

VEHICLE DETECTION AND CLASSIFICATION FOR CLUTTERED URBAN INTERSECTION

Habibu Rabi

Department of Electrical Engineering, Faculty of Technology
Bayero University Kano, Nigerian.
hrabiuele@buk.edu.ng

ABSTRACT

As more vehicles continuously appear on our roads by the day causing congestion and accidents. A traffic monitoring system capable of detecting, counting and classifying the passed vehicles is needed to provide in advance information to relevant authorities on the road traffic demand. Background subtraction and kalman filter algorithm are used to detect and track individual vehicles throughout the detection zone. The detected vehicles blob-area is used to trigger the segmentation unit which inturn extracts the vehicle while at a point closest to the camera. Finally, both geometric and appearance features of the segmented vehicles are passed to the LDA classifier for proper categorisation. The system achieved a high counting performance of 97.37% with corresponding classification rate of 87.82 %

KEYWORDS

Vehicle detection, vehicle tracking, vehicle classification, background subtraction, Kalman filter, LBP features, LDA classifier

1. INTRODUCTION

Visual based vehicle tracking and traffic surveillance have received considerable attention during the last two decades. The subject has a great application in the field of vehicle surveillance system, vehicle tracking and automatic road way and intersection monitoring. Automatic monitoring of traffic helps for early detection of possible road congestion in a timely way. Traffic congestion occurs when the traffic demand exceeds the road capacity, which leads to time wastage, energy lost and unnecessary emission of carbon into the atmosphere. Visual based road management system offer a seamless means of vehicle counts and classification. Vehicle counts and classification are important factors for computation of the number of vehicles and percentages of their classes that run on a particular road. Furthermore, such information helps the relevant authorities to plan for timely expansion of the existing road or construction of new ones.

In this project we use a single camera mounted on a pedestrian bridge directly looking down on traffic scene. The camera is placed at the centre of the road to allow for multi lane detection and minimise vehicle occlusion. In this paper we present a novel vehicle classification method which is based on the both vehicle's geometric and appearance features. The paper is organised as follows. In section 2 we present an overview of the related work. Section 3 gives a brief system's overview. Section 4, describes the methodology. Experimental results are presented in section 5. Finally, conclusions are discussed in section 6.

1.1. Related work

A number of approaches for vehicle detection, tracking and classification have been developed in the past using different attributes of the vehicles. There are several methods for detecting moving vehicles taken from a stationary camera. Generally, they are categorised into three groups; background subtraction, frame by frame differencing and optical flow approaches. In background subtraction method, the object is detected by pixel wise subtraction between the current frame and the background frame. Using some threshold limit, all pixels belonging to object (that are not present in the background image) are detected and grouped together. A comprehensive survey on vehicle detection appears in [1]. Tracking of moving vehicles in video sequences has been an active research topic in field of computer vision. The main idea of the tracking processes is to keep the identity of each detected vehicle over the whole sequences of the video. Thus, eliminates possible multiple counts in the vehicle counting stage. Several vehicle tracking algorithms are presented in [2-7].

Even though, a great deal of research has been reported on the vehicle detection and tracking. However, no much work is done on the vehicle classification. This is due to inherent difficulties associated with vehicle classification problem. The much wide variety of vehicle shapes and colours within same class, make the classification task even more challenging. The main goal of vehicle classification is to categorise the detected vehicles into their respective classes. Majority of the papers reported in the literature on vehicle classification are based on model-based approach [8-10] and recently [11] which is based on the vehicles colour. The model-based classification relies on sets of 3D models that coarsely describe the shapes of each vehicle type. Perhaps the most serious disadvantage of the model-based approach is the need to represent various shapes of vehicles (although belonging to the same class) with a single model.

2. GENERAL SYSTEM OVERVIEW

Figure 1, gives a general overview of the proposed system, which pictorially shows various stages of the system. The system basically consists of three main steps; vehicle detection, tracking and classification.

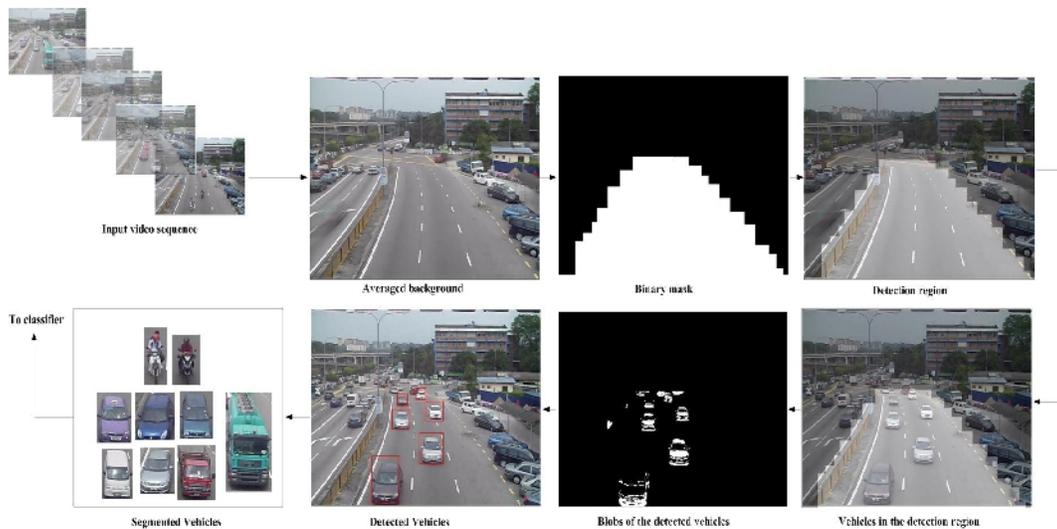


Figure 1. Pictorial flow diagram of the proposed system

2.1. System overview

The proposed system consists of the following stages:

- 1) Background estimation: the background is estimated from few initial frames of the video.
- 2) Detection region: in this stage, we constructed a binary mask to define our region of interest.
- 3) Object detection: here, we employ background differencing to detect the foreground objects and group the connected pixel into blobs.
- 4) Object tracking: bounding boxes are placed on each detected blob and track from frame to frame using Kalman filter.
- 5) Object segmentation: each track object is extract just before it starts leaving the detection zone.
- 6) Object classification: the segmented objects are passed to classifier to categorise them into their appropriate classes.

Section 3 provides detailed description of the stages.

3. MATERIALS AND METHODS

In this work, we utilised a video stream recorded using a single camera placed at high level. Placing camera at high level and looking down to the centre of the scene provides clearer distinctions among the closely moving vehicles. This helps in reducing possible occlusion problem from the objects in the scene. The video used in this work was captured from a busy urban intersection. Unlike high-way, urban intersections present additional challenges, besides the common ones; global illumination changes, shadow handling, occlusion and multiple object detection and tracking. These problems are characterised by cluttered background such as; buildings, pedestrian, parked vehicles, stopping vehicles, slow moving, fast moving, turn and u-turn. Therefore, it is the goal of this paper to develop a low-cost vision-based vehicle detection and classification system that can improve the much desired intelligent-transportation system.

3.1. Background Estimation

The first stage of the algorithm is generation of fairly stable background. Several methods are used in the past for background extraction and update from images of moving vehicles in video. Garcia and Shu, modelled the background using a single frame of the scene containing no object of interest [12]. The image is then smoothed to reduce lighting effects. In [10] Gupe et al, modelled the background using a weighted average of the previous frame and that of the current frame of the video sequence. Since such images might also contain foreground images. Thus, prior to background update the current image is classified into background and foreground. Only pixels belonging to background of the current frame are used in pixel updating stage. Long et al. in [4] used three stages approach to model the background. Firstly, some loosed video frames were selected. Next, frame difference integration was applied to remove pixel movements corresponding to vehicle in the selected frame. Finally, using median value method each pixel value of the background is estimated. Gloyer et al. [6] Proposed an interesting method for background modelling. The assumption here is that, intensity of pixels belonging to the road remains relatively same across initialisation frames as compared to that of the passing vehicles.

Thus, intensity values that have high occurrence within the initialisation frames are considered to be part of the background image. In this work, this method for background modelling is adopted.

3.2. Vehicle Detection

Object detection is usually accomplished by a simple background by current frame differencing and then followed by threshold [2]. The main idea here is to make a pair-wise subtraction between pixels of the background frame and that of the current frame. In this way pixels having value greater than certain threshold are considered to belong to the object (detected vehicle) as shown in figure 2. In this work, we use four distinct steps to detect pixels belonging to moving vehicles.

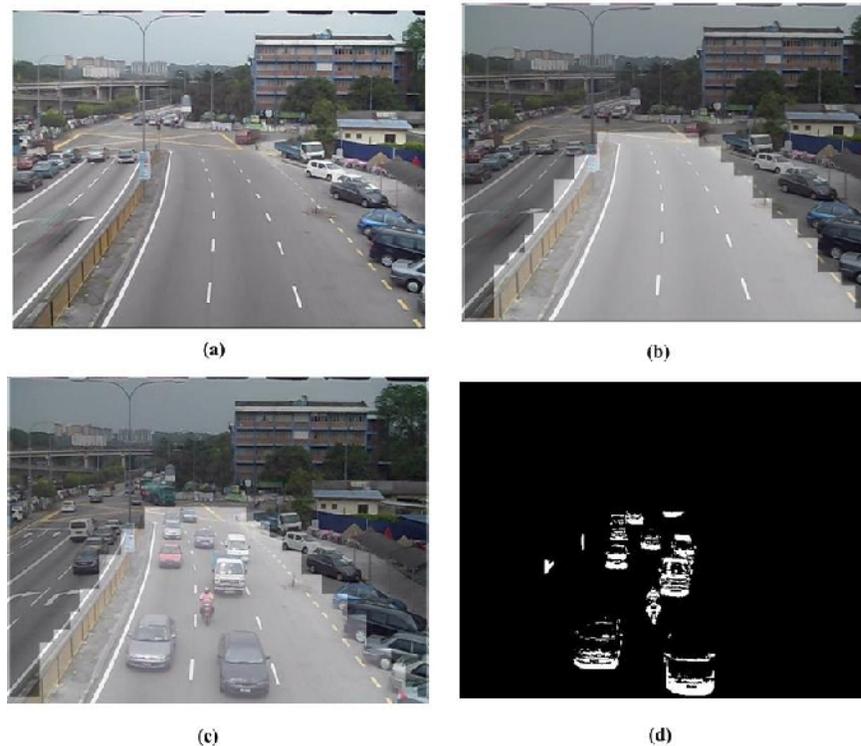


Figure 2. Vehicle detection: (a) averaged background (b) masked background with no object of interest (c) masked current frame containing objects in the detection zone (d) blobs of the detected objects

The first step into the proposed vehicle detection method is to construct a binary mask that defined the detection range; in this case three lanes road from the intersection to the camera location. The mask “M” has same size as the video frames with entries of ones in the region corresponding to the detection zone and zeros elsewhere. Next, we multiply both the background and the current image with the mask figure 2 (b) and (c), this way all areas outside the detection region are simply ignored or rejected (dark areas). We then apply the background subtraction technique to detect locations in the current image that have values greater than the set threshold. To further enhance the shapes of the detected blobs, we apply dilation then followed by erosion.

3.3. Vehicle Tracking

Tracking is the process of labelling the detected object and maintaining its identity across the whole video sequence. Essentially, there are four types of object tracking methods [13]. (1) Model based method, which relies on detailed geometric model of different objects. (2) Feature based method, tracks some selected salient features of the object instead of the whole object. (3) Active contour based method, the main idea here is to represent the detected object by its bounding contour and keep updating it throughout the detection region. (4) Region based method, employs background subtraction technique for initialisation. The detected objects are then traced using region tracking techniques such as Kalman filter.

Kalman filter is a recursive predictive method which is based on the state space techniques. It uses two steps method to estimates the state of a dynamic system; (1) the predict step, which uses dynamic model to predict the object state, (2) the correction step, it corrected the prediction using the measured or observed model, such that the error covariance of the estimator is minimised [14]. In dynamic image processing, Kalman filter plays an important role in predicting the position of a target in the current image frame f_i using the prior information of its location in previous frame f_{i-1} . This procedure is repeated for each time step "t" with state of the previous time as initial value as shown in figure 2. This method is adopted in this project, we employed Kalman filter to track the labelled blobs throughout the whole length of the detection zone.

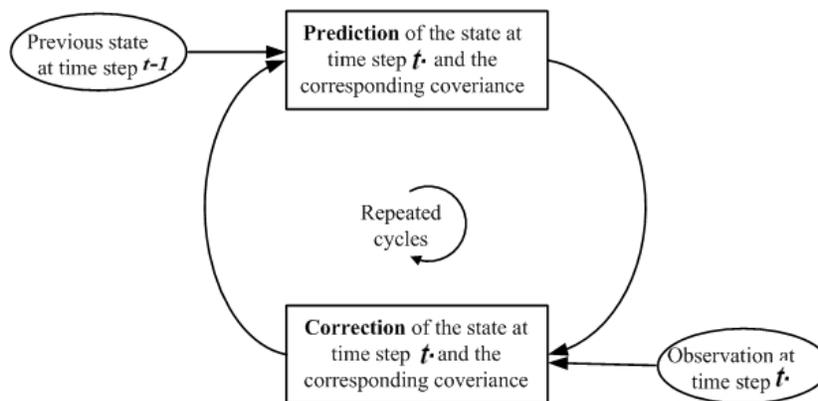


Figure 3. Diagram of Kalman Filter

As vehicles leave the intersection and enter the detection zone (top of the detection window) bounding boxes are added to each detected object. As the video streams, the objects in the scene advances towards the camera's location and their respective blob areas keep growing. We monitor these growing areas frame by frame until the object starts to leave the detection zone, in which case its area begin to fall. We therefore use this reverse in growing blob area as trigger for the segmentation unit. The interesting thing here is that, we segments the tracked vehicles at points when they are closest to the camera, thus their structures and colour are more visible and clearer. Cropping vehicle when it reaches it maximum size not only ease the classification task but also enhances its performance. Figure 4 illustrates the proposed tracking method. In the frame f_{t-5} , five vehicles appear in the detection zone and are detected accordingly. As the vehicles approach the camera position their blob areas increase frame f_{t-4} . However, the blob area of the white car occupying centre lane starts decreasing as the car leaves the detection zone frame f_{t-3} . Thus, we segment this car at frame f_{t-4} just before its blob starts falling. The same procedure is repeated among frames f_{t-2} , f_{t-1} and f_t , leading to extraction of two cars. The segmented vehicles

together with their current blob areas, centroids and bounding boxes coordinates are passed to the classification stage.

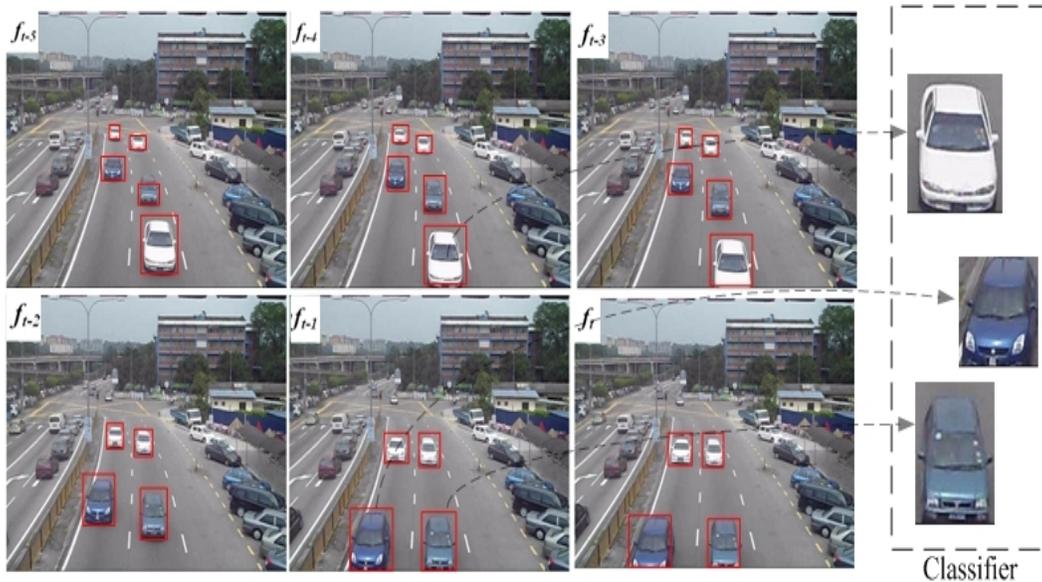


Figure 4. Illustrations of the proposed tracking and segmentation method

3.4. Vehicle Classification

In the field of pattern recognition, a classifier aims to identify the correct class of a given object based on some classification rules and characteristic of the object, which are also called feature vectors. Here, the main goal of this unit is to determine to which class the passing vehicle belongs to and keep track of the number of vehicles that have passed through the detection zone. The vehicles in the video are categorised into five groups; (1) motor cycles (2) cars (3) minibuses (4) trucks and (5) heavy trucks. The selection of these categories is based on the available vehicles types contained in the video. However, the system can easily be made to accommodate other vehicles types by simply including them in the training set. Generally, object recognition consists of three main stages; (1) object detection (2) feature extraction and (3) classification [15]. We have already accomplished the object detection in section 3.2. We now focus on feature extraction and classification.

3.5. Feature Extraction using Local Binary Pattern

The local binary pattern (LBP) operator is a powerful feature extractor which transforms an image into an integer labels statistics using a computationally efficient mean. Prior to the feature extraction process, some image pre-processing steps were applied to standardised and enhance the images. This includes; resizing each image into 64 by 64 pixels. Next, the RGB images are converted into grey images retaining only their intensity values. Furthermore, contrast enhancement technique was also applied to reduced illumination variations in the images. Figure 5 illustrates the pre-processing stage.

Following the image pre-processing, we employed local binary pattern (LBP) to extract unique attributes of each object. The local binary pattern (LBP) was first introduced by [16] as grey-scale and rotation invariant texture descriptor. The basic LBP operator labels the pixels P_n for ($n=0,1,\dots,7$) of an image by thresholding a 3 x 3 neighbourhood of each pixel with the value of

the centre pixel P_c and considers the result as binary number. Given a pixel at location (x_c, y_c) the resulting LBP at that location can be expressed as follows.

$$LBP_{(x_c, y_c)} = \sum_{n=0}^7 S(P_n - P_c)2^n \quad (1)$$

Where P_c is the grey-level value of the centre pixel and P_n are its eight surrounding neighbours and S is given as:

$$S(P_n, P_c) = \begin{cases} 1 & \text{if } P_n > P_c \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The histogram of the labels computed over each region can be used as local primitive descriptors, which robustly described the local changes in the region such as; flat area, curves, edges, spots etc. In this work, the procedure for extracting the LBP descriptors for vehicle representation was implemented as follows: first, the each pre-processed image was partitioned into 36 regions. Second, the LBP histogram from each region was computed. Third, the histogram ratio hr and maximum histogram value hm for each region was determined using equations 3 and 4. Finally, the 36 hr and 36 hm from all the regions were concatenated into a single feature vector. Additionally, the earlier extracted vehicle area, centroid and bounding box coordinates are further fused to the LBP feature to account for the vehicle's geometric features and its relative position on the road equation 5.

$$h_r = \frac{\sum_{i=(k/2)+1}^k h_{(i)}}{\sum_{i=1}^{k/2} h_{(i)}} \quad (3)$$

$$h_m = j, \quad j = \arg\{\max h_{(i)}\} \quad (4)$$

$$V_{LBP} = [h_{r1}, h_{r2}, \dots, h_{r36}, h_{m1}, h_{m2}, \dots, h_{m36}, a, b_1, b_2, b_3, b_4, c] \quad (5)$$

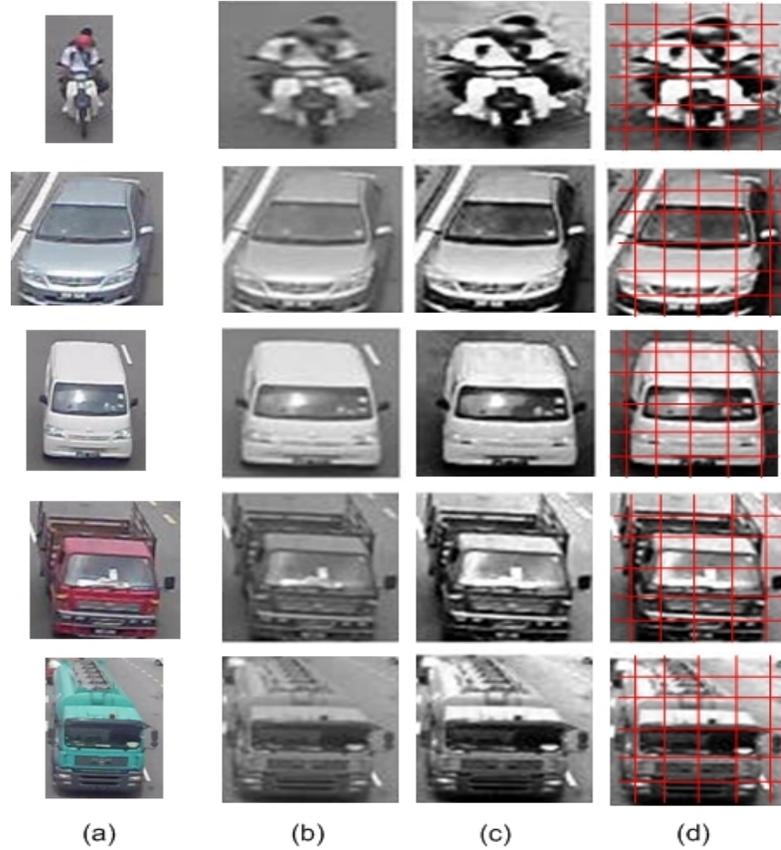


Figure 5. Examples of image pre-processing, (a) original image (b) resized intensity image (c) histogram equalisation (d) partitioned image

The formed feature vectors are then passed to linear classifier to categorise them into their respective categories. Linear discriminate analysis (LDA): attempts to classify a given input object into one of two or more categories based on features that characterize them, LDA attempts to finds linear decision boundaries from the features space that best separate the objects classes. This is achieve by maximizing the between class scatter matrix while minimizing the within class scatter matrix [17]. Mathematically, the scatter matrices are defined as follows:

$$S_W = \sum_c \sum_{i \in c} (x_j - m_i)(x_j - m_j)^T \quad (6)$$

$$S_B = \sum_c (x_i - m)(x_i - m)^T \quad (7)$$

and the class separation measure is given by:

$$P_i = \frac{|S_W|}{|S_B|} \quad (8)$$

Where, S_W and S_B stand for within class and between scatter matrices respectively, n_i is the number of training samples in class i , c is the number of class labels, m_i is samples mean of class i , x_i stands for i^{th} class sample set and x_j represents the j^{th} object of that class. $|S_W|$ and $|S_B|$ represents the determinants of within and between class matrices.

4. EXPERIMENTAL RESULTS

The performance of the proposed method was tested on a video sequence of 3,400 frames containing 76 vehicles. The video was recorded from typical urban intersection characterised by cluttered background. We segmented the video into two parts; the first part contains 2,200 frames was used in training the classifier. The remaining 1,200 frames were used during the testing phase. The system was successfully able to detected and tracked most vehicles correctly (97.37%). However, only 87.82% of the vehicles were correctly classified. Errors in the detection stage were mostly associated with occlusions, especially in the cases where the vehicles moves very closely (bumper to bumper) to each other. Even though, when the leading vehicle leaves the detection zone, the system is robust enough to detect the second vehicle separately. The misclassification cases were mostly between cars and minibuses. Although, both geometric and appearance attributes were used as discriminative descriptors, but due to strong similarities in both shapes and appearances between these classes, the system finds it very difficult to correctly classify them. Tables 1 through 3 summarise the results.

To validate the result, two persons were requested to independently count and classify the vehicles into the considered classes from both training and testing clips. The results from the two persons were simply averaged as shown in table 1 and used as a ground-truth.

Table1. Ground-truth result for the training and testing sequences

	Training sequence	Testing sequence	Total
Motor cycles	9	4	13
cars	30	13	43
Trucks	7	4	11
Minibuses	5	3	8
Heavy trucks	1	0	1
Total number of vehicles			= 76

Table 2. System's counter output for both training and testing sequences

	Training sequence	Testing sequence	Total
Motor cycles	9	4	13
cars	28	13	41
Trucks	7	4	11
Minibuses	5	3	8
Heavy trucks	1	0	1
Total number of vehicles detected			= 74

Table 3. System classifier's output for testing sequences

	Motor cycles	cars	Trucks	Minibuses	Heavy trucks
Motor cycles	4	0	0	0	0
cars	0	11	0	2	0
Trucks	0	0	4	0	0
Minibuses	0	1	0	2	0
Heavy trucks	0	0	0	0	0

At the end of the experiments, the proposed system's detection and classification accuracies were compared against the ground-truth Tables (1). It can be seen from Table 2 that the proposed system had detected, tracked and counted 74 objects as the number of passed vehicles in both the training and testing sequences against the actual number 76 in the ground truth. This corresponds to an error of 2.63%. In the classification stage, the system classified all the motor cycles, trucks and heavy trucks accurately. However, we recorded a global classification error of 12.18%, which resulted due to systems confusion between minibuses and cars. It is obvious that, the passengers sitting capacity of the two vehicle types under discussion is different. However, the frontal appearance of these vehicles looks quite similar. Hence, leading to a higher misclassification recorded between these vehicle types.

5. CONCLUSIONS

The author has presented a novel method for vehicle detection, tracking and classification using both geometric and appearance attributes of the vehicles. The proposed system was tested on a video sequence of 3,400 frames recorded from a typical urban intersection. Background subtraction method and Kalma filter algorithm were employed in the detection and tracking stages of the system. Using the vehicle's blob area, the system segmented the tracked vehicle at a point where its structure is most clear. The geometric and appearance features of the segmented vehicles were fed to LDA classifier for appropriate categorisation. The results indicated a promising trend in the vision-based vehicle detection and classification. Although, the proposed system had demonstrated robustness in the vehicle detection and tracking, further work is still required to help resolve the occlusion issues raised (such as using stereo camera system). We intend to take up this challenge in our future work.

REFERENCES

- [1] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 5, pp. 694–711, 2006
- [2] A. N. Rajagopalan and R. Chellappa, "Vehicle detection and tracking in video," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, 2000, vol. 1, pp. 351–354.
- [3] R. I. R. García and D. Shu, "Vision based Vehicle Tracking using a high angle camera."
- [4] Y. Long, X. Xiao, X. Shu, and S. Chen, "Vehicle Tracking Method Using Background Subtraction and MeanShift Algorithm," in *E-Product E-Service and E-Entertainment (ICEEE), 2010 International Conference on*, 2010, pp. 1–4.
- [5] M. Djalalov, H. Nisar, Y. Salih, and A. S. Malik, "An algorithm for vehicle detection and tracking," in *Intelligent and Advanced Systems (ICIAS), 2010 International Conference on*, 2010, pp. 1–5.
- [6] B. Gloyer, H. K. Aghajan, K. Y. Siu, and T. Kailath, "Vehicle detection and tracking for freeway traffic monitoring," in *Signals, Systems and Computers, 1994. 1994 Conference Record of the Twenty-Eighth Asilomar Conference on*, 1994, vol. 2, pp. 970–974.

- [7] A. Behrad, A. Shahrokni, S. A. Motamedi, and K. Madani, "A robust vision-based moving target detection and tracking system," in the Proceeding of Image and Vision Computing Conference, 2001.
- [8] M. Nieto, L. Unzueta, J. Barandiaran, A. Cortés, O. Otaegui, and P. Sánchez, "Vehicle tracking and classification in challenging scenarios via slice sampling," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, pp. 1–17, 2011.
- [9] S. Messelodi, C. M. Modena, and M. Zanin, "A computer vision system for the detection and classification of vehicles at urban road intersections," *Pattern Analysis & Applications*, vol. 8, no. 1, pp. 17–31, 2005.
- [10] S. Gupte, O. Masoud, R. F. K. Martin, and N. P. Papanikolopoulos, "Detection and classification of vehicles," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 3, no. 1, pp. 37–47, 2002.
- [11] H. Rabiü, "vehicle detection tracking and colour-based classification in video," *International Journal of advanced and innovative Research (IJAIR)*, vol. 2, no.1, pp. 97-100, 2013
- [12] García, R. I. R., and Shu, D. "Vision based Vehicle Tracking using a high angle camera".
- [13] B. Coifman, D. Beymer, P. McLauchlan, and J. Malik, "A real-time computer vision system for vehicle tracking and traffic surveillance," *Transportation Research Part C: Emerging Technologies*, vol. 6, no. 4, pp. 271–288, 1998.
- [14] G. Evensen, "The ensemble Kalman filter: Theoretical formulation and practical implementation," *Ocean dynamics*, vol. 53, no. 4, pp. 343–367, 2003.
- [15] H. Rabiü. and M.I. Saripan, M.I. and S. Mashohor and M.H. Marhaban. "3D facial expression recognition using maximum relevance minimum redundancy geometrical features" *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, pp. 213–220, 2012.
- [16] Ojala, T. and Pietikainen, M. and Maenpaa, T.. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns" *Pattern Analysis and Machine Intelligence IEEE Transactions*, vol. 24, no.7, pp. 971–987, 2002.
- [17] Welling, M., "Fisher linear discriminant analysis" Department of Computer Science, University of Toronto, 2005.

AUTHOR

The author was born in Kano, Nigeria. He received his first degree (Bachelor of Engineering) from the Department of Electrical Engineering, Bayero University Kano, Nigeria in the year 1994. After acquiring five years industrial experience the author in the year 1999 start his Master of Engineering degree program (M.Eng) and graduated in the year 2002 from the same department. In the year 2007 he registered with the Department of Computer and Communication System Engineering, Universiti Putra Malaysia to study PhD degree and still in process. The author's field of specialisation include digital image processing, 3D image processing, pattern recognition, machine learning, face recognition, facial expression recognition, multi media security, Automatic traffic monitoring and control and image forensic.

