section 4 contains the system performance testing and results. Conclusions are summarized in Section 5, while Section 6 contains our plans for future work.

2. NETWORK ARCHITECTURE

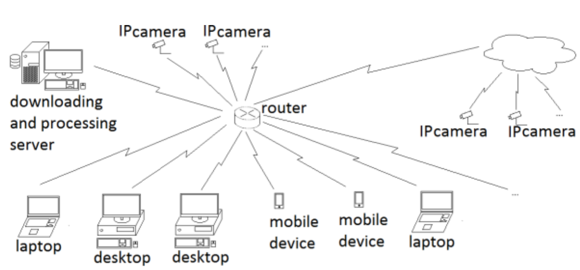

Figure 1: VEMON Network Architecture

Figure 1 shows the network architecture of VEMON. Backend processing is performed at the downloading and processing server. When an internet connection is available, video streams can come from two sources, namely local IP cameras, and Web IP Cameras. If there is no internet connection, only local IP camera video streams can be accessed Data can only be viewed computers within the local network where the server is connected. Managing data can only be performed at the downloading and processing server.

3. METHODOLOGY

The workflow of traffic video processing, shown in Figure 2, is described as follows: An input video module extracts individual image frames from the input video file or stream. Background modeling and removal is next performed as preparation for Object Tracking. The tracked vehicle objects are next classified. Occlusion handling is activated for scenarios wherein detection fails, in order to separate occluding vehicles. The speed estimation module approximates the actual speeds of tracked vehicles in the video. An air quality estimation module next estimates the $PM_{2.5}$ levels based on vehicular counts and ambient weather conditions. Finally, the web-based output module generates reports that can be viewed from devices connected to the local network.

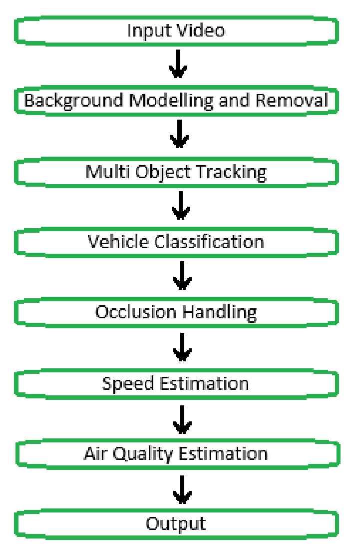

Figure 2: Traffic Video Processing Workflow

3.1 Inputs and Outputs

The input could be any traffic video file or IP camera address. The outputs are the following: traffic video with superimposed object detection visualization, vehicular count per classification in text and pie chart form, histogram of speeds detected, and air quality estimates.

3.2 Background Modeling and Removal

For this system, foreground objects are defined as the vehicles passing by a road segment. For the Background Modelling and Removal module, two moving object detection methods were tested: (1) Background modeling using Approximate Median Method, and Background Subtraction (Aguirre et al., 2013; Batingal et al., 2016, and (2) Background Subtraction using Gaussian Mixture Model. The Gaussian Mixture Model was chosen over the Approximate Median Method because it is faster.

a b

Figure 3: The Background Modeling and Removal process: (a) Image Binarization, and (b) initial selection of candidate vehicles via Blob Analysis

Image Binarization via Thresholding was employed to transform detected vehicle objects into blobs, as shown in Figure 3a. The blob's location and size information are used to compute for the centroid of the vehicle, which is used for tracking the movement of the vehicle throughout the video. The blob's size is also used to determine if a binary blob is a vehicle. The vehicle classification module would determine if the blob is indeed a vehicle.

3.3 Multi-Object Tracking System

Object tracking is used to follow a detected object within successive video frames. The goal of the tracking algorithm was to reduce the number of false re-introduction of an object, such that only a single detection count will be registered for a single object.

a b c

Figure 4: The object tracking process: (a) a sample video frame, (b) a blob detected as a vehicle enclosed in a red box, (c) the vehicle's trajectory indicated by a plot of the corresponding blob's centroid over successive frames

Refer to Figure 4 for an illustration of the Object Tracking process. We implemented our Multi-Object Tracking algorithm by enforcing a motion model constraint when matching blobs in successive frames. The motion model is implemented as follows:

1. *Initialization:* All the blobs in the first frame are considered as new and separate objects. Each object is assigned a tracking identification number (trackId).
2. For successive frames, each blob is checked based on a distance metric relative to blobs in the previous frame:
    a. If the new blob is *close enough* to a blob in the previous frame, it will be assigned with the same trackID as the one from the previous frame.
    b. If the new blob is not near to any of the blob from the previous frame, it will be assigned with a new trackId.

The implemented distance metric matches the centroids of blobs from successive frames if their locations are *close enough* by a fixed threshold of 30 pixels.

Termination of object tracking is not implemented as it is already compensated for by implementing a baseline method for counting, which only counts a vehicle when it reaches the middle of the frame.

3.4 Vehicle Classification and Occlusion Handling

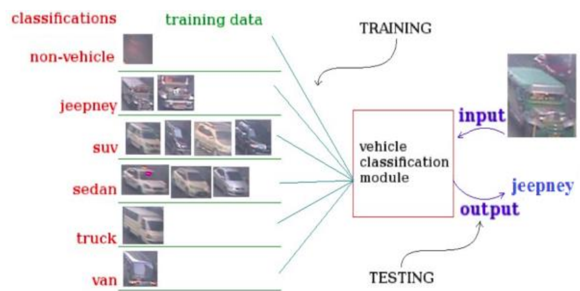

Figure 5: Vehicle Classification Module

Deep Learning Framework was used for vehicle classification according to the following categories: non-vehicle, sedan, Sports Utility Vehicle (SUV), truck, van, bus, motorcycle, and jeepney (Bautista et al. 2016). This is illustrated in Figure 5.

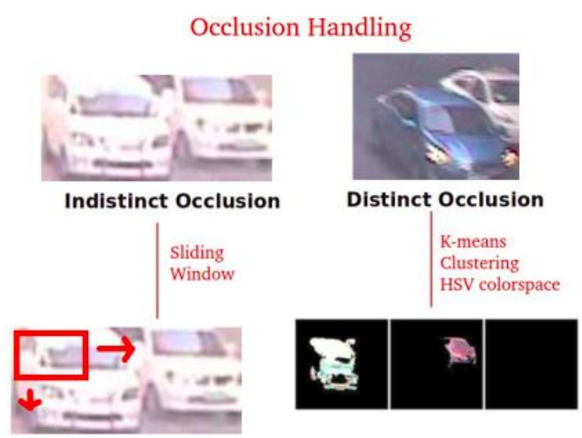

Figure 6: Illustration of Occlusion Handling

When the detection fails to separate cars, occlusion handling is performed, shown in Figure 6. We used K-means clustering in Hue-Saturation-Value (HSV) color space to handle 'distinct' occlusions, defined as when occluding vehicles are of different colors. A sliding window approach was used when occluding vehicles are of similar colors (Castillo, Tejada, Cordel & Ilao, 2017).

3.5 Speed Estimation

Speed was estimated by Vanishing Point Detection and Perspective Transformation. Pixel distance to real-world distance transformations were automatically generated with the assumption that sedan-type tracked objects in the scene have standard metric lengths (Ibarrientos, 2018).

3.7 Air Quality Estimation

For ambient air quality estimation, $PM_{2.5}$ levels of the road vicinity are estimated based on vehicle counts by classification, and ambient meteorological information such as humidity and temperature. The models explored are Multiple Linear Regression (MLR) and Support Vector Machines (SVR) and checked for its Mean Squared Error (MSE) compared to actual sensor reading of $PM_{2.5}$ (Malana, 2018).

3.8 Resource Management

Two daemons manage system resources, for video acquisition and for video processing. A daemon for scheduled downloading of videos from IP cameras was developed to ensure that several videos downloaded will not be encoded at the same time, to avoid resource contention. Another daemon for processing of videos was also developed to enforce queuing such that only a single video will be processed at a time.

3. TESTING AND RESULTS

Two days' worth of video was used to test VEMON. The camera perspective was from Andrew Building facing Taft Avenue. The videos are recorded from October 10, 2015 to October 11, 2015, from 6 AM to 6 PM daily. The video files were cut into 5-minute chunks. Each five-minute chunk was analyzed by counting the number of vehicles per classification in the video. System performance was measured by comparing the counts and classifications made by human encoders and the output of the VEMON system, as shown in Figure 7.

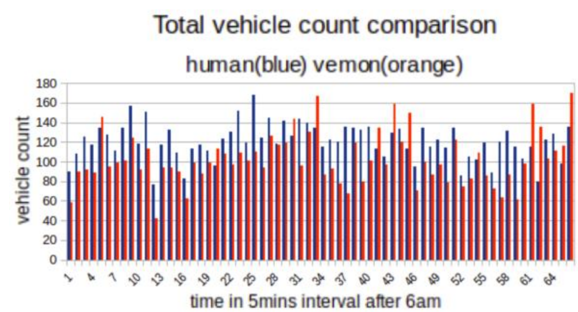

Figure 7: Comparison of VEMON output and Human annotations with respect to vehicle counts

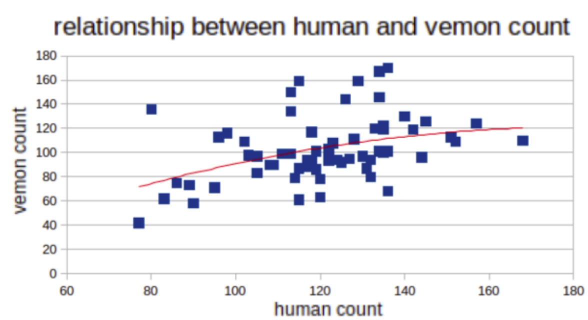

Figure 8: VEMON and Human count correlation

Figure 8 shows a scatterplot of the human counts relative to VEMON system counts. The actual Correlation Coefficient of VEMON count to human count is 0.393.

5. CONCLUSION

In this paper, we described the VEMON traffic monitoring system that can tap into traffic video cameras and report vehicle counts per classification, vehicle speeds, and air quality estimates. Image processing techniques such as image binarization and morphological transformations were used for object detection and tracking. To determine the classification of a vehicle, deep learning approach was used. Perspective transformation and vanishing point detection were used in estimating vehicular speed with minimal calibration. Machine learning was used for air quality estimation.

The completed system's outputs were compared to human counts, yielding a correlation value of 0.393. Cursory inspection of the dataset, shows that human annotations for certain categories, e.g. vans vs. SUVs, get interchanged, which could have affected the system's performance for some of the categories. A controlled environment for traffic scenarios is ideal for setting ground truth data but this is quite costly to actually set up.

While speed estimation and air quality estimation are additional features of the system, only the vehicular counting feature was tested for accuracy. The reason was that speed was difficult for humans to visually estimate. Ground truth for air quality can be sourced from a $PM_{2.5}$ sensor placed beside the camera; however, we have yet to prepare such equipment.

6. FUTURE WORK

For future work, we will expand the capabilities of the system, by designing adaptive algorithms that can handle extreme lighting conditions, such as dim conditions taken during dawn or dusk, or very bright conditions taken during noon-time. Another future work is to develop techniques for non-constant video frame rates caused by faulty internet connections during downloading. To accomplish real-time processing, algorithms particularly the multi-object tracking algorithm, should be improved. To increase classification and detection accuracy, the human-annotated training dataset has to be expanded. Finally, VEMON can be used to assess the quality of videos taken by traffic cameras, which can serve as basis for identifying cameras due for maintenance or replacement, if deployed for organizations that manage a network of surveillance cameras.

8. ACKNOWLEDGMENTS

The VEMON project was funded by the Department of Science and Technology-Philippine Council for Industry, Energy, and Emerging Technology Research and Development (DOST PCIEERD) and De La Salle University Research Coordination Office (DLSU URCO).

9. REFERENCES

N. Aguirre, J. Alcantara, and J. Trinidad. (2013). *TraVis: Web-Based Vehicle Counter with Traffic Congestion Estimation Using Computer Vision.* Undergraduate thesis. Comp. Tech. Dept., De La Salle Univ., Manila, Philippines.

A. Ambardekar, G. Bebis, and M. Nicolescu. (2007). *Efficient Vehicle Tracking and Classification for an Automated Traffic Surveillance System*. Department of Computer Science and Engineering, Nevada.

T. Batingal, A. Camering, R. Naguit, and T. Recato Dy. (2016). *VeMon: Web-Based Vehicle Traffic Monitoring System with Traffic Flow Analysis*. Undergraduate thesis, Comp. Tech. Dept., De La Salle Univ., Manila, Philippines.

C. Bautista, C. Dy, M. Manalac and R. Orbe. (2016). *FastCon: Fast Implementation of Convolutional Neural Network for Vehicle Detection and Vehicle Type Classification*. Undergraduate thesis, Comp. Tech. Dept., De La Salle Univ., Manila, Philippines.

RDM Castillo, MMB Tejada, MO Cordel, AFB Laguna, and JP Ilao (2017), *Vsion: Vehicle Occlusion Handling for Traffic Monitoring*, International Conference on Video and Image Processing, December 27- 29, 2017, NTU, Singapore

A. Conci, E. Nunes, A. Sanchez, and P. Suarez. (2011). *Video-Based Distance Traffic Analysis: Application to Vehicle Tracking and Counting*. Computing in Science and Engineering, vol. XII, no. 3, pp. 38-45.

C. Chiu, M. Ku, and C. Wang, (2010). *Automatic Traffic Surveillance System for Vision-Based Vehicle Recognition and Tracking*. Journal of Information Science and Engineering, no. 26, pp. 611-629.

C. Chen, W. Hu, D. Huang, Y. Lin, and S. Yi, (2012). *Feature-Based Vehicle Flow 1Analysis and Measurement for a Real-Time Traffic Surveillance System*. Journal of Information Hiding and Multimedia Signal Processing, vol. III, no.

M. Dang, A. Daquioag, T. Ramos, F. See. (2006). *Hypothesis Generation of Vision-based Vehicle Tracking*. Undergraduate thesis. College of Computer Studies, De La Salle University. Manila, Philippines.

DLSU. (2011). *Archer's Eye: Looking Out For You*. Date accessed: April 2015. [Online]. http://archers-eye.dlsu.edu.ph/

C. Ibarrientos. (2018). *Automatic Calibration and Speed Estimation on Multi-Directional Vehicle Flow*. M.S. thesis. Soft. Tech. Dept., De La Salle Univ., Manila, Philippines.

K. Kiratiratanapruk and S. Siddhichai. (2009) *Practical Application for Vision-Based Traffic Monitoring System.* 6th International Conference on Electrical Engineering / Electronics, Computer, Telecommunications, and Information Technology, Chonburi, pp. 1138-1141.

F.P. Lai, C. Leong, R. Ortuoste, P. Yu and C.Y. Ong (2010). *Vision-based Intelligent System for Traffic Analysis (VISTA)*. Proceedings of the 5th ERDT Conference, Manila, Phils. 10 Sept 2010.

M. Malana. (2018). *Air Quality Estimation System Using Vehicular Traffic and Meteorological conditions*. M.S. thesis. Software Tech. Dept., De La Salle University, Manila, Philippines

MMDA. (2011). *Interaksyon / TV5-MMDA Traffic Monitoring System*. Date accessed: April 2015. [Online]. http://mmdatraffic.interaksyon.com/system-view.php

Wafi, Z., Ahmad, R., & Paulraj, M. (2010). *Highways Traffic Surveillance System (HTSS) Using OpenCV*. Control and System Graduate Research Colloquium, 44-48.