Occlusion Vehicle Segmentation Algorithm in Crowded Scene for Traffic Surveillance System

Hung Ngoc Phan, Long Hoang Pham, Duong Nguyen-Ngoc Tran, and Synh Viet-Uyen Ha[✉]

School of Computer Science and Engineering, International University, Vietnam National University HCMC, Block 6, Linh Trung, Thu Duc District, Ho Chi Minh City, Vietnam
hvusynh@hcmiu.edu.vn

Abstract. Traffic surveillance system (TSS) is an essential tool to extract necessary information (count, type, speed, etc.) from cameras for further analysis. In this issue, vehicle detection is considered one of the most important studies as it is a vital process from which modules such as vehicle tracking and classification can be built upon. However, detecting moving vehicles in urban areas is difficult because the inter-vehicle space is significantly reduced, which increases the occlusion among vehicles. This issue is more challenging in developing countries where the roads are crowded with 2-wheeled motorbikes in rush hours. This paper proposes a method to improve the occlusion vehicle detection from static surveillance cameras. The main contribution is an overlapping vehicle segmentation algorithm in which undefined blobs of occluded vehicles are examined to extract the vehicles individually based on the geometry and the ellipticity characteristics of objects' shapes. Experiments on real-world data have shown promising results with a detection rate of 84.10% in daytime scenes.

Keywords: Occlusion detection · Blob splitting · Vehicle segmentation · Traffic surveillance system

1 Introduction

The past decade has seen the explosion of intelligent and expert systems, especially in the area of transportation management. TSS has gained popularity among researchers and authorities. A key aspect of TSS is to derive the traffic information (count, average speed, and the density of each vehicle type) for further analysis related to traffic management and planning. In this context, many studies have been conducted in developed countries where the transportation frameworks are constructed primarily for automobiles. These systems [1,2] were

developed with the advanced equipment and sensors to optimize the incoming signal including radar, infrared camera. However, in developing countries, the application of these systems has trouble with high-cost and incompatible infrastructures. On the contrary, the vision-based TSSs which are built from computer vision and image processing techniques [3–5] have shown more superior capability with lower cost and are easier to maintain. Moreover, they are extremely versatile as algorithms can be designed to cope with a broad range of operations such as detecting, identifying, counting, tracking, and classifying vehicles.

In vision-based TSS, vehicle detection is one of the most critical operations since it plays a major role in localizing moving vehicles in traffic videos. This vital tool is a well-defined mechanism that has been studied in a considerable amount of works. Some recent works by Ha and Pham et al. [5,6] have shown remarkable outcomes in detecting and classifying vehicles in urban areas. However, urban vehicle detection has its challenges that cannot be easily addressed by these methods due to the occurrence of vehicle occlusions. The problem is more challenging in rush hour when the traffic gets slower, and the inter-vehicle space is significantly reduced, which increases the occlusion among vehicles. In addition to this, the immense density of motorbikes is the main cause for chaos on urban roads in Vietnam. Furthermore, the non-rigid shapes of this kind of vehicles vary widely when they move throughout the scene. As a consequence, the traffic becomes more complicated and awkward for vehicle detection.

In this paper, we propose a robust vehicle detection algorithm that handles the occlusions of 2-wheeled motorized vehicles in crowded traffic scenes. Our work is an extension of [5,6] which presented a robust TSS comprising of three main components: background subtraction, vehicle detection, and classification. The main contribution of our method is introducing an overlapping vehicle segmentation algorithm which has been developed with a data-driven approach on real-world data. Like previous studies, we use background subtraction to model the background, from which moving vehicles can be detected. Then, blobs of overlapping vehicles are identified based on the geometric characteristics and the spatial features of objects' shapes. This assessment is done by using a decision tree constructed from a large training set of 10,000 vehicle images captured in Ho Chi Minh City, Vietnam. Once occluded vehicles are extracted, we proceed with the overlapping vehicles segmentation process. We propose a novel segmentation method that performs exhaustive checking and pairing of defect points in the object contours. The blobs resulted from each cut are validated with the vehicle model [5] consisting of vehicle size P^C, dimension ratio R^{di}, density ratio R^{de}, and ellipticity characteristics. Experiments have shown promising results with high vehicle detection rates of 84% on the considered data.

The rest of this paper is structured as follows. Firstly, Sect. 2 provides an extensive scrutiny of background subtraction model and describes vehicle detection module which we exploit in our proposed method. This will be followed by Sect. 3 where the main contribution in this research, overlapping vehicle

segmentation, will be presented. Finally, experiments and discussion are stated in Sect. 4 to evaluate, to summarize our work, and to conclude the paper.

2   Vehicle Detection

In video-based approach, the aim of TSS is to process traffic videos which are captured from static pole-mounted cameras. In order to achieve this goal, the input data goes through several intermediate procedures. This section discusses background subtraction model and vehicle detection module.

At the first phase, it is typical to construct a background as well as to extract the moving objects from input videos. However, the background image is not always available as a certain frame in received data. More precisely, background subtraction is an indispensable procedure of modeling static scenery from which we separate the moving objects for further processes. The outcome of this operation which is utilized to evaluate the accuracy and the effectiveness of the model comprises two masks: background and foreground image. While the background is marked as black areas, the fields on the foreground are white blobs which imply the objects that are on the move. Particularly, in the empirical environment, the background image may be affected by external factors including illumination changes, camera vibration, and noise caused by sluggish or motionless vehicles. In order to overcome these issues, we make use of the background subtraction model which is proposed by Nguyen et al. [7] to create a stable background in many circumstances. In addition to this, by narrowing down the examining area, we eliminate unexpected objects and reduce the number of blobs that we need to investigate. Figure 1a–c illustrates the background construction and the background subtraction procedure.
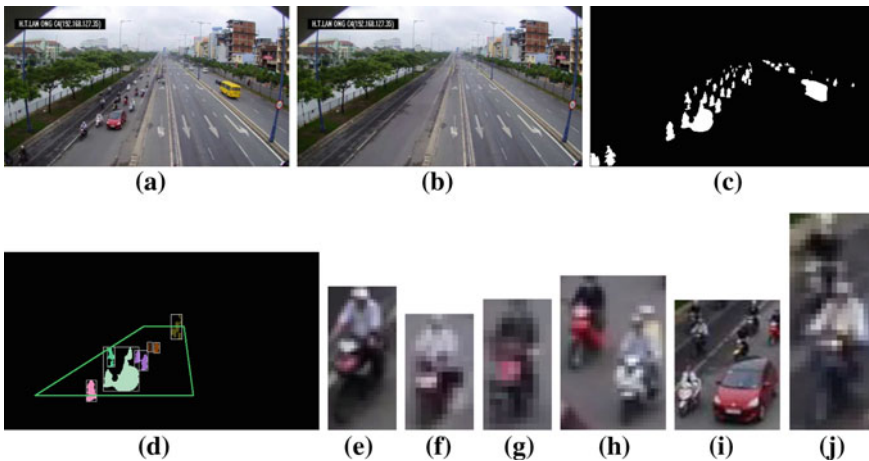

Fig. 1. a Original image. **b** Background image. **c** Foreground image from background subtraction process. **d** Extracted blobs of moving objects inside the examining area. **e–j** Extracted blob images

After moving objects are detached from traffic scene, in order to derive useful features from conveyances, each blob of vehicle has to be characterized individually. Regarding this issue, we adopt the construction of vehicle detection which was proposed by Ha and Pham et al. [5,6,8]. Using this approach, they take advantage of ellipse fitting model, which was discussed by Fitzgibbon and Fisher [9], to bound the isolated blobs of vehicles. As an illustration, Fig. 1d exemplifies the result of this manner. Apparently, their solution has a superiority of lower computation time when exploiting six ellipticity properties of detected blobs. The characteristics of ith candidate at tth frame are outlined and illustrated in Table 1 and Fig. 2.

Table 1. Vehicle's measurement

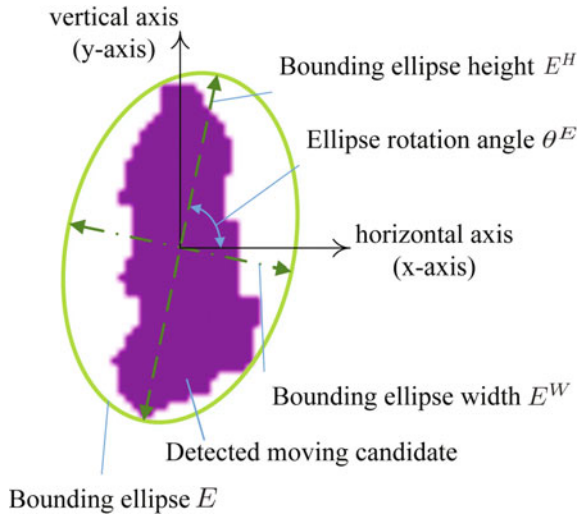| Symbol	Description
$E_i(t)$	Vehicle's bounding ellipse
$\theta_i^E(t)$	Bounding ellipse rotation angle
$E_i^W(t)$	Bounding ellipse width
$E_i^H(t)$	Bounding ellipse height
$P_i^E(t)$	Total pixel of bounding ellipse
$P_i^C(t)$	Vehicle size

Fig. 2. Ellipticity properties

3   Overlapping Vehicle Segmentation

3.1   Occlusion Detection

In the current investigation, from the identified set of moving objects, the occlusion blobs continue to be sorted out to prepare for the further segmentation process. In this research, we categorize the detected blobs of vehicles from Sect. 2 into four classes: light (bike and motorbike), medium (car, sedan, and 12-seater bus), heavy (truck, trailer, 16-to-50-seater bus), and occluded (blob of overlapping vehicles).

Following the essence of vehicle classification, we construct an evaluation procedure generating a tuple of three informative geometric features that depict the four kinds of vehicle blobs. The first assessment is vehicle size P^C, which is the total number of pixels bounded by vehicle's contour. The second measurement is density ratio R^{de}, which is achieved by calculating the proportion of vehicle size to the bounding ellipse size P^E. The last appraisal is dimension ratio R^{di}, that is, the ratio of bounding ellipse width E^W to height E^H:

$$R_i^{di}(t) = \frac{E_i^W(t)}{E_i^H(t)}; R_i^{de}(t) = \frac{P_i^C(t)}{P_i^E(t)} \tag{1}$$

From the above literature, several empirical experiments have been executed. More precisely, the blobs of vehicles which are the outcomes from vehicle detection in Sect. 2 are extracted the necessary properties and labeled for the analysis process. Figure 3a shows a scatter plot which illustrates the distribution of vehicles' features in 3D space containing three axes that correspond to three defined attributes (P^C, R^{di}, R^{de}). In this figure, the green, yellow, blue, and red dots, respectively, describe light, medium, heavy, and occluded vehicles. The scatter diagram intelligibly confirms an observation that there is a substantial separation in dimension ratio among these groups. Actually, at FOV of surveillance cameras, the length of medium and heavy vehicles which is considered as the height of detected blobs seems to be shorter than the one in reality. On the other hand, there is a disparity between the vertical and horizontal linear measurement when considering motorized two-wheelers. Moreover, the height of candidates in this class is also affected by the presence of motorists. Therefore, R^{di} of automobiles is greater than the value of motorbikes. Nevertheless, the dimensions of blobs of obscured vehicles notably fluctuate because of the dependence on the state of occlusion, which is the primary issue causing the confusion of distinguishing with the other classes. In order to resolve this problem, density ratio R^{de} is a noteworthy investigation. Obviously, the density ratio of occlusion blob is the least when compared with three other categories of conveyances. The cause of this fact is the portion of inner spaces among occluded candidates that form the gaps interleaving in blobs detected. However, another essential aspect of handling the borderline candidates is still indecisive. By inspecting the vehicle size P^C, four groups are obviously separated by surfaces which are parallel to the plane (R^{di}, R^{de}). In other words, considering the size of blobs, there is a manifest difference between any two classes.
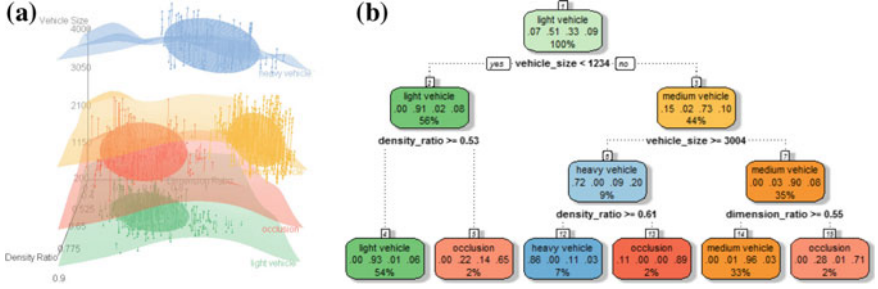
Fig. 3. a The scatter plot of vehicles from dataset VVK1. **b** The decision tree categorizing the vehicles and occlusion part in dataset VVK1

With this inspection of informative features in separating four classes of blobs detected, a set of tuples (P^C, R^{di}, R^{de}) which are extracted from an amount of data is utilized to construct a categorizing structure. In this research, the decision tree approach that is presented by Thomas [10] is adopted to build up a classification model. In the literature, the decision tree is a method classifying a batch of discrete samples. More precisely, the input data is distributed on an architectural tree with a group of nodes consisting of a root node, some internal nodes, and several leaf nodes. This structure can be obtained by using a top-down, greedy search algorithm, especially ID3, which focuses on finding out the best classifiers among the candidate's attributes to disassemble a large collection into smaller identified groups via a statistical assessment. Figure 3b shows the decision tree that corresponds to our training dataset. Starting at the root node, on the left branch, the light vehicles and the occlusion blobs share the same size criteria, but they distinguish each other with the specification of the density ratio. On the contrary, at the right branch, the remaining candidates are distributed with a particular vehicle size evaluation. Continuously, the smaller subset finally completes the compartmentalization through appropriate constraints on the density and the dimension ratio. From these observations, the decision tree can be utilized as a mechanism resolving both occlusion detection and vehicle classification.

3.2 Vehicle Segmentation

Once the occlusion blobs of vehicles are categorized as an exceptional instance of experimental subjects, at this stage, we continue to examine the bundle segmentation of detected candidates into the individual vehicles. Dealing with this issue, we present a robust solution to handle the overlapping vehicle segmentation.

As stated earlier in the introduction of this research, our proposed method is a video-based manner that processes directly on a sequence of captured images from surveillance cameras. Undeniably, pictorial signals received through the system can be affected by a variety of external factors including camera vibration, illumination change, and interference from other devices. For these reasons,

before initiating our principal procedure, the received data which is the set of occlusion blobs designated from Sect. 3.1 has to undergo a preliminary convention. As the foreground blobs are presented in binary image, in order to get rid of jagged edges and to refine the bounding contours of objects, morphology operators are appropriate solutions in this circumstance. After undertaking these pre-processing operations, the target subjects are steady and stable enough for the primary section that is overlapping vehicle segmentation.

In our approach, the occlusion segmentation is a repetitive process which partitions the obscured objects alternately. In normal conditions, regardless of small defects, the appearances of individual vehicles are considered as curving outward shapes. Therefore, in case of overlapping among vehicles, there are concave spots on the edge of detected blob, which is a corollary of the overlapping phenomena among moving conveyances. These positions are effective indications for blob splitting. Hence, in the first step of this method, we construct a convex bounding polygon ρ for the border of each detected blob μ_k that comprising a set of k vertices p_i, which is mathematically indicated as:

$$\rho = \left\{ \sum_{i=1}^{k} \lambda_i p_i | p_i \in \mu_k \wedge \lambda_i \geq 0 \wedge \sum_{i=1}^{k} \lambda_i = 1 \right\} \tag{2}$$

In this matter, an optimal algorithm is thoroughly presented by Sklansky [11]. Step (a) at each iteration in Fig. 4 illustrates the convex bounding polygon of occlusion blob. From that outcome, we continue to determine potential points for later process. In particular, this set σ_m consists of m concave spots which are on the outer boundary of examining objects but do not belong to the polygon ρ. Step (b) shows two selected concave spots that are utilized to segment the occlusion blob at the initial stage. In this figure, two attributes are presented to describe the importance of a detected locality. The first characteristic, denoted by a red line segment, is defect width that is the length of the bounding convex polygon's edge of inspecting spot. The other property, indicated as a green line,
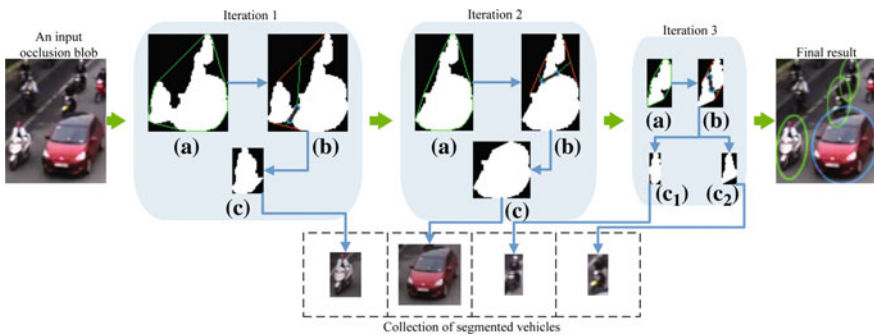

Fig. 4. The overlapping vehicle segmentation. **a** Convex bounding polygon computation. **b** Concave spot localization. **c** Segmented individual vehicles

is defect depth which is the distance from the concave spot to the midpoint of the corresponding bounding convex polygon's edge.

Furthermore, in this investigation, to select the correct segment points and to eliminate unnecessary impurities, two constraints attained from practical experiments are considered on detected set σ_m:

(1) The defect width must be greater than certain threshold.

$$\|\rho_i - \rho_j\| \geq TH_1 \text{ where } \rho_i, \rho_j \in \rho \tag{3}$$

where TH_1 is the minimum defect width and set to 7 pixels.

(2) The defect depth is limited with an interval bounded by two thresholds.

$$TH_2 \leq \left\| \sigma_t - \left(\frac{\rho_i + \rho_j}{2} \right) \right\| \leq TH_3 \text{ where } \rho_i, \rho_j \in \rho \text{ and } \sigma_t \in \sigma_m \tag{4}$$

where TH_2 and TH_3 are, respectively, the minimum and the maximum defect depths that are alternately set to 3 pixels and 30 pixels.

Afterward, for each pair of defined points, we form a cutting line which is utilized to separate obscured vehicles as step (c) at each iteration in Fig. 4. Subsequently, the segmented candidate after reconstructing the necessary vehicle's measurement is verified through the decision tree mentioned in Sect. 3.1. This manner takes 2–17 iterations of processes to attain the convergence of results. The procedure ends at the stage when all individual vehicles are identified. The rightmost image in Fig. 4 shows the final result of segmentation procedure.

4 Experiments and Discussion

Experiments have been performed on the selected data to evaluate the proposed method. In these examinations, the testing datasets are captured from static pole-mounted surveillance cameras in Ho Chi Minh City at the rate of 30 fps with a resolution of 640×480 to assess the accuracy objectively and to test the performance of our method during crowded scenes. Technically, these cameras are set up at the height of 8–9 m and inclined at an angle of $12°$–$15°$. In addition to this, the testing system has a configuration consisting of Intel Core i7 2630QM and 8 GB of RAM.

In the previous studies, Ha and Pham et al. [5] demonstrated a robust classification algorithm in daytime surveillance environment with a remarkable accomplishment. In this paper, we continue to improve prior attainments by initiating a novel method to proceed the occlusion in vehicle detection. Table 2 summarizes the results of the previous studies and our proposed method. Both solutions are examined on three different datasets HMD01, NTL01, and COL01 which depict the different levels of occlusion. Dataset HMD01 was captured on the highway during rush hour in the morning. In this dataset, because there is lane separation between motorbike and automobile, two approaches mainly concentrate on detecting and classifying overlapping vehicles of the same group. Dataset NTL01, which describes the high density of vehicles in the residential area, is

Table 2. Comparison of results between Ha's method and our proposed method

Dataset	Class	Actual	Ha's method [5]				Our method			
			Detect	Deviation	Percent (%)	T.Acc (%)	Detect	Deviation	Percent (%)	T.Acc (%)
HMD01	1	498	250	248	50.20	54.06	427	71	**85.74**	**84.60**
	2	68	47	21	69.12		56	12	**82.35**	
	3	7	3	4	42.86		6	1	**85.71**	
NTL01	1	652	364	288	55.83	62.54	538	114	**82.52**	**84.97**
	2	168	115	26	68.45		144	12	**85.71**	
	3	30	19	2	63.33		26	4	**86.67**	
COL01	1	574	392	182	68.30	66.97	483	91	**84.15**	**82.70**
	2	32	21	11	65.63		26	6	**81.25**	
	3	0	0	0	–		0	0	–	
Overall Accuracy			61.19				84.09			

also tested to assess a complicated circumstance with mixed-flow lanes. Dataset COL01 plays a major role in scrutinizing the higher level of occlusion among the light and the medium conveyances. Figure 5 shows some examples of three experiments.

Fig. 5. Examples of occlusion vehicle detection and classification result with three experimental datasets. Green, yellow, blue ellipses, respectively, indicate light, medium, and heavy vehicles. On the 1st column: results from dataset HMD01. On the 2nd column: results from dataset NTL01. On the 3rd column: results from dataset COL01

As shown in Table 2, the results obtained from the analysis of two methods on different datasets show that our algorithm improves the overlapping vehicle detection and accurately classifies the segmented candidates into three classes with confidence up to 85%. When compared with Ha's method whose studies are most relevant to ours, in intricate circumstances where there is a considerable presence of obscured vehicle, our overlapping vehicle segmentation algorithm significantly enhances the capability of TSS by 20% and increases adaptation in complicated situations. The proposed method can detect at least 82% of moving overlapping motorbikes and over 81% of medium vehicles. In particular, this result primarily depends greatly on the verification procedure as we mentioned in Sect. 3.1. Moreover, our approach of overlapping vehicle segmentation relies much on the geometric characteristics and the ellipticity attributes of detected candidates on the whole. Hence, in some minor cases, unexpected objects are erroneously gathered as desired vehicles such as pedestrians, rudimentary means of transportation, and slabs of foreground faults caused by sudden illumination change in background subtraction model. In general, our method not only provides the result with high accuracy, but also retains the low deviation when classifying light, medium, and heavy vehicles. This result is only achieved by the contributing an effective vehicle detection, a steady verification model, and a robust segmentation algorithm.

Besides outperforming on the outcome with an overall accuracy of roughly 84.1%, the proposed method maintains a small amount of computational time on the whole solution. To be more specific, the tests were performed with a

processing of approximately 9,000 frames in each execution. With the high density of occlusion which requires such more computations, the process rate is around 26 fps because the system needs to run extraordinary computation. On the contrary, with a lower degree of occlusion as in dataset HMD01, the frame rate can reach up to over 27 fps. Accordingly, our method can process traffic data in real-time and possibly integrates into the existing TSS.

5   Conclusion

This paper proposes a new method for overlapping vehicle segmentation to handle vehicle detection in circumstances of occlusion. The contribution of our research is a vision-based approach which utilizes the typical geometric features and the ellipticity characteristics to localize the conveyances inside the occlusion blob individually and to classify vehicles into 3 classes: light, medium, and heavy vehicles. In this investigation, the experiments on the suggested algorithm show some promising of results with the average accuracy over 84% and robust adaptability to the real-time performance at the overall frame rate of 27 fps.

References

1. R. Mobus and U. Kolbe: Multi-target multi-object tracking, sensor fusion of radar and infrared. IEEE Intelligent Vehicles Symposium, pp. 732–737 (2004)
2. U. Meis, W. Ritter and H. Neumann: Detection and classification of obstacles in night vision traffic scenes based on infrared imagery. Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems, 2, 1140–1144 (2003)
3. Duong Nguyen-Ngoc Tran, Tien Phuoc Nguyen, Tai Nhu Do, and Synh Viet-Uyen Ha: Subsequent Processing of Background Modeling for Traffic Surveillance System. International Journal of Computer Theory and Engineering, vol. 8, no. 3, pp. 235–239 (2016)
4. Synh Ha, Huy Hung Nguyen, Tu Kha Huynh and Phong Ho-Thanh: Lane detection in intelligent traffic system using probabilistic model. 16th Asia Pacific Ind. Eng. Manag. Syst. Conf. (APIEMS 2015), 1856–1863 (2015)
5. Ha, S.V.-U., Pham, L.H., Phan, H.N., Ho-Thanh, P.: A robust algorithm for vehicle detection and classification in intelligent traffic system. 16th Asia Pacific Ind. Eng. Manag. Syst. Conf. (APIEMS 2015), 1832–1838 (2015)
6. Ha, S.V.-U., Pham, L.H., Tran, H.M., Ho-Thanh, P.: Improved vehicles detection & classification algorithm for traffic surveillance system. J. Inf. Assur. Secur., 9, 268–277 (2014)
7. Nguyen, T.P., Tran, D.N.-N., Huynh, T.K., Ha, S.V.-U.: Disorder detection approach to background modeling in traffic surveillance system. J. Sci. Technol. Vietnamese Acad. Sci. Technol., 52, 140–149 (2014)

8. L. H. Pham, T. T. Duong, H. M. Tran and S. V. U. Ha: Vision-based approach for urban vehicle detection & classification. 2013 Third World Congress on Information and Communication Technologies (WICT 2013), pp. 305–310, Hanoi (2013)
9. Andrew W. Fitzgibbon and Robert B. Fisher: A buyer's guide to conic fitting. In Proceedings of the 6th British conference on Machine vision (BMVC '95), David Pycock (Ed.). BMVA Press, Surrey, vol. 2, pp. 513–522, UK (1995)
10. Thomas M. Mitchell: Decision Tree Learning. In: Machine Learning (1 ed.), pp. 52–80. McGraw-Hill, Inc., New York, NY, USA (1997)
11. Jack Sklansky: Finding the convex hull of a simple polygon. Pattern Recognition Letters, vol. 1, pp. 79-83, ISSN 0167–8655 (1982)