HISTOGRAMS OF ORIENTED GRADIENTS FOR FAST ON-BOARD VEHICLE VERIFICATION

Gonzalo Ballesteros , Luis Salgado

ABSTRACT

Histograms of Oriented Gradients (HoGs) provide excellent results in object detection and verification. However, their demanding processing requirements bound their applicability in some critical real-time scenarios, such as for video-based on-board vehicle detection systems. In this work, an efficient HOG configuration for pose-based on-board vehicle verification is proposed, which alleviates both the processing requirements and required feature vector length without reducing classification performance. The impact on classification of some critical configuration and processing parameters is in depth analyzed to propose a baseline efficient descriptor. Based on the analysis of its cells contribution to classification, new view-dependent cell-configuration patterns are proposed, resulting in reduced descriptors which provide an excellent balance between performance and computational requirements, rendering higher verification rates than other works in the literature.

Index Terms— HOGs, vehicle verification, efficient descriptor configuration, view-dependent classification.

1. INTRODUCTION

Vision-based object detection from a moving platform becomes particularly challenging in the field of advanced driver assistance systems (ADAS). On-board vision systems are an excellent source of information where real-time detection of vehicles becomes a critical task, facing challenges derived from the variability of vehicles appearance, illumination, shadows, and vehicle speed.

The most used methodology for vehicles detection consists of two stages. First, the whole image is fast analyzed using knowledge-based [1][2], motion [3][4] or stereovision [5] methods to identify regions potentially containing vehicles. Second, these candidates are verified using features related to their appearance through model-based approaches or, more recently, learning-based methods [6] in which the vehicles characteristics are learned from a training set, and new candidates are classified according to learned patterns. Learning-based hypothesis verification is approached as a two-class classification problem: a feature vector is extracted from the image, and the sample is classified as vehicle or non-vehicle. Among the most commonly used descriptors, such as those based on PCA [7] or Gabor filters [8], HOGs stands out for their excellent performance in object detection. However, their performance is bounded by a tradeoff between complexity of descriptor configuration and real-time operation.

In this work, an efficient HOG configuration for pose-based on-board vehicle verification is proposed, which alleviates both the processing requirements and required feature vector length without reducing classification performance. The impact on classification of some critical configuration and processing parameters is in depth analyzed to propose a baseline efficient descriptor. Furthermore, subsets of cells corresponding to view-dependent patterns are investigated, resulting in reduced descriptors where only the most significant cells are considered for classification. The hypothesis that there are areas in the images which do not contain useful information or even gather misleading information for classification is verified. The classification accuracy is demonstrated on a large public database, outperforming other approaches recently proposed.

2. PREVIOUS WORK

Although originally proposed for people detection [9], HOGs were rapidly expanded to other fields such as face recognition [10]. In recent studies, HOGs have been adopted for video-based vehicle detection and verification, although with a limited exploration of the descriptor configuration as typically that for people and other objects detection are directly applied. This so-called standard HOG is present in works such as [11] for vehicle detection in aerial views, [12] for preceding vehicle detection, [13] for rear collision avoidance, or [14] for view-dependent vehicle verification. Although works either provide only qualitative results [12], or use limited non-public sequences and databases [13][11], in [14] a large public database of vehicle hypothesis obtained from an on-board forward looking camera is considered. Computational efficiency is addressed differently in the literature: from standard HOG ad-hoc hardware implementations [15], to descriptor simplifications [12][14][16] reducing the orientation range considered, modifying the weighted contributions to adjacent bins, or proposing alternative cell and blocks configurations to alleviate the cost of classification. The use of different block sizes is explored in [17], with fairly low accuracy results, while in [18] the use of masks adapted to the vehicle shape is proposed to speed up classification with good results in the classification between different types of vehicles. In [14], different reduced configurations are evaluated and the V-HOG, that uses only vertical cells, is proposed: it provides better verification accuracy than other approaches with a reduction of the computational cost that allows real-time operation. Combination with other features, such as Haar-like [11][16], is also proposed to speed up detection.

3. FEATURE EXTRACTION AND CLASIFICATION BASED ON HOG

3.1. Feature extraction phase

The main idea of using HOG is that objects appearance can be characterized by the local distribution of its edges orientation. The HOG descriptor results from the computation of local histograms of orientation of the image gradients in a grid. A scheme summarizing the feature extraction system using HOG is shown in Figure 1.

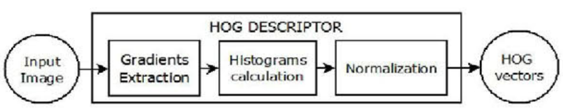


Figure 1. HOG descriptor generation

The gradients extraction phase computes, for each image pixel, the edge magnitude and orientation. The image is divided into cells, and for each cell a histogram of its pixels orientation is obtained. The last step is the normalization of the histograms to account for uneven illumination and shadows. As proposed in [9], cells are grouped into larger structures called blocks. For each block the non-normalized vector holding the histograms of its cells is normalized using any standard norm. The overlap of blocks is also suggested to make this step more robust. The final HOG descriptor is the resulting vector of the concatenation of the normalized blocks. Among the cell configurations proposed, the rectangular (R-HOG) geometry is assumed in this work, as it naturally adapts to the dominant vehicle geometry.

3.2. Classification phase

For HOG based classification of the input samples into vehicles and non-vehicles, Support Vector Machines (SVMs) have been extensively proposed: they render excellent results, and provide better generalization involving lower number of parameters than other discriminative approaches such as Neural Networks [20]. A linear SVM is adopted in this work.

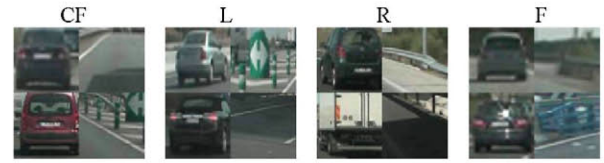


Figure 2. GTI Database images. For frontal (CF), left (L), right (R), and far (F) views: two vehicle (left) and two non-vehicle (right) samples.

Evaluation is carried out in the only extensive public database for vehicle verification proposed in the literature [18]. The GTI vehicle database is a complete data set with 4000 vehicles and 4000 non-vehicles images of size 64x64 pixels. Images were obtained from videos taken with an on-board forward looking camera, considering a large variability of situations typical from on-line hypothesis generation systems. Furthermore, to explore classification taking into account the vehicles pose, images are also organized into four classes according to their relative position and distance to the camera: frontal (CF), left (L) and right (R) views in the middle-close distance, and far distance (F). Examples of vehicle and non-vehicle images are shown in Figure 2. Experiments are carried out based on 5-fold 50% cross-validation methodology, and the classification accuracy, average percentage of correctly classified samples, is evaluated.

3.3. Experiments on HOG

The configuration of the parameters of HOG descriptor is important to perform a good classification. Table I shows the best results obtained with the standard HOG (S-HOG) training individual classifiers for each view. Different values of the number of cells ($\eta\times\eta$=4, 16, 64, 256) and the number of orientation bins (β=8, 12, 16, 32) are evaluated. The L2-norm is used to normalize blocks of 2x2 cells.

View	Acc (%)	η	β
CF	99.48	4	8
L	97.64	4	8
R	96.22	4	16
F	97.76	4	8

Table I: Best accuracy results for standard HOG

Best results are obtained for η=4 (16 cells), the lowest spatial resolution considered, which serve as a starting point to reduce the processing requirements in this work. As expected, highest scores correspond to the frontal view: hypothesis generated for vehicles located in front of the own vehicle show well defined and quite stable geometrical patterns that adapt perfectly to the HOG topology. Regarding orientation resolution, low values (the lowest ones for three of the four views) are enough for a good rating. Larger granularity does not imply a better outcome.
In the following sections, first alternative configurations of the HOG descriptor parameters to reduce computation without loosing performance are proposed. Then, based on the hypothesis that there are cells in the descriptor not

η=4	S-HOG		[-π, π]		NCCG		2x1		1x1	
View	β=8	β=16	β=8	β=16	β=8	β=16	β=8	β=16	β=8	β=16
CF	99.48	99.18	99.00	99.28	99.06	99.14	99.14	99.52	99.20	99.56
L	97.64	97.04	98.72	99.04	98.40	99.00	98.50	98.84	98.42	98.86
R	96.00	96.22	98.14	98.36	97.44	98.28	97.96	98.02	97.94	98.22
F	97.76	97.04	97.38	98.50	97.54	98.02	97.70	98.16	97.60	97.98

Table II: Accuracy rates for efficient configuration of parameters

Frontal View (a)				Left View (b)				Right View (c)				Far View (d)			
A(%)	NC	DC	Δ(%)	A(%)	NC	DC	Δ(%)	A(%)	NC	DC	Δ(%)	A(%)	NC	DC	Δ(%)
99.56	16	-	-	98.86	16	-	-	98.22	16	-	-	97.98	16	-	-
99.78	15	3	+0.22	99.12	15	5	+0.26	98.38	15	4	+0.16	98.32	15	2	+0.34
99.62	14	14	+0.08	99.16	14	2	+0.30	98.48	14	15	+0.26	98.28	14	3	+0.30
99.70	13	11	+0.14	99.08	13	9	+0.22	98.34	13	8	+0.12	98.36	13	6	+0.38
99.68	12	2	+0.12	99.20	12	6	+0.36	98.50	12	3	+0.28	98.20	12	10	+0.22
99.70	11	15	+0.14	99.26	11	11	+0.40	98.44	11	5	+0.22	98.16	11	4	+0.18
99.62	10	10	+0.08	99.14	10	15	+0.28	98.36	10	7	+0.14	98.34	10	5	+0.36
99.56	9	12	0	99.06	9	1	+0.20	98.30	9	10	+0.08	98.34	9	11	+0.36
99.48	8	9	-0.08	98.90	8	3	+0.04	98.38	8	13	+0.16	98.36	8	1	+0.38

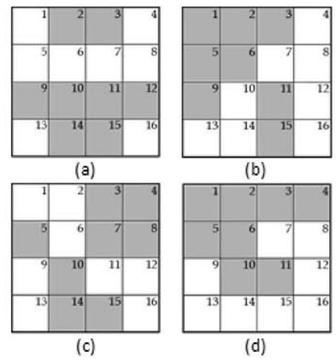


Figure 3: Classification accuracy evolution removing less significant cells from the descriptor: NC indicates the number of cells composing the descriptor, DC the deleted cells identification number, and Δ(%) the accuracy deviation with respect to the performance of the entire descriptor.

relevant for classification, we analyze their influence and new view-dependent cell configurations are proposed.

4. FEATURE EXTRACTION AND CLASIFICATION BASED ON HOG

The impact on the performance of alternative processing and configuration parameters of the descriptor is studied.

A. *Gradients extraction*

Although orientation sign is typically disregarded for efficiency purposes [13][14][16], Table II demonstrates its high impact (columns [-π, π]) particularly on non-frontal vehicle views. Compared with S-HOG (no sign considered), better results are obtained for the same granularity, being more significant for β=16 in all views. It seems more important to exploit the whole range of gradient orientations than have smaller orientation bins to describe the structure of the vehicles. Improved results are achieved keeping the same computational cost.

B. *Histograms Calculation*

The next improvement focuses on reducing computational requirements of the histograms in S-HOG. According to the edge orientation, weighted contributions are provided to the corresponding bin and the closest one. This results on a costly interpolation process whose impact on the verification is evaluated. Column NCCG in Table II shows the results achieved when interpolation is removed: each gradient contributes only to the bin that corresponds to its orientation. Results are stable, improving for most configurations while slightly underperforming for the others. The non-interpolation causes better accuracy values for higher values of β, having significant impact on computational saving.

C. *Normalization*

As proposed in [9], the use of overlapped 2x2 blocks is typically assumed [14], resulting on a new histogram generation and longer feature vectors. To reduce the vector size, 2x1 and 1x1 blocks are explored. Table II shows that for 1x1 blocks, performance of the descriptor is still maintained, not requiring the computation of a second histogram, with significantly better accuracy than S-HOG.
Therefore, the descriptor with the proposed modifications improves the performance with respect to S-HOG while significantly reducing the computational requirements. In average (i.e. computing the average accuracy of the four classifiers for each β), there is a 0.57% and 1.3% gain for β=8 and 16 respectively. As a conclusion, the values of η and β chosen to provide a good balance between performance and computational cost are η=4 and β=16.

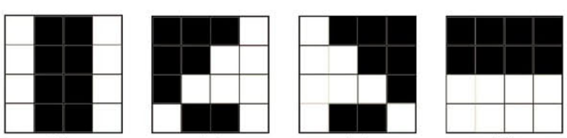
5. INFLUENCE OF HOG CELLS

In this section we carried out a study to analyze the impact of the different cells of the HOG descriptor in the classification. The goal is to eliminate cells that provide less or misleading information for classification, thus reducing both, the length of the descriptor – faster classification - and the cost of feature extraction.
Figure 3 shows, for each view, the verification accuracy evolution when cells are sequentially removed from the descriptor (Δ(%) in left tables), and the corresponding topology of removed cells (right images, grey colored).
The first conclusion is that for all views, removing the less significant cells do improve verification. Deviations are

positive, with a gain in classification which is less significant for the frontal view (as expected) than for the more complex lateral and far views. Therefore, with a much smaller histogram, a good discrimination is still reached, improving of performance through the elimination of cells.
Another relevant conclusion is that the topology of the cells having negative impact in verification is related to the vehicle view particularities. Most important cells for the frontal view (Figure 3(a)) are the outer ones, as they typically hold main edges of vehicle rear when frontally observed, and thus provide more discriminating information. For the left view (Figure 3 (b)), most of the relevant information is within and below the right to left diagonal cells, which agrees with the image areas likely holding vehicle information when observed from its right. Similar conclusions can be obtained from the cells distribution for the right view, but considering the left to right diagonal. Finally, for the far view images (Figure 3 (d)), which show smoothed edges due to interpolation (images are scaled to 64x64 pixels), the lower cells are more discriminative as they hold more stable and contrasted edge information.
In conclusion, the study of the cells' influence in the HOG descriptor helps to understand which image parts are more decisive for the classification. In addition, we can assure that the use of full HOG descriptors, which conveys a high computational cost, is unnecessary. Using a part of the HOG descriptor is sufficient for the correct classification.

6. DISCUSSION

Firstly in this work changes in processing and parameters of the S-HOG descriptor are proposed: modifying the gradient extraction phase improves performance while maintaining the computational cost, and simplifying the histogram generation and normalization steps manage to reduce the computational cost while marginally affecting verification. However, motivated from the conclusions of the previous section, view-dependent cell configurations are here proposed (Figure 4) which result in faster to compute and shorter HOG descriptors.



(a) (b) (c) (d)
Figure 4. HOG cells – in white - considered for (a) frontal view; (b) left view; (c) right view; (d) far view.

As already mentioned, in the middle close frontal view the most significant cells are those corresponding to the sides of the vehicle, so it was decided to use these cells only for classification. Regarding the left and right views the same analysis was made, and the cells in and below each view diagonal are kept, removing also those that typically hold shadows cast by the vehicles. For the far view only the lower-half cells are considered: edges corresponding to vehicle wheels, lights or underneath shadows are here more important.
Table III compares the performance of S-HOG, our view dependent HOG descriptors and V-HOG [14]. Our method outperforms S-HOG while removing costly processing steps: no interpolation and no multi-cell normalization are applied, halving the number of cells to compute. As a result, in our non-optimized implementation, average feature extraction computational time savings ~60% are achieved.

S-HOG		Our method		V-HOG[14]	
CF	99.48	CF	99.40	CF	97.68
L	97.64	L	98.96	L	97.02
R	96.22	R	98.14	R	95.54
F	97.76	F	98.24	F	95.60

Table III: Comparison of accuracy (%) between different methods

Furthermore, in terms of descriptor length, the average 360 components required for S-HOG go down to 128 with our proposal. In [14], an efficient HOG descriptor using only vertical cells, V-HOG, is proposed to carry out view dependent vehicle verification. Compared with the best results achieved with V-HOG, our proposal also largely outperforms for all views with a similar cost. Figure 5 shows some false positives examples that illustrate the dependency of classification with the quality of the generated hypothesis. In terms of average verification rate, i.e. averaging the accuracy for all views, our proposal reaches 98.69%, an excellent score that largely outperforms the results reported in [13] (92.9%), and [16] (94%), that proposes a cascade of boosted classifiers combining Haar-like and HOG features.

(a) (b) (c)
Figure 5. (a) Partial view, (b) low contrast, and (c) distorted hypothesis.

7. CONCLUSIONS

The adaptation of the HOG descriptor for fast vehicle verification is proposed. An adequate configuration of the descriptor and the simplification or elimination of some processing steps lower the complexity bounding the loss in verification accuracy. A study of the influence of cells in classification is carried out, showing that a significant number of the cells do not contribute positively to verification, and that their spatial configuration relates to the view of the vehicles to be verified. This information has been used to propose new view-dependent HOG cells configurations which provide a suitable balance between performance and processing steps simplification, rendering higher verification rates than other works in the literature.

8. REFERENCES

[1] Junyeon Hwang, Kunsoo Huh, and Donghwi Lee. Vision based vehicle detection and tracking algorithm design.Optical Engineering, 48(12):127201–127201–10, 2009.

[2] Gwang Yul Song, Ki Yong Lee, and Joon Woong Lee. Vehicle detection by edge-based candidate generation and appearance-based classification. In Intelligent Vehicles Symposium, 2008 IEEE, pages 428–433, 2008.

[3] F. Woelk and R. Koch. Robust monocular detection of independent motion by a moving observer. In Proceedings of the 26th DAGM Symposium, ser. Lecture Notes in Computer Science, pages 27–35.

[4] J. Arróspide, L. Salgado, M. Nieto and R. Mohedano, "Homography based ground plane detection using single on-board camera", IET Intelligent Transport Systems, vol. 4, no. 2, pp. 149–160, 2010.

[5] F. Oniga and S. Nedevschi. Processing dense stereo data using elevation maps: Road surface, traffic isle,and obstacle detection. *Vehicular Technology, IEEE Transactions on*, 59(3):1172–1182, 2010.

[6] Sivaraman Sayanan and Trivedi, Mohan Manubhai, "A General Active-Learning Framework for On-Road Vehicle Recognition and Tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, pp. 267-276, 2010.

[7] J. Zhou, D. Gao and D. Zhang, "Moving Vehicle Detection for Automatic Traffic Monitoring," *IEEE Transactions on Vehicular Technology*, vol. 56, pp. 51-59, 2007.

[8] S. Zehang, G. Bebis and R. Miller, "On-road vehicle detection using evolutionary Gabor filter optimization," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, pp. 125-137, 2005.

[9] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *IEEE Conf. on Computer Vision and Pattern Recognition,* pp. 886–893, Apr. 2005.

[10] O. Daniz and G. Bueno and J. Salido and F. De la Torre, "Face recognition using Histograms of Oriented Gradients," *Pattern Recognition Letters*, vol. 32, pp. 1598 - 1603, 2011.

[11] Pengpeng Liang and Gregory Teodoro and Haibin Ling and Erik Blasch and Genshe Chen and Li Bai, "Multiple Kernel Learning for Vehicle Detection in Wide Area Motion Imagery," *Information Fusion (FUSION), 2012 15th International Conference on*, pp. 1629-1636, 2012.

[12] L. Mao, M. Xie, Y. Huang, and Y. Zhang, "Preceding vehicle detection using Histograms of Oriented Gradients," in *Proc. International Conference on Communications, Circuits and Systems*, 2010, pp. 354–358.

[13] D. Balcones, D. F. Llorca, M. A. Sotelo, M. Gavilán, S. Álvarez, I. Parra, and M. Ocaña, "Real-time vision-based vehicle detection for rear-end Collision mitigation systems," in Computer Aided Systems Theory - EUROCAST, LNCS, 2009, vol. 5717, pp. 320–325.

[14] Jon Arrospide, Luis Salgado, and Massimo Camplani, "Image-based on-road vehicle detection using cost-effective Histograms of Oriented Gradients," *Journal of Visual Communication and Image Representation*, vol. 24, pp. 1182 - 1190, 2013.

[15] Seonyoung Lee and Haengseon Son and Jong Chan Choi and Kyoungwon Min, "HOG feature extractor circuit for real-time human and vehicle detection," *ENCON 2012 - 2012 IEEE Region 10 Conference*, pp. 1-5, 2012.

[16] Pablo Negri, Xavier Clady, ShehzadMuhammad Hanif, and Lionel Prevost. A cascade of boosted generative and discriminative classifiers for vehicle detection. EURASIP Journal on Advances in Signal Processing, 2008(1):782432.

[17] Xianbin Cao and Changxia Wu and Pingkun Yan and Xuelong Li, "Linear SVM classification using boosting HOG features for vehicle detection in low-altitude airborne videos," *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 2011.

[18] Yang Lv, Benjamin Yao, Yongtian Wang, and Song-Chun Zhu, "Reconfigurable Templates for Robust Vehicle Detection and Classification," *Applications of Computer Vision (WACV), 2012 IEEE Workshop on*, pp. 321-328, 2012.

[19] GTI Vehicle Image Database, Image Processing Group at UPM, <http://www.gti.ssr.upm.es/data>, 2011.

[20] C. M. Bishop, Pattern Recognition and Machine Learning. Springer Science+Business Media, 2006.