Salient Features Based on Visual Attention for Multi-View Vehicle Classification

Ana-Maria Cretu, Pierre Payeur and Robert Laganière
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada
[acretu, ppayeur, laganier] @ site.uottawa.ca

Abstract— **The continuous rise in the amount of vehicles in circulation brings an increasing need for automatically and efficiently recognizing vehicle categories for multiple applications such as optimizing available parking spaces, balancing ferry load, planning infrastructure and managing traffic, or servicing vehicles. This paper describes the design and implementation of a vehicle classification system using a set of images collected from 6 views. The proposed computational system combines human visual attention mechanisms to identify a set of salient discriminative features and a series of binary support vector machines to achieve fast automated classification. An average classification rate of 96% is achieved for 3 vehicle categories. An improvement to 99.13% is achieved by using additional measurement on the width and height of the vehicles.**

Keywords-visual attention; saliency; machine learning; support vector machines; vehicle classification.

I. INTRODUCTION

The increase in population and economic prosperity has led to a huge increase in the number of vehicles. This reality brings a growing need for automated and efficient classification techniques for different vehicle categories for a multitude of applications such as optimizing available parking lots and spaces, balancing ferry loads, managing traffic and planning infrastructure or servicing vehicles. Vision systems are relatively cheap, easy to install and configure and offer direct visual feedback and flexibility in mounting. They are therefore an appropriate sensing solution for vehicle classification. However, the issue of vehicle classification from images is not trivial. Due to the ever increasing number of vehicle models and sizes and the aesthetic similarities between them, the main problem is the identification of a set of representative and discriminative features that allow for the best possible classification of the vehicle type.

Taking inspiration from the significantly superior performance of humans to extract and interpret visual information, the exploitation of biological and psychological knowledge could contribute to improve artificial vision systems [1]. While still in their infancy, early cognitive vision-inspired algorithms for object recognition have already reached performance comparable to the best computer vision systems [2]. Of particular interest is the role of attention. Computational models of visual attention have been shown to significantly improve the speed of scene understanding and object recognition [3] by attending only the regions of interest and distributing the resources where they are required. As well, it was proven that attention systems are especially well suited to detect discriminative features and that the repeatability of salient regions is higher than the repeatability of non-salient regions provided by classical feature descriptors such as corners or SIFT keypoints [4]. Therefore such attention models are a promising direction of research for identifying discriminative features to be used as basis for classification.

This paper evaluates low-level features inspired from human visual attention for image-based vehicle classification. It proposes an original technique to identify the number of salient features to be considered for classification purposes. It also proposes and evaluates a viable design for a multiple camera system and the corresponding software solution for multi-view vehicle classification.

II. RELATED WORK

There are a few solutions for vehicle classification proposed in the literature. Yoshida *et. al* [5] use computer generated images of vehicles viewed from the top and their local features obtained by a corner detector to perform recognition of 4 vehicle types: sedan, wagon, minivan, hatchback. They obtain a limited 54% classification rate. Petrovic and Cootes [6] classify vehicles into 77 distinct classes (based on vehicle make and model) using the principle of locating, extracting and recognizing normalized structure samples taken from a reference image patch on the front of the vehicle and obtain about 93% recognition rates using only frontal views of vehicles. In [7], a neural network takes as input a reduced wavelet transform of the image of a vehicle and outputs one single element of the feature set that is considered relevant for classification purposes. An overall 83% classification rate is obtained for 5 vehicle types: motorcycle, car, bus, trailer 1 and trailer 2 type. Ji *et al.* [8] report performances between 93% and 95% when using a partial Gabor filter bank to represent sedan, van, hatchback, bus and truck vehicle categories. In [9] edge points and modified SIFT descriptors are combined to obtain a rich representation for vehicle object classes. Classification rates of 98% are obtained for car vs. minivan and 96% for car vs. taxi.

Regarding human visual attention, psychological studies have shown that there are two major categories of features that drive the deployment of attention: bottom-up features, derived

directly from the visual scene, and top-down features detected by cognitive factors such as knowledge, expectation, or current goals. Most computational implementations are based on bottom-up features, that can capture attention during free viewing conditions. A measure that has been shown to be particularly relevant is the local image saliency, which corresponds to the degree of conspicuity between that location and its surround. In other words, the responsible feature needs to be sufficiently discriminative with respect to the surroundings in order to guide the deployment of attention. In spite on the fact that opinions on features that guide human visual attention are still controversial [10], the intensity, color, orientation and motion are undoubted attributed that guide the deployment of attention. A full survey on attention-based computational systems is presented in [11].

Most of the proposed computational solutions have been tested mainly on indoor scenes or for a limited number of images. It is only in the latest years that attention-based computational systems started to be studied in practical applications dealing with real data. Frintrop and Jensfelt [4] use a sparse set of landmarks based on a biologically attention-based feature-selection strategy and active gaze control to achieve simultaneous localization and mapping of a robot circulating in an office environment and in an atrium area. In a similar manner, Siagian and Itti [12, 13] use salient features derived from attention together with context information to build a system for mobile robotic applications that can differentiate outdoor scenes from various sites on a campus [12] and for localization of a robot [13]. In Rasolzadeh *et al.* [14], a stereoscopic vision system framework identifies attention-based features that are then utilized for robotic object grasping. Rotenstein *et al.* [15] propose the use of mechanisms of visual attention to be integrated in a smart wheelchair for disabled children to help in visual search tasks.

In this work, salient features derived from the bottom-up computational model proposed by Itti *et al.* [16] are used as a basis to perform fast 3-category vehicle classification.

III. Computational System for Multi-view Vehicle Classification based on Vision-attention Salient Features

In order to achieve multi-view classification, it is considered that images from 6 views, of each vehicle are available as illustrated in Fig. 1a. Fig. 1 shows the 6 cameras and the images collected by each, namely straight front and rear view (camera 1 and 2), driver and passenger side profiles (camera 3 and 4) and front and rear three quarter view (camera 5 and 6), as well as examples from the dataset used for an initial experimentation that contains images of 122 cars from the following 3 categories [17]: sedan (Fig. 1b), sports car (Fig. 1c) and SUV (Fig. 1d). The size of each image is 99×150 pixels. For each of the images, the Itti *et al.*'s computational model of visual attention [16] is employed to identify a feature set, containing a predetermined number of features for each view and therefore for each camera.

The selection of the numbers of features is performed separately for the different views to cope with the differences in the number of salient features that might occur from a view to the other (e.g. more features might be required to discriminate vehicles from a view than from another view).

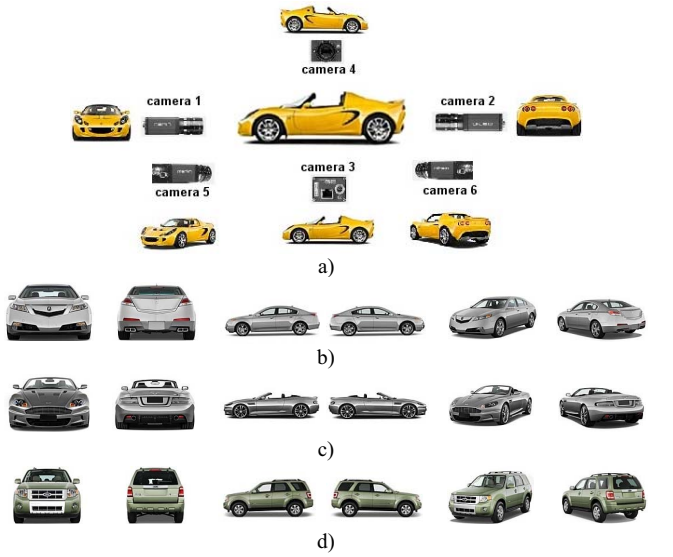


Figure 1. Multi-view vehicle classification: a) camera positioning and examples of vehicle categories in the dataset: b) sedan, c) sports car, d) SUV.

The feature sets, denoted Feature Set 1 to Feature Set 6 in Fig. 2 can therefore have different sizes.

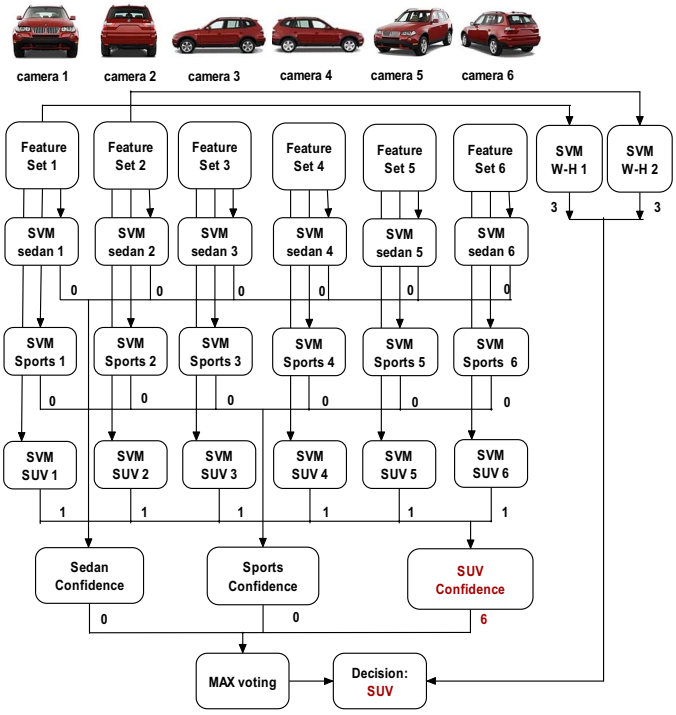


Figure 2. Multi-view vehicle classification: information flowchart.

The automated selection for the number of salient features to represent each view will be discussed in further detail in the next sections. The set of features is then transformed into a vector for each image. A set of 6 support vector machine (SVM) classifiers, one per each view of a given category, is

trained to perform a binary classification of vectors representing images of vehicles coming from a given camera, as illustrated in Fig. 2. This implies that the overall number of SVM classifiers equals 6 times the number of categories to be classified by the system (e.g. 18 in this work).

The output result of each classifier is a 1 if the classifier recognizes the vehicle in the image, from a given viewpoint, as belonging to the category that the classifier has learnt and 0 otherwise. For example a SVM that has been trained for the SUV class (e.g. SVM SUV 1) and that recognizes a vehicle in a test image coming from a given camera (e.g. camera 1) as a SUV, will prompt a 1 at the output. The results of the 6 classifiers representing the 6 available viewpoints are composed into what is called a confidence measure by adding the decision of all 6 classifiers for a certain category from the different views. The minimum confidence is 0, when none of the classifiers identifies the vehicle in the image as belonging to the category that the respective classifier has been trained with, and the highest is 6 when all the classifiers recognize the vehicle in the test image as belonging to a certain category. Such confidence measures are built for all the categories of vehicles studied when a certain test image is presented to the system. For example in Fig. 2, the sedan classifiers do not recognize the vehicle being a sedan and therefore the confidence measure is 0, while all the SUV classifiers recognize the vehicle as being a SUV. In order to provide the final decision, a MAX voting is performed on the resulting confidence measures. The vehicle in the test image is recognized as belonging to the category that provided the highest confidence measure. In Fig. 2, the highest confidence comes from the SUV and therefore the vehicle is classified (correctly) as a SUV.

When no decision can be produced because 2 or more categories provided the same confidence measure, an additional set of two SVM classifiers (denoted SVM W-H 1 and SVM W-H 2 in Fig. 2), is used to provide a decision based on the width and height of the vehicle derived from the front view (SVM W-H 1) and the rear view (SVM W-H 2) respectively. In this case, the classifier receives at the input a vector composed of the width and height of a vehicle in each image, computed based on the feature set, as it will be described in the next sections, and maps it to a corresponding category (1 for sedan, 2 for sports, 3 for SUV). The decision produced by these SVMs is used only for those cases where an ambiguous decision is reached by the salient feature SVMs.

IV. SALIENT FEATURE EXTRACTION

A. Extraction of Visual-Attention Inspired Salient Features

The main idea behind the bottom-up computational systems proposed in the literature in general, and for Itti *et al.*'s system in particular, is to compute several features derived from a color image provided as input and fuse their saliencies into a representation called saliency map [11, 16]. Initially, one or several image pyramids are created from the input image to enable the computation at different scales. Several features are then computed in parallel and feature-dependent saliencies are computed for each channel. Itti's computational attention model considers as features the intensity ($I = (R+G+B)/3$ where R, G and B are the red, green and blue color channels respectively), color (color maps are represented by the RG and BY color opponency) and orientation (local orientation information is obtained from the intensity image I using oriented Gabor pyramids of different scales and different preferred orientations). Center-surround operations, modeled as a difference between fine and coarse scales, are applied on all features. Each set of features is stored in feature dependent saliency maps, called conspicuity maps in form of grayscale images where the intensity of each pixel is proportional to its saliency. After normalization, these maps are summed up linearly in the final saliency map. The full implementation details are available in [16].

This computational attention model is employed in the context of this work to detect the salient features in each of the images in the dataset. The model described in [16] with 4 orientation channels (0, 45, 90, and 135 degrees), a small blur radius of 0.01 and a single center scale (instead of the 9 in the original model) is used to compute the saliency map, *SM*. To remove the distortions due to the symmetrical Gabor filtering, the map is resized at the original size of the input image. Examples of saliency maps, *SM,* are illustrated in Fig. 3 for the sedan in Fig. 1b. The images are presented as negatives to better visualize the results by showing the areas of highest saliency with darker shades.

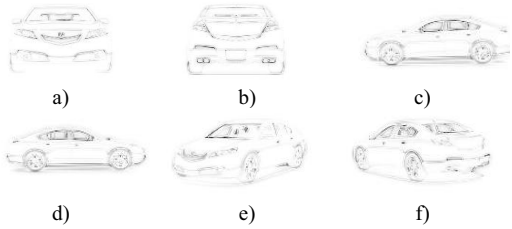


Figure 3. Saliency maps for the sedan in Fig. 1b obtained using Itti *et al.*'s visual attention computational model: straight a) front and b) rear, c) driver side profile, d) passenger side profile, e) front three quarter view and f) rear three quarter view.

B. Selection of the Number of Features for Classification

In order to select automatically the number of salient features to be used for classification of all the images in a given dataset, a saliency threshold is defined as the ratio between the saliency contained in the most salient m points of an image as computed from the saliency map and the saliency of the whole image:

$$s_T = \sum_{i=1}^{m} s_i \Big/ \sum_{j=1}^{n} s_j \tag{1}$$

where

$$s \in S,\ S = \left\{ s_k \,\middle|\, s_k = SM(x_k, y_k),\ s_k > s_{k+1},\ k = 1..n \right\} \tag{2}$$

and m is the number of salient pixels to be considered, n is the total number of pixels in SM with SM being the saliency map as obtained in section IV.A, with $m < n$, and S is a list in decreasing order of saliency of all the pixels in SM from the most salient to the least salient.

The number of salient features m to be used in the classification is selected such that all (at least 99.8% of) the

images in a dataset reach a saliency threshold s_T of at least 0.5. In other words, for all the images, the set of selected salient features represents at least 50% of the whole image saliency. To achieve this, values for m are increased gradually, initially with a step of 100. In each step, the first m values are selected from S. Eq. (1) is then used to compute the saliency threshold for each value of m for all the images in the dataset (in the context of this work, the images collected from a given viewpoint). The percentage of the number of images that have their saliency threshold s_T larger than 0.5 with respect to the number of all images in the dataset is computed for each value of m. The procedure is repeated for larger values of m until the threshold of 95% is reached for the percentage computed. The step size for m is then decreased to 50 for a better tuning. The computation is stopped when for a given value of m, all the images (at least 99.8%) from the dataset have their saliency threshold s_T at least 0.5. This value of m represents the optimal number of features to be selected for a given dataset and the set of salient features, S_m, is the set of the first m elements in the set S defined in eq. (2).

Fig. 4 illustrates the number of salient features for the three quarter front side view dataset (e.g. images from camera 5). It can be observed that for 900 features, the percentage of images with the saliency threshold larger than 0.5 computed using eq. (1) is about 23% and that at 1300 features more than 95% of images have the threshold larger than 0.5. For fine tuning, the step size is decreased to 50 and the computation is stopped when all images (99.8%) from the dataset contain at least 50% of their whole saliency, that is for m=1450 salient feature points in this case.

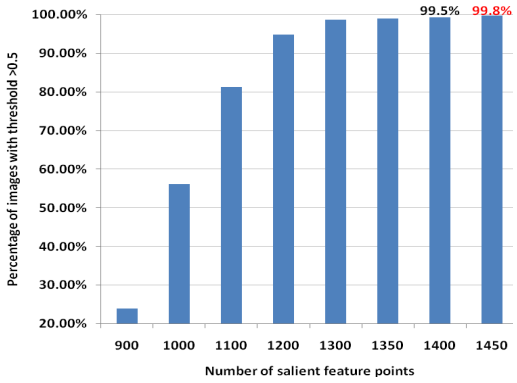


Figure 4. Saliency thresholding to identify the optimal number of salient features for classification for the three quarter front view dataset.

The same procedure is repeated for all the categories of vehicles being viewed from a given direction (e.g. front view, rear view, etc.) since some views might contain a higher number of discriminative features than others. This fine tuning for each view is possible in the context of this work because it is known that all the images from a given viewpoint are provided by a given camera and therefore the number of features identified can be used for all images coming from that camera.

The number of points identified with the above mentioned procedure is 1250 salient features points for the front view image dataset, 1450 for the rear view and three quarter front view datasets, 1650 points for the lateral profiles (driver and passenger sides) and 1350 for the three quarter back view dataset. After the number of salient features is identified for each view of a vehicle, all the selected salient features S_m are replaced with 1s in SM and all the rest of the points with 0 to build a limited m-feature saliency map, SM_l:

$$SM_l(x,y) = \begin{cases} 1, & if\ SM(x,y) \in S_m \\ 0, & otherwise \end{cases} \quad (3)$$

The limited saliency maps SM_l with black pixels representing 1s, obtained for the 6 views of the sedan are shown in Fig. 5. They produce a sketched shape of the vehicle which provides rich inputs to the SVMs for classification.

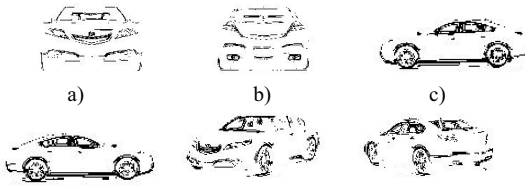


Figure 5. Salient regions used for classification for the sedan in Fig. 1b.

V. TRAINING, TESTING AND EVALUATION OF SVM CLASSIFICATION

A. SVM Classification Based on Salient Features

A binary SVM classifier is trained to recognize a specific type of vehicle from a given view against all the other types of vehicles viewed from the same direction. In order to build the training and the test sets, the limited m-feature saliency map SM_l is downsampled to one third of the size (e.g. 30×50 pixels) and transformed into a vector that is used as input in the classifier. The target set is built for each category by assigning 1 to all the vehicles representing that category (positive examples) and 0 to all vehicles belonging to other categories (negative examples). On the average, there about 40 positive examples and 82 negative examples for each classifier. A 5-fold cross-validation procedure is used for building the classifier's training and testing sets respectively. In the first fold, 80% of randomly selected input vectors built from all the images representing vehicles from a certain view and their corresponding targets are initially used for training and the rest of 20% for testing. In the next folds, the data used for testing is moved back into the training set and another 20% is selected for testing in order to ensure better performance evaluation.

The set of vectors is classified using least-squares SVMs (LSSVM) [18] for each given viewpoint and the results are added to compute the confidence for a given category, as illustrated in Fig. 2. A LSSVM classifier with a Gaussian RBF kernel, the regularization parameter γ=10 and the squared bandwidth σ^2=0.4 is used. The training for the 122 vehicles from a given viewpoint takes about 0.09 s. The testing per test image takes on average 0.03s. The classification rate is based on the confidence measure and computed as the number of test images correctly classified over the number of test images and

averaged over the 5 folds. For each view, the classifier's performance is reported in Table I. It can be observed that the average classification rate for every classifier is over 90%. The reason to use binary classifiers instead of a single multiclass classifier is that, during the experiments performed, the latter demonstrated lower performance, that is classification rates of about 75% were obtained.

TABLE I. AVERAGE CLASSIFICATION RATES PER VIEW

	Sedan	Sports car	SUV	Per view
View 1	86.9%	96.5%	93.9%	**92.5%**
View 2	88.7%	93.0%	94.8%	**92.0%**
View 3	90.4%	93.0%	89.6%	**90.4%**
View 4	90.4%	93.0%	93.0%	**92.2%**
View 5	91.3%	93.0%	96.5%	**93.6%**
View 6	87.8%	96.5%	98.3%	**94.2%**

To demonstrate that the correct number of salient features are selected, experiments are performed to show the increase in per view classification rate. Fig. 6 illustrates the change in the classification rate with the number of salient features for the three quarter front side view dataset. It can be observed that the classification rate increases up to 1450 salient features after which it starts to decrease. This conclusion is similar with the one derived in Fig. 4 where for 1450 saliency features, all images in the dataset contain at least 50% of their saliency. Similar results are obtained for all the other views, showing that the method for selecting the number of features to represent the vehicles is adequate.

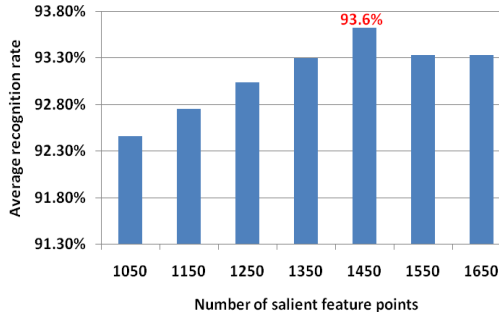


Figure 6. Average classification rates for all vehicle categories for increasing number of saliency points for the three-quarter front view.

To compute the final decision of the system, a MAX voting is performed on the results provided by the confidence scores of the multiple view classifiers for all the vehicle categories. The category that corresponds to the highest confidence classifier is the winner. An average result of 96% is achieved by considering all the views available for the 3 vehicle categories. The improvement over the lower per view rates illustrated in Table I is obtained by taking advantage of decision based on multiple views.

There is a limited set of situations when the system does not perform as expected, as illustrated in Fig. 7. Fig. 7a shows a sedan misclassified as an SUV, and cases where the system is not able to provide a decision based on the salient features are shown in Fig. 7b-7d. For example, for the sedan in Fig, 7b the confidence measures computed are 0 for all the 3 categories, and for the sedan in Fig. 7c the measures of confidence provided by sedan equals the one provided by the sports cars. Finally, for the SUV in Fig. 7d, two of the confidence measures provided by the sedan and SUV are equal.

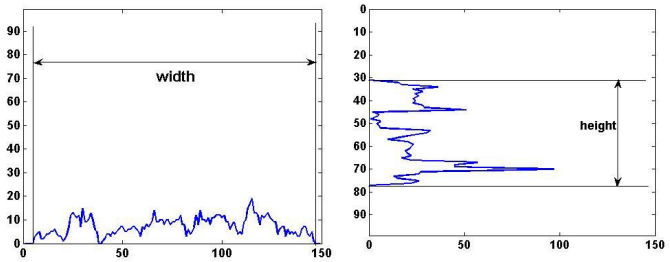
a) b) c) d)

Figure 7. Sample cases where no decision is produced based on salient features.

To cope with these situations, an additional set of two SVM classifiers is used to provide a decision based on the width and height of the vehicle derived from the front view and the rear view respectively.

B. *SVM Classification Based on Width and Height*

The saliency map obtained in section IV.A can be used at the same time to compute the width and height of the vehicle because it delimitates the contour of each vehicle. In order to compute the width and height, the image representing the saliency map, *SM*, is initially converted to binary and denoted SM_{bin}. The vertical, *W*, and horizontal, *H*, projection vectors are built by summing all the columns of SM_{bin} to obtain *W* and all the rows of SM_{bin} to obtain *H*. The projection vectors are shown in Fig. 8.



a) b)

Figure 8. a) Width and b) height projection vectors used to compute the vehicle's width and height.

The initial width *w* is set to the width of the initial image that is also equal to the one of the saliency map. Assuming a clean background of SM_{bin}, the *W* vector is then searched starting from the left until a value different from 0 is identified (left vertical line in Fig. 8a). Each time a value of 0 (empty background) is found, the value of the *w* is decreased by 1. When the first value different from 0 is encountered, the search from the left side is stopped. The same procedure is used from the right side by decreasing the value of *w* until a value different than 0 (right vertical line in Fig. 8a) is found. In this way *w* will define the width of the car.

A similar top and down search is performed on the *H* vector to compute the height *h* of the vehicle in the image. The vectors consisting of width and height [*w h*] for each image in view 1 and view 2 respectively are associated with the corresponding vehicle type (1 for sedan, 2 for sports, 3 for SUV) and classified with SVM W-H 1 for the front view coming from camera 1 and SVM W-H 2 for the rear view coming from camera 2 respectively, as shown in Fig. 2. The best views for computing width and height are the front and

rear views. The lateral profiles can be misleading because of a large variance in length within each class on one side and because length alone is not discriminative enough to allow the distinction between different categories on the other side.

As in the previous case, a 5-fold cross validation procedure is applied for training and testing of the SVMs. A simple SVM with a Gaussian RBF kernel and the regularization parameter γ=10 and the squared bandwidth σ^2=0.4 is used for each view to learn and then provide categories for the testing set. The average classification rate for all vehicle categories based on width and height information only is 93%. Correct classification cases are considered as those where the results provided by the two views are the same.

C. *Combined Classification Based on Salient Features and Width-Height Information*

The results when only the salient features are used for the 6 views as in section V.A and when the results from the width-height SVMs in section V.B are added are shown in Table II.

TABLE II. AVERAGE CLASSIFICATION RATES FOR ALL CATEGORIES

Salient features	**96%**
Salient features + W-H information	**99.13%**

TABLE III. STATISTICAL EVALUATION PER CATEGORY

	Sedan		**Sports car**		**SUV**	
	Sal. feat.	**Sal.+ W-H**	**Sal. feat.**	**Sal.+ W-H**	**Sal. feat.**	**Sal.+ W-H**
precision	0.68	0.97	1.00	1.00	0.96	0.98
recall	0.90	0.97	1.00	1.00	1.00	1.00
accuracy	0.82	0.98	1.00	1.00	0.98	0.99
F_1-score	0.77	0.97	1.00	1.00	0.98	0.99

The average recognition rate is significantly higher when width and height information is used as well, reaching 99.13%. The statistical evaluation of the results in terms of precision, recall, accuracy and F_1-score, reported as an average over the 5-folds, is comparatively presented in Table III. A perfect classification is denoted by a value of F_1-score equal to 1. The values in Table III are very close to this value. The same improvement can be noticed when the width and height information is used in addition to the salient features for the sedan category. All the ambiguous cases illustrated in Fig. 7b-d are correctly classified by the joint decision of SVM W-H 1 and SVM W-H 2. The only misclassified vehicle that remains using the proposed combined classification is the one illustrated in Fig. 7a.

It can be seen that the proposed classification technique obtains better results than the best solutions found in the literature [6, 8, 9] for a case where the discrimination between the classes is less clear. For example it is easier to differentiate between a sedan and a truck [7-9] than between a sedan, a sports car and an SUV.

VI. CONCLUSION

The work in this paper demonstrates that biologically-inspired features derived from visual attention combined with series of binary support vector machines can achieve fast classification, with exceptional recognition rates for the task of multi-view vehicle classification. As future work, the system will be expanded to include more classes of vehicles into the categorization task and experiments will be performed to reduce the number of cameras used by identifying the viewpoints that make the most important contribution to the classification, in order to decrease the implementation cost while maintaining the high performance.

REFERENCES

[1] T. C. Kietzmann, S. Lange and M. Riedmiller, "Computational Object Recognition: A Biologically Motivated Approach", *Biological Cybernetics*, vol. 100, pp. 59-79, 2009.

[2] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust Object Recognition with Cortex-Like Mechanisms", *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 29, no. 3, pp. 411-426, 2007.

[3] D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch, "Attentional Selection for Object Recognition – a Gentle Way", *Biologically Motivated Computer Vision*, Lecture Notes in Computer Science, Springer, vol. 2525, pp. 472-479, 2002.

[4] S. Frintrop, and P. Jensfelt, "Attentional Landmarks and Active Gaze Control for Visual SLAM", *IEEE Trans. Robotics*, vol. 24, no. 5, pp. 1054-1065, 2008.

[5] T. Yoshida, S. Mohottala, M. Kagesawa and K. Ikeuchi, "Vehicle Classification System with Local-Feature Based Algorithm Using CG Model Images", *IEICE Trans.*, vol. E00A, no. 12, pp. 1-8, 2002.

[6] V.S. Petrovic, and T.F. Cootes, "Analysis of Features for Rigid Structure Vehicle Type Recognition", *British Machine Vision Conf.*, Kingston, pp. 587-596, 2004.

[7] N. Xiong, J. He, J. H. Park, D. Cooley, and Y. Li, "A Neural Network Based Vehicle Classification System for Pervasive Smart Road Security", *Universal Computer Science,* vol. 15, no. 5, pp. 1119-1142, 2009.

[8] P. Ji, L. Jin, and X. Li, "Vision-based Vehicle Type Classification Using Partial Gabor Filter Bank", *Proc. IEEE Int. Conf. Automation and Logistics*, Jinan, China, pp. 1037-1040, 2007.

[9] X. Ma, W. Eric, and L. Grimson, "Edge-based rich representation for vehicle classification", *Int. Conf. Computer Vision*, vol. 2, pp. 1185-1192, 2005.

[10] J.M. Wolfe, and T.S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?", *Nature Reviews Neuroscience*, vol. 5, pp. 1-7, 2004.

[11] S. Frintrop, E. Rome and H. Christensen, "Computational Visual Attention Systems and their Cognitive Foundations: A Survey", *ACM Trans. Applied Perception*, vol. 7, no. 11, pp. 1-46, 2010.

[12] C. Siagian and L. Itti, "Rapid Biologically-Inspired Scene Classification Using Features Shared with Visual Attention", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 300- 312, 2007.

[13] C. Siagian, and L. Itti, "Biologically Inspired Mobile Robot Vision Localization", *IEEE Trans. Robotics*, vol. 25, no. 4, pp. 861-873, 2009.

[14] B. Rasolzadeh, M. Björkman, K. Huebner and D. Kragic, "An Active Vision System for Detecting, Fixating and Manipulating Objects in the Real World", *Int. Journal of Robotics Research*, vol. 29, issue 2-3, pp. 133-154, 2010.

[15] A. M. Rotenstein, A. Andreopoulos, E. Fazl, D. Jacob, M. Robinson, K. Shubina, Y. Zhu, and J. K. Tsotsos, "Towards the Dream of an Intelligent, Visually-Guided Wheelchair", *Proc. 2nd Int'l Conf. on Technology and Aging*, Toronto, Canada, 2007.

[16] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[17] Available online, www.izmostock.com/.

[18] Least-Squares Support Vector Machines (LSSVM) Matlab Toolbox, available online, http://www.esat.kuleuven.be/sista/lssvmlab/.