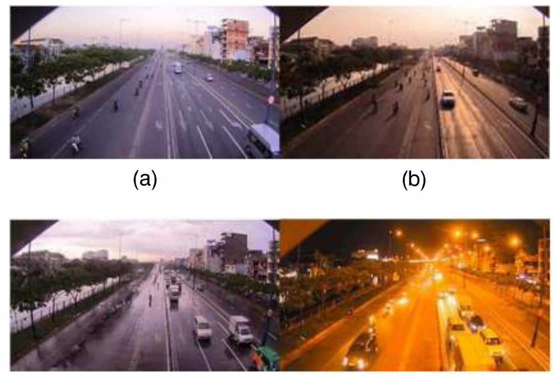
Scene Recognition in Traffic Surveillance System using Neural Network and Probabilistic Model

Duong Nguyen-Ngoc Tran*, Long Hoang Pham*, Ha Manh Tran, and Synh Viet-Uyen Ha†, *Member, IEEE*

Abstract— **In the traffic surveillance system (TSS), there are many factors affect the qualities of the result. Through practical application, it is difficult to determine which scene changing during the day period, from the daylight to nighttime, the conversion of the sunny and overcast, wet and dry scene. However, there have been no controlled studies which illustrate the method to distinguish environment scene, which is the one of six main challenges in TSS. Therefore, this paper presents the method to detect and recognize the change of scene during all-day surveillance; Thus, TSS adopt the recognition to determine the appropriate method for each scene, for increasing performance. Our recognition model is based on the combination of the CIE-Lab color space and the histogram of the region-of-interest (ROI) in each frame, which used for extracting the feature for the Feed Forward Neural Network to perform the detection. In the experiment section, our results show that the benefits of our proposed method in the real-world traffic surveillance system.**

Keywords— **scene recognition, traffic surveillance system, probabilistic model, artificial neural network.**



(a) (b)

(c) (d)

Fig. 1. Illustration of four typical outdoor scenes captured from a surveillance cameras. (a) Overcast. (b) Clear sky. (c) Rain. (d) Nighttime.

I. INTRODUCTION

Traffic surveillance system (TSS) has become a crucial tool for many intelligent traffic systems. As progress have been made in the field of computer vision and image processing [1], TSSs have been targeting to obtain an understanding of the traffic flow through extracting information (counts, speed, vehicle type, and density) [2], [3]. The camera used as the sensors takes a large variety of advantages; However, it faces a couple of challenges as well. The challenges include incompatible outdoor scenes because of the changing environment conditions. To consider these challenges, specialized techniques can be proposed to deal with specific conditions. Therefore, a vision-based system should be capable of detecting the current outdoor condition of the surveillance video or stream and determine the best method for the situation. The knowledge obtained by these systems can provide decision supports regarding traffic managements and urban planning, and enhances the reliability of the system.

Over the past decade, a significant amount of works has been done to make the TSSs cope with a wide range of outdoor environments. Each outdoor condition demands variety of handling process. Therefore, vision-based traffic surveillance systems require the ability to observe the current outdoor environment, for which the daytime (clear sky, overcast, and rain) modules or nighttime modules should be performed. At the moment, there are several works done for recognition the outdoor conditions. For the nighttime surveillance videos, the headlight is the important features. In [4], they use separated low-level modules to determine the daytime and nighttime conditions. The vehicles at night are tracked by executing headlight paring using template matching the image frames with the codebook, which was resized to the respect of the image regions. More headlight pairing method can be found in more recent research [5]–[8]. Additionally, besides suing pair headlights to determine if daytime or nighttime, the work [9] examine the average intensity and the red, green, and blue (RGB) color components in the surveillance frame. The work observes that the average intensity has a tendency of being lower and the deviation of each RGB color component tend to be small at night. Using two thresholds for histogram intensity and variation of the color component, they choose the proper procedures for TSS. One such scenario is the detection and classification of vehicles in daytime [10]; they only work on the overcast condition to detect and tracking the vehicle. In the overcast condition, the vehicle has no shadow go along with; Therefore, the object extracted by segmentation has the exact size and using the street observation zone, they extract the vehicle types and its speed. In the presence of shadows created by different lighting conditions [11], the outdoor environment in the condition is a clear sky. By shadow detection, the process classifies the moving pixels as shadow points based on their appearance with respect to the static frame. The challenge in the shadow detection is the misclassified the moving object (vehicle) as the moving shadow. They introduce the Sakbot aims to solve the under-segmentation problem. The method analyzes pixels in the Hue-Saturation-Value (HSV) color space, for which they can separate the chromaticity and

D. N.-N. Tran, L. H. Pham, H. M. Tran, and H. N. Phan are with the School of Computer Science and Engineering, International University, Vietnam National University, Ho Chi Minh City, Vietnam.

978-1-5386-3422-6/17/$31.00 ©2017 IEEE

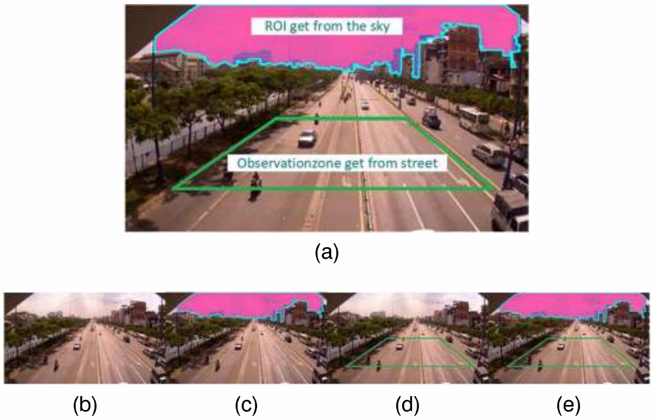


Fig. 2. Illustration of the dual sampling regions approach. (a) The detail of the dual sampling regions. The green part represents the observation zone. The pink area represents the sky region. (b) The original frame. (c) The watershed segmentation for getting the sky region. (d) The manual setting observation zone. (e) The combination of the sky region and the observation zone.

luminosity explicitly. Others proposed methods to remove vehicles's reflections on the wet road in rainy conditions. The main point of the process is to detect and remove the reflected areas in the input frame. In the rainy conditions, the light intensity is low, the difference between shadow pixels (reflected pixel) and normal ones is low, which leads to failing to distinguish the shadow region and vehicle. In [12], they propose a combined technique to remove both vague and hard shadows in a single image. However, the resulting image still contains halos and the areas around the shadow edge would have unnatural artifacts. To improve the remove reflection, in [13], the method uses the combination of Lab and HSV color spaces. They focus on L and H channels by combining the values, a certain threshold extracted from the practical experiment would determine reflection or object pixel. In addition, one challenge in the rainy condition is that the headlight reflection occurs in low lighting conditions while the vehicle runs on the wet road. They used the pair headlight while checking whether one is above the other with a reasonable amount of displacement. As can be seen, these studies have proven to work well in their specific defined scenarios. However, in real-world surveillance, the outdoor scene can change drastically from overcast to sunny or rainy, or from daytime to nighttime. There have been little studies on a mechanism to recognize the gradually changing environment, from which algorithms can be interchanged and applied to appropriate scenes throughout all-day surveillance. This problem hinders the autonomous capability of TSSs and reduces the performance of vehicle detection and classification.

In this paper, we present a scene recognition algorithm that provides TSSs with the ability to adapt different algorithms to all-day surveillance. The proposed method detects and classifies outdoor surveillance scenes into four types (Fig. 1): overcast (OC), clear sky (CS), rain (RA), and nighttime (NT). To accurately determine the scene, we create a scene feature model consisting of the histogram of RBG color space and the probability model of CIE-Lab color space. We then use a feedforward neural network (FFNN) to train a dataset of several all-day surveillance videos. The trained FFNN has been validated with the real-time video stream from the surveillance cameras in Ho Chi Minh City, Vietnam. Early experiments have shown promising results with the accuracy of 87.38%, 90.85%, 86.7%, and 96.81% for overcast, clear sky, rain, and nighttime respectively.

The remainder of this paper is organized as follows. In Section II, we present the detail of the proposed method. The benchmark results to corroborate the advantages of the new method is described in Section III. For more detail, we show the qualitative performance in Section III-A to prove to the efficiency of the method and the time-consuming in Section III-B to make a point that the TSS would run in real-time include the proposed method. Finally, Section IV summarizes the proposed method and concludes this paper.

II. The Proposed Method

A. Dual Sampling Region

In this section, we introduce our approach and its implementation. From the dataset, two observations can be made: the camera field-of-view (FOV) contains both the sky part and the road:

- Sky region: the color intensity of the sky is distinguishable between daytime and nighttime.
- Observation zone: the road surface has different luminosity between overcast, clear sky (cast shadows), and rainy weather (reflection).

For the dual sampling regions, as shown in Fig. 2, consisting of an observation zone and a sky region is applied. The observation zone can be obtained as suggested by [10]. The observation zone is the area on the input frame that after consecutive images the size of vehicles does not change sharply. In order to determine the zone, we use both optical flow and statistical estimation. One of many methods of optical flow described in [14] is lane detection, from which we can determine the left and right border lines. To specify the top and bottom borders, we track the vehicle throughout the input frame along with their sizes and location; then we approximate the second derivative to measure the min of function size-location [15] (Fig. 2d).

Additionally, it is a common practice to get the sky region using watershed segmentation in combination with the horizon line [16]–[18]. To specify the region, the watershed segmentation method extracts the sky region from the input frame, which results in the binary image. The white pixels represent the sky, and the black ones are not belonging the region. The Fig. 2c shows the result of the segmentation; the pink region is the sky region we mention above. Additionally, the main reason for covering all the sky is that the small part could make the wrong decision. For example, the sun could overshadow by the cloud, which makes the overcast scene; However, in the other time, the cloud changes to position and the sun continues shining, the scene in this situation is the clear sky. Therefore, the small sky region could cover the small part, in this case, the part is a cloud, which results in false predictions.

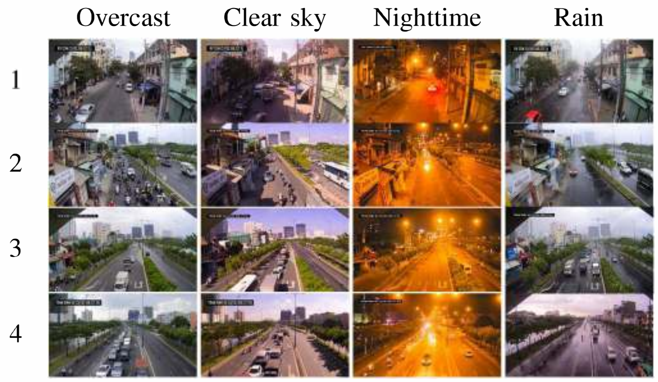


Fig. 3. Frames captured from five difference cameras in the Traffic Surveillance System at the same time on Vo Van Kiet Avenue, Ho Chi Minh City. The first column is show the camera in the overcast condition. The second column is in the clear sky. The third column is in rain, and the last column is in nighttime.

B. Scene Model

In our approach, we extract features from the background image obtained using the algorithm in [19]. In the sky region, we evaluate the scenes based on the histogram on the RGB color space. The histogram of the sky zone is suitable to detect the daylight and nighttime. All the traffic surveillance camera always have sky (Fig. 2); Therefore, the feature gotten from sky always be cogitate in training.

In the observation zone, we use the CIE-Lab color space. In the CIE-Lab color space, L stand for the lightness intensity; *a* and *b* stand for the average color, from which, we calculation the mean of *ab* value to model the change of the illumination. The shadow and overcast have their own range value; therefore, the mean of *ab* is the primary value when training data as shown in Eq. 1:

$$\sigma_L = \frac{1}{K}\sum_{i=0}^{K}(\rho_L - \mu_L)^2 \tag{1}$$

where K is the number of pixels in observation zone; μ_L, μ_a, and μ_b is the mean color value of all pixels; σ_L is the standard deviation of color value.

To determine the occurrence of shadows in the current frame, the sum of μ_a and μ_b is less or equal than the T (Eq. 2).

$$\mu_{ab} = \mu_a + \mu_b \leq T \tag{2}$$

$$\rho_{shadow} = \begin{cases} 255, \; if \; \rho_L \leq \mu_L - \sigma_L/3 \\ 0 \end{cases} \tag{3}$$

where $T = 250$ based on practical experiments,

In addition, to get the shadow frame in binary image, the pixel which have the L color less or equal then $\mu_L - \sigma_L/3$

For the determinant of the rainy scene, we use combination of μ_{ab} on the observation zone and the histogram in sky region.

C. Feed Forward Neural Network

Feed forward neural networks are preferably appropriate for modeling relationships between input variables or set of predictors and response or output variables. Otherwise stated, they are suitable for any practical mapping problem which shows how a number of input variables affect the output variable. The feed forward neural networks are the most widely studied and used neural network model in practice. In our approach, we feed the neural network with the features extracted from the scene model section above. Each camera has to train individually, because the size of the sky region, the observation zone is different from others. Moreover, the building alongside the road could affect the result, specially the clear sky and the overcast scene. The weight in result of the FFNN should be updated once per week to make the point that the annual weather could not lower the positive rate of the decision.

III. Experiments and Discussion

A. Quantitative Evaluation

In this section, we show the evaluation of our approach and prove the benefit of the method for the TSS. We evaluate the proposed method on the datasets captured on the Vo Van Kiet Avenue in the Ho Chi Minh City, Viet Nam. The videos were recorded at several locations from single pole-mounted cameras with the frame rate of 30 fps and the resolution of 640x360. Moreover, we also performed online testing with the video stream. The cameras should be from along the avenue to avoid that the local outdoor environment affects the detection. The scene of the camera for testing illustrates in Fig. 3, each row represents for each camera in the experiment, and the columns show the cases in each scene: overcast, clear sky, nighttime, and rain.

Table I shows the results of quantitative evaluations, each row corresponds the same order row in Fig. 3. The first three cases are recorded, and the last one is online video stream. The measure approach is mentioned in [20]. The accuracy of recognition via three videos were more than 85%, especially, the night time were more than 95%. For the testing in real-time dataset, the accuracy was lower than running from captured video; However, the deviation was just approximately 5%.

As can be seen in Table I, the result of the clear sky and nighttime scene get the better result than the rain and overcast scene. The overcast scene happens when the cloud that passed across and nearly covers the sky. In some case, the great cloud that almost darkened the sky which makes the method wrongly detect the nighttime instead of the overcast, or the time of cloud covering the sky is too short; the feature was not clear, then the method recognizes the clear sky as an alternative. The rainy scene got the lowest results in Table I. Because the rainy scene always comes with the thick cloud with a gloomy sky that leads to mistakenly detect the night as a substitute. Moreover, the shadow rain, which makes the reflection on the road, could be confused with the clear sky scene.

TABLE I
QUANTITATIVE EVALUATION

Case		OC	CS	RA	NT
1	Groundtruth	1000	5000	500	3500
	Result	935	4680	423	3412
	Accuracy(%)	93.5	93.6	84.6	97.5
2	Groundtruth	1000	3000	600	3500
	Result	893	2856	553	3364
	Accuracy(%)	89.3	95.2	92.2	96.1
3	Groundtruth	1000	3500	600	3000
	Result	874	3361	546	2927
	Accuracy(%)	87.4	96	91	97.6
4	Groundtruth	1000	1000	300	2000
	Result	793	786	237	1921
	Accuracy(%)	79.3	78.6	79	96.05
Average accuracy(%)		**87.38**	**90.85**	**86.7**	**96.81**

B. Speed Evaluation

In this section, we illustrate the speed performance of the proposed approach. Because the scene estimation is one of the processes in TSS, the non-functional requirement of the method is that the speed must be taken as short as possible. According to experiments, the numbers of frame per second we execute the method in each case is high. As can be seen in Fig. 4, the uppermost is 130, and the lowermost is 110. There are some factors affect the time-consuming, and the following part will explain.

The speed performance of the proposed method is conditional on the magnitude of the sky region and the observation zone. As can be seen, the first and the fourth cases observe the small streets and the narrow part of the sky; Therefore, the highest speed is in those cases. With the sky, regions cover nearly one-third of the frame, and the observation zones are double with two cases above, the second, the third, and the fifth cases take more time for processing. To conclude, the running time of the proposed is satisfied for performing in real-time with other processes in TSS.

IV. CONCLUSION

For an intelligent traffic surveillance system, it is crucial point that the system should be able to determine different scenes and outdoor traffic conditions. In the paper, we illustrate the method to detect and recognize the scene of the environment (overcast, clear sky, rain, nighttime). Our most significant contribution to this study is that the traffics surveillance system uses our method to handle the suitable method running in real-world (motion segmentation, vehicle tracking and classification). The novel approach using Feed Forward Neural Network with CIE-Lab color space and histogram feature has shown the improvement in performance over running real-time system. Base on the proposed method, the future studies is that the technique can give more detail the anomaly scene such as rush-hour [21], accident [22], raindrop-tampered [23],..., and the transition time between scenes.

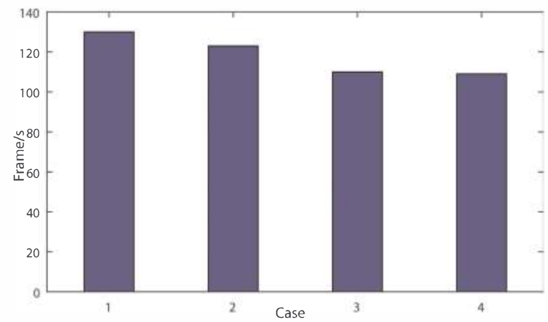


Fig. 4. Speed Evaluation from the experiment. The x-axis presents each case for testing, and the y-axis describes the running time. The columns show the frame rate for each case.

REFERENCES

[1] B. Tian, B. T. Morris, M. Tang, Y. Liu, Y. Yao, C. Gou, D. Shen, and S. Tang, "Hierarchical and networked vehicle surveillance in its: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 25–48, Jan 2017.

[2] N. Buch, S. A. Velastin, and J. Orwell, "A Review of Computer Vision Techniques for the Analysis of Urban Traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 920–939, sep 2011.

[3] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognition Letters*, vol. 34, no. 1, pp. 3–19, jan 2013.

[4] R. Cucchiara, M. Piccardi, and P. Mello, "Image analysis and rule-based reasoning for a traffic monitoring system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 2, pp. 119–130, jun 2000.

[5] U. Meis, W. Ritter, and H. Neumann, "Detection and classification of obstacles in night vision traffic scenes based on infrared imagery," in *Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems*, vol. 2. IEEE, 2003, pp. 1140–1144.

[6] Y.-L. Chen, C.-T. Lin, C.-J. Fan, C.-M. Hsieh, and B.-F. Wu, "Vision-based nighttime vehicle detection and range estimation for driver assistance," in *2008 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, oct 2008, pp. 2988–2993.

[7] R. O Malley, E. Jones, and M. Glavin, "Rear-lamp vehicle detection and tracking in low-exposure color video for night conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 453–462, 2010.

[8] K. Robert, "Night-Time Traffic Surveillance: A Robust Framework for Multi-vehicle Detection, Classification and Tracking," in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, sep 2009, pp. 1–6.

[9] H.-Y. Cheng and S.-H. Hsu, "Intelligent Highway Traffic Surveillance With Self-Diagnosis Abilities," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1462–1472, dec 2011.

[10] S. V.-U. Ha, L. H. Pham, H. N. Phan, and P. Ho-Thanh, "A robust algorithm for vehicle detection and classification in intelligent traffic system," *16th Asia Pacific Industrial Engineering & Management Systems Conference (APIEMS 2015)*, no. 1, pp. 1832–1838, 2015.

[11] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337–1342, oct 2003.

[12] L. Xu, F. Qi, and R. Jiang, "Shadow Removal from a Single Image," in *Sixth International Conference on Intelligent Systems Design and Applications*, vol. 2. IEEE, oct 2006, pp. 1049–1054.

[13] S. V.-U. Ha, N. T. Pham, L. H. Pham, and H. M. Tran, "Robust Reflection Detection and Removal in Rainy Conditions using LAB and HSV Color Spaces," *REV Journal on Electronics and Communications*, vol. 6, no. 1, pp. 13–19, aug 2016.

[14] H.-H. Nguyen, P.-T. Tran-Huu, H. M. Tran, and S. V.-U. Ha, "Improved optical flow estimation in traffic monitoring system," in *2013 Third World Congress on Information and Communication Technologies (WICT 2013)*. IEEE, dec 2013, pp. 165–169.

[15] S. V.-U. Ha, L. H. Pham, H. M. Tran, and P. Ho-Thanh, "Improved vehicles detection & classification algorithm for traffic surveillance system," *Journal of Information Assurance and Security*, vol. 9, no. 5, pp. 268–277, 2014.

[16] S. Beucher and F. Meyer, "The morphological approach to segmentation: the watershed transformation," *Mathematical Morphology in Image Processing*, vol. 34, pp. 433–481, 1993.

[17] A. Cohen, C. Meurie, Y. Ruichek, J. Marais, and A. Flancquart, "Quantification of GNSS signals accuracy: An image segmentation method for estimating the percentage of sky," in *2009 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*. IEEE, 2009, pp. 35–40.

[18] C. Couprie, L. Grady, L. Najman, and H. Talbot, "Power Watershed: A Unifying Graph-Based Optimization Framework," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, pp. 1384–1399, jul 2011.

[19] X.-Y. Wang, J.-F. Wu, and H.-Y. Yang, "Robust image retrieval based on color histogram of local feature regions," *Multimedia Tools and Applications*, vol. 49, no. 2, pp. 323–345, 2010.

[20] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, jul 2009.

[21] X. Song and R. Nevatia, "Detection and Tracking of Moving Vehicles in Crowded Scenes," in *2007 IEEE Workshop on Motion and Video Computing (WMVC'07)*, no. Mcmc. IEEE, feb 2007, pp. 4–4.

[22] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi, "Traffic monitoring and accident detection at intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 2, pp. 108–118, jun 2000.

[23] C.-C. Yu, H.-Y. Cheng, and Y.-F. Jian, "Raindrop-Tampered Scene Detection and Traffic Flow Estimation for Nighttime Traffic Surveillance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1518–1527, jun 2015.