VISION-BASED APPROACH FOR URBAN VEHICLE DETECTION & CLASSIFICATION

Long Hoang Pham*
International University - Vietnam National University
Ho Chi Minh, Vietnam
longpham3105@gmail.com

Tin Trung Duong*
International University - Vietnam National University
Ho Chi Minh, Vietnam
tin.duongtrung@gmail.com

Ha Manh Tran
International University - Vietnam National University
Ho Chi Minh, Vietnam
tmha@hcmiu.edu.vn

Synh Viet-Uyen Ha
International University - Vietnam National University
Ho Chi Minh, Vietnam
hvusynh@hcmiu.edu.vn

Abstract - **Vehicles detection and classification are the most popular subjects in the computer vision researching field, and also are the most important parts in any traffic monitoring or surveillance system. Although there has been a considerable amount of ideas to accommodate this problem since the 90s, many problems are still unresolved due to the complexity of traffic systems and the variety of vehicles. This paper is a work-in-process that proposes a new approach to detect and classify vehicles based on the traffic system in Vietnam. The main goal of this method is to group vehicles into 2 main classes, which are 2-wheeled and 4-wheeled vehicles, based on low-level traffic parameters in urban areas.**

Keywords - Segmentation, object detection, contours detection, distance transform, vehicle classification.

I. INTRODUCTION

In recent years, there has been an increasing interest in using traffic monitoring systems, especially in developing countries such as Vietnam. The basic function of these systems is to detect and classify vehicles. Most of the existing technologies monitor traffics by using transportation tag reader and antennas, or inductive loop detector to obtain traffic information. The vehicles must also be integrated with many signal-supporting devices such as GPS, IR laser, magnetic loop detector, etc..., which are only compatible with cars. However, these systems create an enormous amount of data, not to mention the huge investment that cannot be afforded by many developing countries. Additionally, in the developing countries, especially Vietnam, the large amount of small vehicles (e.g. bicycle, motorcycle) as well as the disorder movement of these vehicles makes these systems inoperative. Under these circumstances, a new trend has been emerged, based on the usage of traffic video processing. The computer vision-based traffic monitoring system is becoming popular due to its flexibility in installation, maintenance and upgrade. This method can utilize the existing surveillance system on some big routes to reduce costs while still reserve the function of count and classifying vehicles.

Recently, there is noticeable work in the area of vehicles detection and classification such as [1, 2, 3]. The most recent study on vehicle segmentation is [4], in which a model-based approach is employed using 3-D box models. There are also many other studies on the model-based method [5, 6, 19], which compare and match the detected vehicles with a set of sample images and models. Another approach for this problem is using features detection [7]. By using this approach, [8] presented a quite successful solution to classify vehicles into three main classes, Sedan, Semi, and an additional class, Truck + SUV + Van. However, these methods only work on specific traffic systems, which are unique for each country. About two years ago, [9] provided a traffic monitoring system, which is capable of analyzing the vehicle flow on urban streets in Vietnam. However, this system has two disadvantages; (a) it can only work with the camera which is installed in the middle-above of the street; (b) the result from this classifying method cannot clearly distinguish bike and car.

The objective of this research is to achieve an improvement method in detection and classification vehicles into two main class, 2-wheeled and 4-wheeled vehicles, using data of traffic video. The research inherits one core idea from [8], which is using the low-level features from the video to detect the best observation zone. This model-based system can give us a good result, providing a set of models for every vehicle. However, in developing countries where routes are not properly divided into lanes for different means of transport (e.g. cars, bikes, motorbikes, pedestrians, and others), these models seem to be impossible to create. Moreover, the model-based system requires the vehicles to be in the right traffic lane, which is uncommon in these countries, especially Vietnam. Compared to [9], our system can find a better observation zone, which can remove most unclear objects, and can work with any camera position toward which the traffic flow heads. Furthermore, the system can utilize the low quality videos from the surveillance system of Vietnam.

Because the location of the surveillance camera is unknown until the system runs, we apply one simple technique but it can give a significant effect to our method for finding the best observation zone. We name it "best-zone tracking technique", and the idea can be described as follow. Firstly, we divide the input frame into particular zones. Further explanations of the zones are shown in section II-B. Secondly, we use some statistical method to get the best zone in which we can detect and classify the vehicles almost perfectly. After having the best observation zone, we apply the ratio estimation to classify vehicles. What makes our method stand out from the conventional ones is that it classifies vehicles using only the low-level features from the objects' ellipse. The evaluation function takes into account 2 criteria, which are dimensional ratio $R_{Dimension}$ and density ratio $R_{Density}$, dividing vehicles into 2 classes (2-wheeled and 4-wheeled vehicles). Further information of this function is show in section II-C.

The rest of the paper is organized as follow. Section II-A describes briefly the preparation step including the optical flow and contour detection. Then section II.B introduces the technique to detect the best observation zone automatically and section II.C provides the evaluation function for detecting vehicle. Experiments are discussed in section III, followed by the conclusion in section IV.

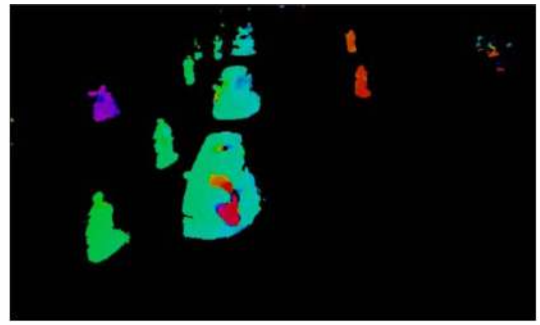
II. THE PROPOSED METHOD

A. *Vehicle Detection*

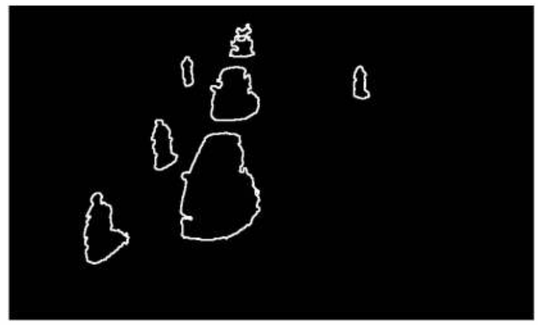
The main goal of the proposed system is to segment, count and classify moving vehicles in surveillance videos. Therefore, static objects are obsolete and need to be removed. For this study, the optical flow and background subtraction modules from [10, 11, 12, 13] were used to produce the traffic flow of the street from the input video. In this paper, the result of optical flow is represented by the color flow, which is the moving vehicles or objects being extracted from the background (Fig. 1 - c).

(a)

(b)



(c)



(d)

Fig. 1: Traffic flow example. (a), (b) Original images (sequence VVK1 – frame 350 and 355). (c) The traffic flow. (d) Contour extraction.

After obtaining the traffic flow, each object is extracted from the image by determining its contour. These contours are prepared by adapting the procedure used by [14]. More details about contour tracing algorithm discussion can be found in there.

The contour extraction method can also filter noises and obsolete objects to obtain a better foreground blobs. Then, we calculate the area of each contour and eliminate the object that has too small area by using the Green formula. This small step helps reducing the potential noises and eliminates obsolete objects. In Fig. 1, this process is fully described.

B. *Observation Zone*

Since the cameras are installed along the streets, the best observation zone is the one where the sizes of vehicles do not change significantly in several consecutive frames. Generally, the surveillance video can be divided into 3 distinct zones: the top, the middle and the bottom one. The vehicles in the top zone are further from the camera, which are smaller to the vehicles located in the bottom zone. One common characteristic of these 2 zones is that the sizes of objects change significantly after two or more consecutive frames. On the other hand, when the vehicles travel through the middle zone, their sizes slightly change. Since the vehicles classification method significantly depends on the low-level information of objects such as the shape size or the middle zone, the observation zone must be defined.

It is so difficult to define the observation zone because not all surveillance cameras are set up at the same heights and angles. In fact, most cameras are installed on the 4.5-meter traffic lights, while the rests are positioned on the pedestrian walking bridges on the street. In order to address this issue, we propose a method to automatically and accurately define the best observation zone. Let x be the location of a vehicle on a frame and $f(x)$ be its contour size. We will track the location and size of some vehicles when they travel through the camera range. Then we calculate the second derivative to obtain how fast the objects' size is changing. By using these values, we can determine the range of locations where the magnitudes of second derivative are between 0 and 1 ($0 \leq f''(x) \leq 1$).

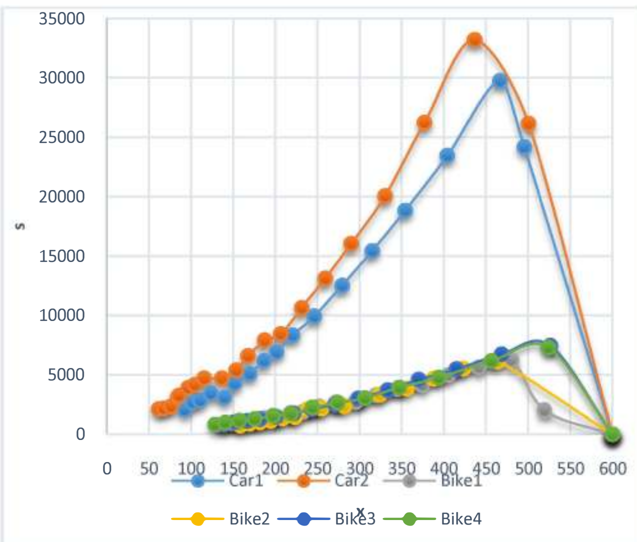


Fig. 2: Vehicle Size vs. Location

TABLE 1: The result of $f(x)$ and $f''(x)$ of Bike 1

x	$f(x)$	$f'(x)$	$\lvert f''(x)\rvert$	Description
158	732	15.40	0.6964	Top zone (Vehicle appear)
169	901	7.74	0.0621	
181	994	8.48	0.4677	
194	1104	14.56	0.4945	
208	1308	7.64	2.4488	
223	1423	44.37	2.6807	
238	2088	4.16	0.1799	Observation zone
255	2159	7.22	1.1465	
282	2354	38.18	1.5194	
298	**2965**	**13.87**	**0.0693**	**Best stable location**
324	3325	15.67	0.3394	Observation zone
356	3827	26.53	0.0920	
388	4676	23.59	0.2619	
423	5501	14.42	1.4697	Bottom Zone
463	6078	-44.37		
600	0			

Fig. 2 represents the function $f(x)$ of both car and motorbike from the sample video VVK1. Next, Fig. 3, 4 shows the magnitude of $f''(x)$. In TABLE 1, at $x = 298$, the value of $f''(x)$ is the smallest and almost equal to 0. Furthermore, the best observation for bike 1 is from 223 to 423 ($223 < x < 423$). From Fig. 3 and Fig. 4, it is concluded that the observation zone for the camera in sequence VVK1 is within the range of $200\, and\, 400$ ($200 < x < 400$), where $|f''(x)|$ is in the range of $[0,1]$.

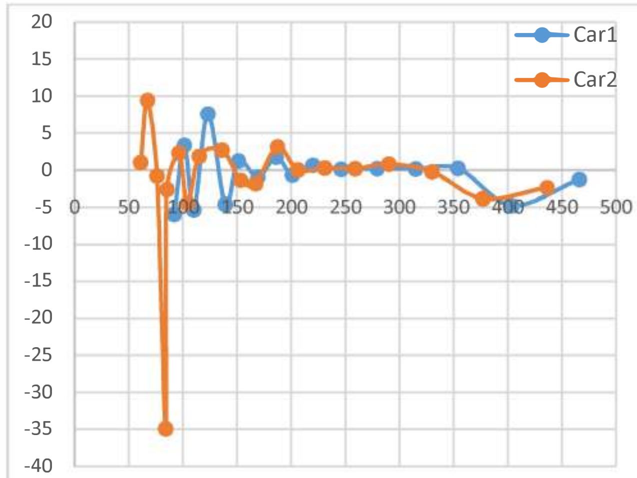


Fig. 3: Second Derivative for 4-wheeled

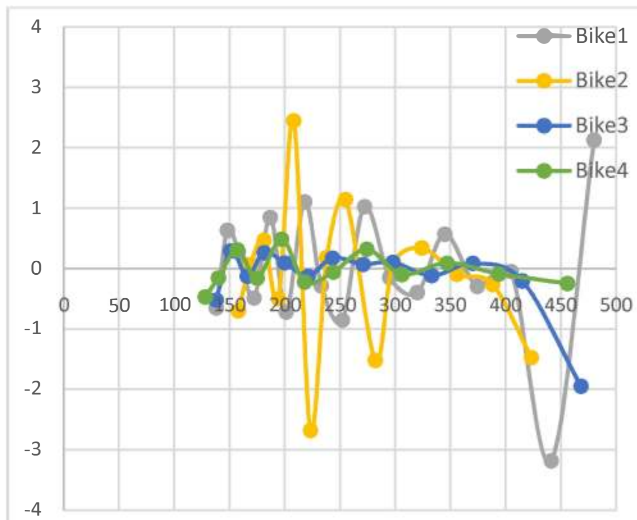
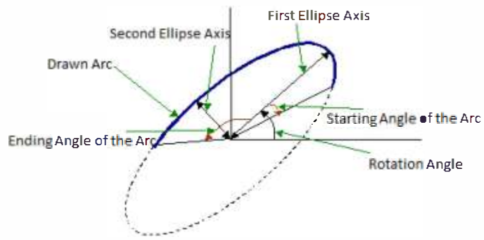


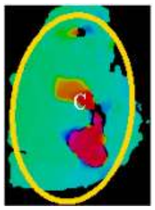
Fig. 4: Second Derivative for 2-wheeled

C. *Vehicles Classification*

After some preparation steps to identify the objects and extract their contours information from the input frame, the system performs a classification process. In order to extract useful low-level features, each vehicle is bounded by ellipse. [15] provides an algorithm to calculate the ellipse that fits best (in a least-squares sense) to a set of 2D points. The extracted ellipse-bounding vehicles can be visualized in Fig. 5-a:



(a)



(b) (c)

Fig. 5: (a) The extracted ellipse's features. (b) Extracted car. (c) Extracted motorbike (sequence VVK1 – frame 355)

What makes our method stand out from the conventional ones is that it classifies vehicles using only the low-level features from the objects' ellipse. The proposed method uses ratio estimation to classify vehicles. The evaluation function take into account 2 criteria, which are dimensional ratio $R_{Dimension}$ and density ratio $R_{Density}$, dividing vehicles into 2 classes (2-wheeled and 4-wheeled vehicles).

Let O be the object that need to be classify and E is the bounding ellipse of O. The dimension ratio is the fraction of the ellipse's width and height:

$$R_{Dimension} = \frac{E_W}{E_H}$$

Where E_W is the ellipse's width and E_H is the ellipse's height

Since the ellipse has a property that its width is always smaller than height, the dimensional ratio is in the range of $0 < R_{Dimension} < 1$. The dimensional ratio is used as the primary criterion to classify 2-wheeled and 4-wheeled vehicles.

As suggested by Fig 6-c, the ellipse bouding a 2-wheeled vehicle, for instance a motorbike, has a small $R_{Dimension}$ value due to the great difference between dimensions. The reason is that the motorbike and the rider are grouped together as one moving object which makes the bounding ellipse thinner. In most cases, the ellipse's height is 2 times larger than its width. In constrast, the value $R_{Dimension}$ of a car is large since its height and width are not so different.

The dimensional criterion alone can work well in most cases, but there are remaining problems due to the variety of motorbike. When 2-wheeled vehicles are far from the camera, their shapes can be similar similar to 4-wheeled ones. In order to separate vehicles efficiently, the second criterion – desity ration, must be used. The density ratio is calculated by dividing the number of the object's pixels (non-zero pixels) by the total pixels of the ellipse. According to the image above, cars are uniform rectangles that are fitted into the ellipses, while motorbikes are not. So the density ratio of 4-wheeled vehicles will be larger than 2-wheeled ratio.

$$R_{Density} = \frac{\sum_{p \in O} p}{\sum_{p \in E} p}$$

The Fig. 6 shows the distribution of 2-wheeled and 4-wheeled vehicles ratios taking from a set of samples.

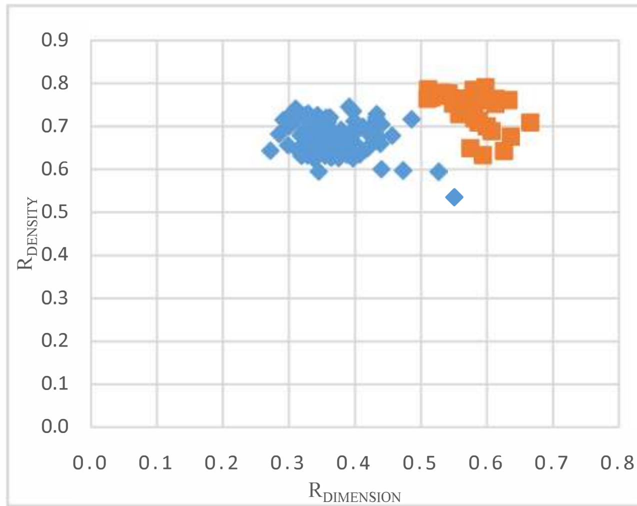


Fig. 6: The blue dots represent 4-wheeled vehicle. The orange dots represent 2-wheeled vehicles.

Along the dimensional axis, most of the 2-wheeled vehicles distribute in the range of $0.2 < R_{Dimension} < 0.5$ and the 4-wheeled vehicles are in the range of $0.6 < R_{Dimension} < 1$. The range $0.5 < R_{Dimension} < 0.6$ is the ambiguous zone, where the dimensional ratio alone cannot clearly distinguish 2-wheeled and 4-wheeled vehicles. Therefore, the density ratio could be used to improve the classification process. In the ambiguous zone, the value of 4-wheeled $R_{Density}$ is larger than that of 2-wheeled $R_{Density}$. The means and variances are as are as Table 2:

TABLE 2: Table of Mean and Standard Deviation

	$R_{Dimension}$		$R_{Density}$	
Vehicle's type	2-wheeled	4-wheeled	2-wheeled	4-wheeled
Mean	0.3591	0.6640	0.6126	0.7858
Variance	0.00252	0.00161	0.00295	0.00406
Standard Deviation	0.0502	0.0401	0.0544	0.0637

III. RESULTS AND DISCUSSION

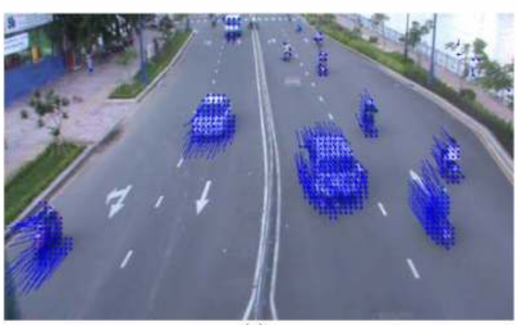
The experiment input data is selected specially for testing the proposed algorithm. Each testing video sequence includes a mix of 4-wheeled (cars) and 2-wheeled (motorbike) vehicles. Since the camera is installed on top of a bridge, the best observation areas are the one that is near the camera, and the bottom half. The final goal is to classify the vehicles at this observation area.

The 2-wheeled vehicles are marked with the letter ”M” (stand for motorbike) and bounded by green ellipses. The 4-wheeled vehicles are labelled by the letter ”C” (stand for car) and bounded by orange ellipses (Fig. 7). The system can clearly detect and classify well when the vehicles in the observation area.

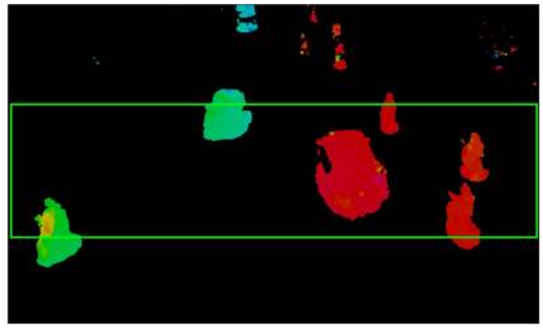
Sometimes, bad situations happen in the input sequence. In particular, two or more motorbikes driving so close to each other with the same speed will create a shape similar to the shape of a car, which causes incorrect results. In Table 3, we make some statistic calculations for a random set of frames from the sequence VVK1 to test the exactness of the proposed algorithm. The results show that most of the errors are caused by the occlusion of 2-wheeled vehicles. Since it is still a work-in-process, more improvement will be made to enhance the accuracy of the classification algorithm.

TABLE 3: Statistical calculation base on a random set of 1470 frames from sequence VVK1

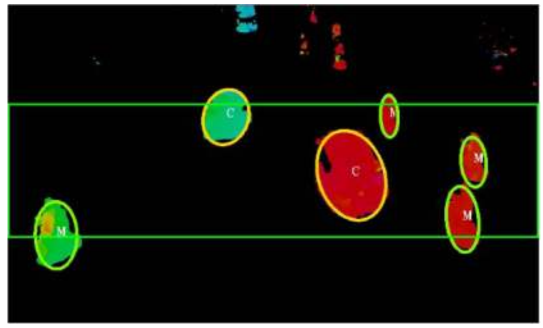
	Count Manually	**Count Automatically**	**Deviation**	**Percent**
Total Vehicles Detected	736	684	52	92.9347%
4-wheeled	72	64	8	88.8889%
2-wheeled	664	620	44	93.3735%



(a)



(b)



(c)

Fig. 7: Experiment result on frame 8855 and 8860 of sequence VVK1. (a) optical flow. (b) trafic flow. (c) Segmentation and Classification

IV. CONCLUDING REMARKS

In this paper, we presented a new method that can detect and classify vehicles in low quality monitoring videos. Our proposed method first finds motion vectors associated with the moving vehicles and then marks and extracts them. Then the system bounds each vehicle with an ellipse and obtains low-level features. Moreover, we also apply the “best-zone tracking technique” to help our method to get a better result. Although the idea of this one is simple, it helps the system a lot by removing all of the noise zones from the input scene. For example, the bottom zone can contain some cars that have gone out from the middle of the scene. Therefore the shape that we get is not a full-car shape, which can lead to an error in computing the result. Using these features, the classification function applies 2 ratio criteria to divide the vehicles into 2 main groups: 2-wheeled and 4-wheeled. The experiments on the suggested algorithm show some promising of results. Most the vehicles are detected and can be classified successfully. The future works will focus on solving the case when two or more vehicles overlapping each other to improve the accuracy of the proposed algorithm.

V. REFERENCES

[1] S. Nandhini, P. G. S. Parthiban, “Automatic Vehicle Detection during Nighttime Using Bright Pixel Segment with Spatial Temporal Technique”, IJECSE Volume 1, Number 3, June 25, 2012.

[2] L. Vasu and D. M Chandler, "Vehicle Tracking Using Human-Vision-Based Model of Visual Similarity", 2010 Southwest Symposium on Image Analysis and Interpretation, May 2010.

[3] X. Song and R. Nevatia, "Detection and tracking of moving vehicles in crowded scenes", IEEE Workshop on Motion and Video Computing, 2007.

[4] N. Buch, M. Cracknell, J. Orwell, and S. A. Velastin, "Vehicle localisation and classification in urban CCTV streams", United Kingdom, IT World Congress 2009.

[5] R. T. Collins, A. J. Lipton, and T. Kanade, "A system for video surveillance and monitoring". InProc. American Nuclear Society on the International Topical Meeting on Robotics and Remote Systems, Pittsburgh, PA, pages 1 – 15, April 1999.

[6] M. Nieto, L. Unzueta, J. Barandiaran, A. Corte1s, O. Otaegui and P. Sa1nchez, "Vehicle tracking and classification in challenging scenarios via slice sampling". EURASIP Journal on Advances in Signal Processing, 2011.

[7] J. Y. Ng, Y. H. Tay, "Image-based vehicle classification system", the 11th Asia-Pacfic ITS Forum & Exhibition, Kaohshiung, Taiwan, June 8-11, 2011.

[8] B. Morris and M. Trivedi, "Robust Classification and Tracking of Vehicles in Traffic Video Streams," in Proc. IEEE Conf. on Intell. Transport. Syst., pp. 1078-1083, Toronto, Canada, Sept. 2006.

[9] Nam Tang, Cuong Do, Tien Ba Dinh, and Thang Ba Dinh, "Urban traffic Monitoring system", ICIC 2, volume 6839 of Lecture Notes in Computer Science, page 573-580. Springer, 2011.

[10] S. V.-U. Ha and J. W. Jeon, "Readjusting unstable regions to improve the quality of high accuracy optical flow," IEEE Transaction on Circuits and Systems for Video Technology (CSVT), vol. 20, August 2010.

[11] Stefano Messelodi, Carla Maria Modena, Michele Zanin, "A computer vision system for the detection and classification of vehicles at urban road intersections", Pattern Analysis and Applications, Volume 8, Issue 1-2, pp. 17-31, September 2005.

[12] Brox T., Bruhn A., Papenberg N., Weickert J. – High Accuracy Optical flow Estimation Based on a Theory for Warping, European Conference on Computer Vision (ECCV), pp. 25–36, 2004.

[13] C. Stauffer and W.E.L. Grimson, Adaptive background mixture models for real-time tracking," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 246-252, 1999.

[14] S. Suzuki and K. Abe, "Topological structural analysis of digitized binary images by border following," COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING 30, pp. 32–46, 1985.

[15] W. Fitzgibbon and R. B. Fisher, "A buyer's guide to conic fitting", BMVC 1995 doi:10.5244/C.9.51.

[16] N. K. Kanhere, S. J. Pundlik, and S. T. Birchfield, "Vehicle segmentation and tracking from a low-angle off-axis camera," Computer Vision and Pattern Recognition, IEEE Conference on, vol. 2, pp. 1152–1157, 2005.

[17] G. J. J., K. Aggarwal, and M. Gokmen, "Tracking and segmentation of highway vehicles in cluttered and crowded scenes," IEEE 2008 Workshops on Applications of Computer Vision Copper, Jan. 2008.

[18] M. Nieto, L. Unzueta, J. Barandiaran, A. Corte1s, O. Otaegui and P. Sa1nchez, "Vehicle tracking and classification in challenging scenarios via slice sampling". EURASIP Journal on Advances in Signal Processing, 2011: 95.

[19] Amol Ambardekar, Mircea Nicolescu, and George Bebis, "Efficient Vehicle Tracking and Classification for an Automated Traffic Surveillance System," the Proceedings of Signal and Image Processing, Kailua-Kona, Hawaii, August 2008.

[20] Atkočiūnas, R. Blake, A. Juozapavičius, M. Kazimianec, "Image Processing in Road Traffic Analysis", Nonlinear Analysis: Modelling and Control, Vol. 10, No. 4, 315–332, 2005.