

Dynamic Features Descriptor for Road User Recognition Using Hierarchical Graph Dynamic Gradient Pattern

Ma'moun Al-Smadi, Khairi Abdulrahim, Rosalina Abdul Salama

Abstract: Accurate and precise vehicle recognition and classification play a major role in analyzing and understanding traffic surveillance systems. This paper proposes a dynamic feature descriptor to recognize and classify road users based on graph representation. Local gradient patterns are computed based on the grayscale difference on the four directions across the center pixels. Dynamic gradients are determined according to the effective gradient computed as the mean value of all gradients. Hierarchical Graph using angular rotation pattern are applied to extract Dynamic Gradient Patterns (DGP). The central pixel is represented by Hierarchical Graph of Dynamic Gradient Patterns (HG-DGP). The proposed method learns dynamic representation adaptively to achieve efficient recognition with higher accuracy and lower pre-processing. The experimental results show that the proposed technique combined with support vector machine is efficient and discriminative for road user recognition and classification.

I. INTRODUCTION

With the massive increase of road users and vehicles, there has been a great interest in vehicle recognition and classification for Intelligent transportation systems (ITS). Therefore, it is important to develop techniques or methods that allow accurate recognition and classification of urban road users. Conventional technologies collect information using various types of sensors like inductive loops that have limitations in cost, maintenance and discrimination capability. However, these techniques are highly affected by intrusive, malfunctions and environmental conditions. In recent years, video cameras combined with image processing and computer vision techniques offer an attractive capability for data acquisition, since it provides a lot of valuable information about the traffic. These systems are easy to install, maintain and upgrade in urban environments with relatively low cost and wide variety of applications. Video-based applications may include vehicle recognition, classification, counting, speed measurement and incident detection [1]. In general, vision-based techniques are used for feature extraction to achieve object recognition and classification [2].

Revised Manuscript Received on December 22, 2018.

Ma'moun Al-Smadi, Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Negeri Sembilan, Bandar Baru Nilai, Nilai 71800, Malaysia

Khairi Abdulrahim, Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia (USIM), Negeri Sembilan, Bandar Baru Nilai, Nilai 71800, Malaysia

Rosalina Abdul Salama, Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Negeri Sembilan, Bandar Baru Nilai, Nilai 71800, Malaysia

Most video recognition and classification systems utilize either motion segmentation or appearance feature extraction. Motion features are measured using traffic dynamics. Background subtraction, frame differencing and optical flow are the main techniques used in vehicle recognition. In contrast to motion segmentation techniques that detect motion only, appearance-based techniques detect stationary objects in images or videos [3]. Visual appearance features include local symmetry edge operators, geometrical features, three-dimensional modelling and spatial invariance edge-based histogram [5,6].

In recent years, these simple features evolve into more robust and stable features that allow direct recognition and classification of vehicles. Scale Invariant Feature Transformation (SIFT) [4], Speeded Up Robust Features (SURF) [7], Histogram of Oriented Gradient (HOG) [2] and Haar-like features [8] are used extensively in the literature for vehicle recognition and classification. A comprehensive review of vehicle recognition techniques is described in [1].

Vehicle make and model classification using support vector machine (SVM) and a combination of SIFT and bag-of-words was addressed in [9]. SIFT was applied in [10] to extract and match features from the segmentation obtained by background subtraction. Symmetrical SURF descriptor was used to recognize vehicle make and model in [11] by identifying object symmetry about vertical axis. It does not require segmentation and performs well in real time. In [12] HOG descriptor and SVM was applied to distinguish between vehicles and background by detecting the underneath vehicles shadow. They detect the forward view of the vehicle to adapt variations in lighting conditions. Similar to the previous work, symmetrical HOG features were applied to extract shadows underneath vehicles in [13]. HOG descriptor and SVM was applied to distinguish between people and vehicles from different viewpoints in [14]. They use stereo cue to distinguish between person or vehicle candidates. Separate classifiers were developed and applied simultaneously to achieve joined result for all viewpoints. More accurate results were obtained by combining HOG and Haar-like features in [15].

Advanced techniques like deep neural networks (DNN) was used for vehicle recognition and classification in [16]. They classify vehicles into cars, sedans, and vans using small data set. SVM was used with geometric and appearance-based features in [17] for multiclass and intra class vehicle type classifications.

Its application was limited to small, medium, and large categories. In [18], DNN was used for feature extraction and linear SVM for classification.

In this work, vision-based system is proposed to recognize and classify different types of road users. The proposed technique make use of graph representation and dynamic gradient calculation. Graph based theory is utilized to describe the spatial angular relation between image pixels and their neighbors. Dynamic gradient pattern is used to represent effective gradients that are above the mean of all gradients surrounding the central pixel. This paper is organized as follows: section 2 introduce the proposed method, which include graph based pixel representation and the dynamic gradient pattern transformation. Experimental results and discussion are demonstrated in section 3. Finally, section 4 conclude this work and propose directions for future work.

II. HIERARCHAL GRAPH REPRESENTATION OF DYNAMIC GRADIENT PATTERN

This section describe the Hierarchical Graph Representation of Dynamic Gradient Pattern (HG-DGP) generation. An overview of the proposed approach is illustrated in figure 1. The descriptor operates on gray scale difference. The descriptor for the center pixel is represented

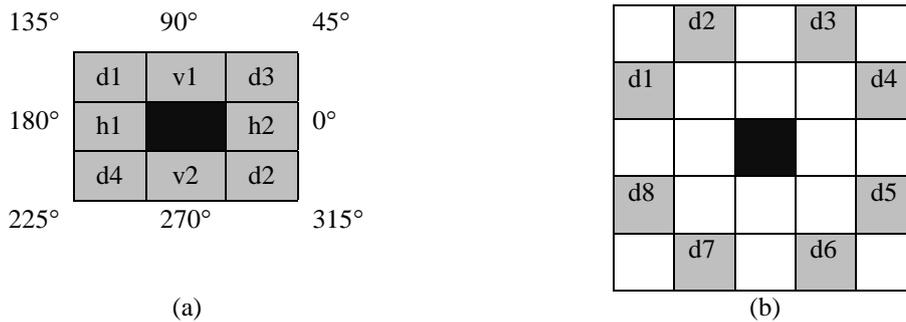


Fig. 1 Graph based dynamic gradient pattern descriptor

The descriptor operates on gray scale images. The pixels are represented by a graph of effective gradients e. g., the maximum directional gradient and the directional gradients above the mean of all surrounding gradients. Thus, 8-bins are needed to represent four bidirectional gradients. The proposed graph structures

This new descriptor is called graph based directional gradient pattern

It includes the main gradient an any effective gradient that is above the mean of all gradients. thus, it will provide better representation for corners according to the corner angle and the gray scale variation.

Object edges can be considered as vectors with magnitude and direction representation. line, curve and corner edges can have single or multiple magnitudes and directions as shown in figure. Thus, edge pixels can have a single or multiple (corner or curve) magnitude and direction representation. The gradient magnitude g_{ij} of the graph is computed as the maximum absolute gray scale difference across the central pixel. This will provide four gradients. The direction of the gradient is determined by the gradient orientation and sign.

by two gradient graphs gradient1 and gradient2. Gradient1 has for gradient magnitudes $|d_1 - d_2|$, $|v_1 - v_2|$, $|d_3 - d_4|$ and $|h_1 - h_2|$. Gradient2 is calculated as $|d_1 - d_5|$, $|d_2 - d_6|$, $|d_3 - d_7|$ and $|d_4 - d_8|$. Each gradient magnitude has two direction according to the sign of the difference. Therefore, 8-bins are required to represent 8 equally spaced angles for gradient1 and another 8-bins are required to represent another 8 angles for gradient2 that are dividing the angles of pattern1 equally.

The proposed descriptor is based on hierarchical graph of weighted directional gradient within a local area around the pixel. Angular rotation pattern mapping is proposed for multi-scale analysis by varying the scale and rotation of the pixel connectivity. An angular graph can be expanded to incorporate multiple vertices based on the rotational angle across the center pixel. A pair of pixels are denoted as a vertex v_i in the graph that pass through the center pixel.

Two pattern mapping are generated with neighborhood of 8 pixels each. For the first pattern the angle between vertices is 45° and all vertices is $45n$ from the vertical or horizontal line, where $n = 0, 1, \dots$. Two vertices are connected if the angles between the them is equally spaced. for the second hierarchical pattern non-orthogonal vertices that pass through the central pixel is used.

$$g_{ij} = |d_{ni} - n_{nj}|$$

Where x_{ni} and x_{nj} are corresponds to gray scale values of a pair of neighboring pixels around the central pixel. The dynamic gradient can be generated using threshold value as the average of the for gradient values.

$$T = \left(\frac{1}{4}\right) \sum_{a=1}^4 g_a$$

Graph gradients that will be included in the representation are dynamically selected as those gradients with absolute value greater than threshold T as:

$$g_a = \begin{cases} w_a \times g_{ij} & g_{ij} \geq T \\ 0 & otherwise \end{cases}$$

Where w_a is a weighting factor computed as:

$$w_a = \frac{g_{ij}}{\max(g_{ij})}$$

After pixel representation is calculated and accumulated into bins, feature vector is generated by concatenating the bins of image regions. The size of the feature vector is computed as:

$$N = \left(\frac{R_w}{C_w} - 1\right) \times \left(\frac{R_h}{C_h} - 1\right) \times B \times H$$

where R is the image dimension w width and h height, C is the cell dimension, B is the number of cells per block, and H is the number of histograms bins.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed HG-DGP is evaluated and compared on a standard benchmark. Experiments were performed on a laptop (2.3 GHz, 4th generation core i5 CPU with 8-GB RAM, SSD storage and windows 10 64-bit). The data set used in this paper was taken from MIO-TCD classification

dataset [16], it contains a total of 648,959 colored images of various categories, sizes and orientations. Images are divided into 11 categories which include: Pedestrian, Bicycle, Motorcycle, Car, Bus, Pickup truck, Work van, Single unit truck, Articulated truck, Non-motorized vehicle and Background. A total of 2065 images were selected from the first 8 class in the data set, which represent urban road users only. The total image samples for each class is shown in Figure 2, which include 106 Pedestrian, 126 Bicycle, 139 Motorcycle, 1002 Car, 213 Bus, 119 Pickup truck, 160 Work van and 200 Single unit truck.

The classification was performed using support vector machine with linear, quadratic, cubic and Gaussian kernel. Ten rounds of experiments were conducted, and the average value of the results were reported.

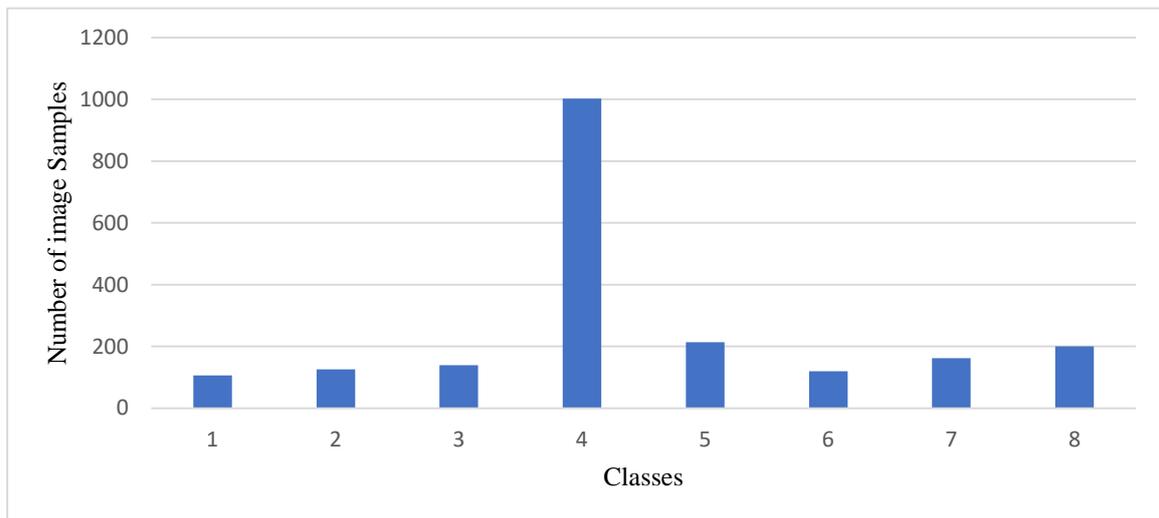


Fig. 2 Histogram of training image set per class category

The input images are converted to gray scale and resized into 32*32 pixel. Then it is divided into several equal sized cells. For the single cell case, the bins are used to represent the whole cell. A ten-fold cross validation using multiclass support vector machine with linear, quadratic, cubic and

Gaussian kernel were used for comparison. The classification accuracy for various SVM kernels, cell dimension and feature vector size is shown in table 1 and table 2 for HOG feature descriptor and the proposed HG-DGP respectively.

Table. 1 Average classification accuracy in (%) for 8-bins HOG descriptor using Linear, Quadratic, Cubic and Gaussian SVM

| | | | | | |
|--------------|-------|-------|------|-------------|------|
| LinerSVM | 52.1 | 67.8 | 80.4 | 83.8 | 82.3 |
| QuadraticSVM | 42.9 | 75.4 | 84.3 | 87.8 | 86.4 |
| CubicSVM | 21.1 | 67.5 | 83.5 | 88.7 | 86.9 |
| GaussianSVM | 48.5 | 76.5 | 86.5 | 87.3 | 82.8 |
| Cell size | 32*32 | 16*16 | 8*8 | 4*4 | 2*2 |
| Features # | 8 | 32 | 128 | 512 | 2048 |

In the single cell experiment, the feature vector size equal to the number of bins multiplied by number of cells. The proposed HG-DGP descriptor achieves a classification accuracy of 93.2 for 256 features vector and cubic SVM. While the HOG descriptor achieves 88.7 classification accuracy at 512 features vector and cubic SVM. From tables

1 and 2, it is clear that the proposed descriptor can extract more discriminative information than HOG descriptor. Moreover, the 4.5% improvement in the classification accuracy was obtained with 50% smaller feature vector with the same classification technique (cubic SVM).

Table. 2 Average classification accuracy in (%) for 16-bins HDG descriptor using Linear, Quadratic, Cubic and Gaussian SVM

| | | | | | |
|--------------|-------|-------|------|------|------|
| LinerSVM | 58.6 | 79.2 | 87.7 | 89.2 | 87.8 |
| QuadraticSVM | 62.9 | 86.3 | 92.8 | 92.5 | 91.6 |
| CubicSVM | 63.9 | 87.5 | 93.2 | 93.1 | 91.3 |
| GaussianSVM | 63 | 86.1 | 91.4 | 90.7 | 88.3 |
| Cell size | 32*32 | 16*16 | 8*8 | 4*4 | 2*2 |
| Features # | 16 | 64 | 256 | 1024 | 4096 |

In the case of block with multiple cells, the bins for each block is the sum of cells bins constituting the block. The classification accuracy using various cell dimension with

2*2 cell block and feature vector size is shown in table 3 and table 4 for HOG feature descriptor and the proposed HG-DGP respectively.

Table. 3 Average classification accuracy in (%) for 8-bins HOG descriptor using Linear, Quadratic, Cubic and Gaussian SVM

| | | | | | | | | |
|--------------|-------|------|------|------|-------|------|-------------|------|
| LinerSVM | 72.6 | 84.6 | 87.7 | 85.7 | 73.1 | 86.6 | 89.1 | 87.0 |
| QuadraticSVM | 83.7 | 91.1 | 91.2 | 90.0 | 83.3 | 91.9 | 92.6 | 90.6 |
| CubicSVM | 81.5 | 91.1 | 92.3 | 90.5 | 82.2 | 92.5 | 93.1 | 91.4 |
| GaussianSVM | 85.4 | 91.6 | 90.1 | 87.1 | 85.3 | 92.6 | 91.7 | 88.1 |
| Cell size | 16*16 | 8*8 | 4*4 | 2*2 | 16*16 | 8*8 | 4*4 | 2*2 |
| Features # | 32 | 128 | 512 | 2048 | 32 | 288 | 1568 | 7200 |

In the multiple cells experiment, non-overlapping and overlapping blocks were tested. For the non-overlapping blocks the proposed technique achieves a classification accuracy of 95.0 for 256features vector, and cubic SVM. While the HOG descriptor achieves 92.3 classification accuracy at 512 feature vector and cubic SVM. On the other hand, the proposed technique with overlapping blocks

achieves a classification accuracy of 95.3 for 576features vector, and cubic SVM. While the HOG descriptor achieves 93.1 classification accuracy at 1568 feature vector and cubic SVM. From tables 3-4, the proposed HG-DGP descriptor outperform HOG descriptor in extracting discriminative information with smaller feature vector size.

Table. 4 Average classification accuracy in (%) for 16-bins HDG descriptor using Linear, Quadratic, Cubic and Gaussian SVM

| | | | | | | | | |
|--------------|-------|-------------|------|------|-------|------|------|------|
| LinerSVM | 81.7 | 88.8 | 89.8 | 88.2 | 82.3 | 89.8 | 89.3 | 88.5 |
| QuadraticSVM | 89.1 | 94.1 | 92.6 | 91.0 | 89.1 | 94.7 | 93.4 | 91.6 |
| CubicSVM | 89.8 | 95.0 | 93.6 | 91.7 | 90.5 | 95.3 | 94.1 | 92.2 |
| GaussianSVM | 88.7 | 92.9 | 91.6 | 88.8 | 89.1 | 93.1 | 91.5 | 89.3 |
| Cell size | 16*16 | 8*8 | 4*4 | 2*2 | 16*16 | 8*8 | 4*4 | 2*2 |
| Features # | 64 | 256 | 1024 | 4096 | 64 | 576 | 3136 | |

The proposed feature descriptor archives the best score of 95.3 using cubic SVM with a cell size of 8*8bit and overlapping blocks of 2*2 cell each. On the other hand, HOG descriptor achieves93.1 using cubic SVM with a cell size of 4*4bit and overlapping blocks of 2*2 cell each. For the best classification accuracy, the size of feature vector for the proposed HG-DGP descriptor is 576, while the size of HOG descriptor is 1568, which is about triple size.

Finally, the experimental results confirmed that the proposed HG-DGP feature descriptor is more efficient, accurate and compact in classifying road users. The results also showed that the use of dynamin gradient and expanding the local structure provide higher discrimination ability at lower feature vector size.

IV. CONCLUSION

The dynamic feature descriptor proposed in this paper HG-DGP achieve better classification accuracy with lower feature vector size as compared to the HOG descriptor using cubic SVM. It computes the gradient pattern for each pixel based on the gray scale difference across pixel. Hierarical angular graph combined with dynamic gradient pattern provide compact, accurate and more discriminative descriptor even for small size feature vector. Experimental results show that the classification accuracy based on the



proposed feature descriptor using multiclass cubic SVM outperform HOG for various block sizes. In future work, the proposed HG-DGP descriptor will be trained using deep learning and compared with other descriptors like local binary pattern LBP.

V. ACKNOWLEDGMENTS

This work was supported by University Science Islam Malaysia (USIM), Faculty of Engineering and Built Environment and Research and Innovation Management Center under the Research Acculturation Grant number: PPPI/FKAB/0118/051000/14618

REFERENCES

1. M. Al-Smadi, K., Abdulrahim and R.A., Salam, 2016. Traffic surveillance: A review of vision based vehicle detection, recognition and tracking. *International journal of applied engineering research*, 11(1), pp.713-726.
2. N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection," in *proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, Jun. 2005, pp. 886-893.
3. M. Al-Smadi M., Khairi, and R. A. Salam, "Cumulative frame differencing for urban vehicle detection." p. 100110G.
4. D.G.Lowe, "Object recognition from local scale-invariant features", in *proc. IEEE 7th international conference on Computer vision*, Jan. 1999, vol. 2, pp. 1150-1157.
5. S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1475-1490, 2004.
6. X. Ma, and W.E.L.Grimson, "Edge-based rich representation for vehicle classification," in *Proc. IEEE 10th International Conference on Computer Vision*, vol. 2, Oct. 2005, pp. 1185-1192.
7. H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *Computer vision (ECCV)*, Jan. 2006, pp. 404-417. Springer Berlin Heidelberg.
8. P. Viola, and M.Jones, "Rapid object detection using a boosted cascade of simple features," in *proc. of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, vol. 1, pp. I-511.
9. M. A. Manzoor and Y. Morgan, "Vehicle Make and Model classification system using bag of SIFT features," in *Computing and Communication Workshop and Conference (CCWC)*, 2017 IEEE 7th Annual, pp. 1-5, IEEE, 2017.
10. M. C. Narhe and M. Nagmode, "Vehicle classification using SIFT," *International Journal of Engineering Research and Technology ESRSA Publications*, 2014.
11. J.-W. Hsieh, L.-C. Chen, and D.-Y. Chen, "Symmetrical SURF and its applications to vehicle detection and vehicle make and model recognition," *IEEE Transactions on intelligent transportation systems*, vol. 15, no. 1, pp. 6-20, 2014.
12. X. Li and X. Guo, "A HOG feature and SVM based method for forward vehicle detection with single camera," in *Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2013 5th International Conference on, vol. 1, pp. 263-266, IEEE, 2013.
13. S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1773-1795, 2013.
14. F. Han, Y. Shan, R. Cekander, H. S. Sawhney, and R. Kumar, "A two-stage approach to people and vehicle detection with HOG-based SVM," in *Performance Metrics for Intelligent Systems 2006 Workshop*, pp. 133-140, 2006.
15. P. Negri, X. Clady, S. M. Hanif, and L. Prevost, "A cascade of boosted generative and discriminative classifiers for vehicle detection," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, p. 136, 2008.
16. Zhou, Y., and N. Cheung. Vehicle Classification Using Transferrable Deep Neural Network Features. <http://arxiv.org/pdf/1601.01145.pdf>. Accessed July 2, 2016.
17. Moussa, G. S. Vehicle Type Classification with Geometric and Appearance Attributes. *International Journal of Civil, Environmental,*

- Structural, Construction, and Architectural Engineering, Vol. 8, No. 3, 2014, pp. 277-282.
18. Y. Adu-Gyamfi, S. Asare, and A. Sharma, "Automated vehicle recognition with deep convolutional neural networks," vol. 2645, 01 2017.