www.s2is.org

International Journal on Smart Sensing and Intelligent Systems

ISSN 1178-5608

BIOLOGICALLY-INSPIRED VISUAL ATTENTION FEATURES FOR A VEHICLE CLASSIFICATION TASK

A.-M. Cretu, and P. Payeur

School of Electrical Engineering and Computer Science

University of Ottawa, 800 King Edward

Ottawa, Canada

Emails: acretu@site.uottawa.ca, ppayeur@site.uottawa.ca

Submitted: June 15, 2011 Accepted: August 2, 2011 Published: September 1, 2011

Abstract- The continuous rise in the number 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 loads, planning infrastructure and managing traffic, or servicing vehicles. This paper explores the use of human visual attention mechanisms to identify a set of features that allows for fast automated classification of vehicles based on images taken from 6 viewpoints. Salient visual features classified with a series of binary support vector machines and complemented by a dissimilarity score achieve average classification rates between 94% and 97.3% for five-category vehicle classification depending on the combination of viewpoints used. The viewpoints that make the most important contribution to the classification are identified in order to decrease the implementation cost. The evaluation of performance against other feature descriptors and various

approaches for vehicle classification shows that the proposed solution obtains results comparable to the best ones reported in the literature.

Index terms*:* Visual attention, saliency, machine learning, support vector machines, vehicle classification, dissimilarity, feature extraction.

I. INTRODUCTION

The growth of 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.

Humans show a significantly superior performance in extracting and interpreting visual information to any state-of-the-art artificial vision model. Therefore the exploitation of biological and psychological knowledge derived from human visual mechanisms can contribute to the improvement of computational vision systems [1]. Early vision-inspired algorithms for object recognition, in spite of their relative novelty, have already reached performance comparable to the best computer vision systems [2] and biologically-inspired visual features have been successfully applied for different tasks in image processing [3]. Computational models of visual attention have been shown to significantly improve the speed of scene understanding and object recognition [4] by attending only the regions of interest and distributing the resources where they are required. Moreover, recent research showed that attention systems are well suited to detect more repeatable discriminative features than other classical feature descriptors such as corners or SIFT key points [5].

This paper uses salient features derived from the low-level, bottom-up visual attention and originally combines them with a series of support vector machines and a dissimilarity score to achieve fast automated classification of vehicles from 5 categories based on images taken from different viewpoints. The organization of the paper is as follows: Section II discusses related work on computational vision and visual attention on one side and on vehicle classification on the other side. Section III summarizes the proposed solution for multi-view vehicle classification. Further details on the extraction of salient features based on visual attention are given in Section IV. Section V presents the training and evaluation of the classification performance. Experimental results are presented for each viewpoint, combination of viewpoints and category of vehicle.

II. RELATED WORK

Most computational implementations of human visual attention are based on bottom-up features, derived directly from the visual scene, and that can capture attention during free viewing conditions. A full survey on attention-based computational systems is presented in [6]. In order to guide the deployment of attention, the responsible features need to be salient or in other words sufficiently discriminative with respect to their surroundings. The intensity, color, orientation and motion are undoubted attributes that guide human visual attention and are used in almost all current computational visual attention models. Most of the proposed computational attention models have been tested for a limited number of images in simplistic scenarios. There are only a few attention-based computational systems that have been used in practical applications dealing with real data. In [5], a sparse set of landmarks based on a biologically inspired attention-based feature selection strategy and active gaze control are used to achieve simultaneous localization and mapping of a robot. In a similar manner, Siagian and Itti [7, 8] employ salient features derived from attention and context information to build a system for mobile robotic applications that can differentiate outdoor scenes [7] and that can help in the localization of a robot [8]. Rasolzadeh *et al.* [9] propose a stereoscopic vision framework that uses attention-based features for robotic object grasping. Mechanisms of visual attention are integrated in [10] in a smart wheelchair application to help visual search tasks.

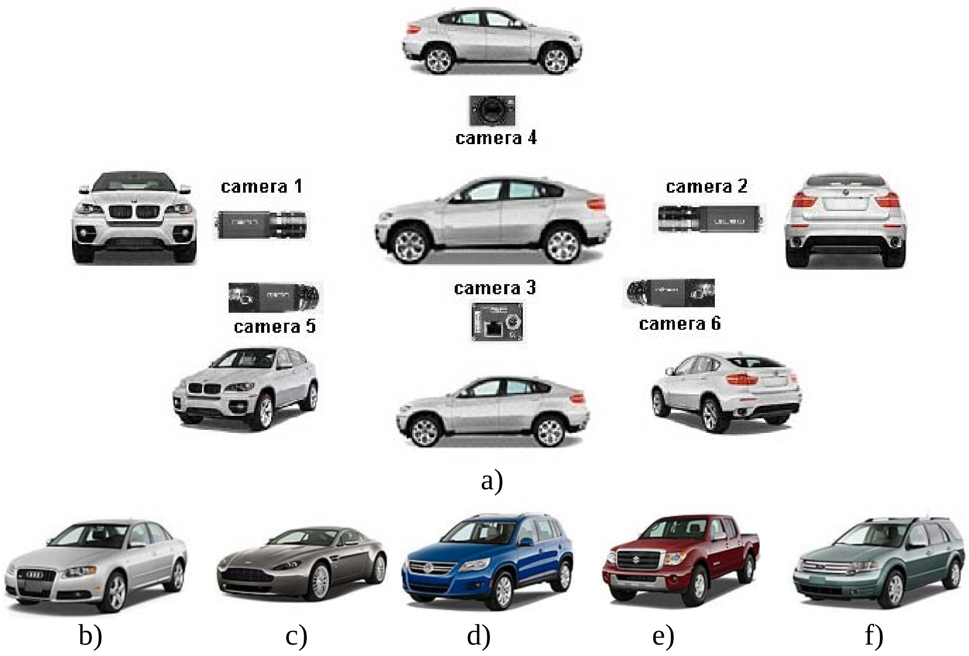
Regarding the topic of vehicle classification, there are several solutions proposed in the literature. In [11], a neural network takes as input a reduced wavelet transform of the image of a vehicle and outputs a single element of the feature set that is considered relevant for classification purposes.

An overall 83% classification rate is obtained for 5 vehicle types: motorcycle, car, bus, trailer type 1 and trailer type 2. The cascade of classifiers based on Adaboost proposed in [12] to categorize from front and rear views of vehicles belonging to car, van or truck categories achieves an overall 93.51% classification rate. Ji *et al.* [13] report classification performances between 93% and 95% when using a partial Gabor filter bank to represent sedan, van, hatchback, bus and truck vehicle categories. A maximum 92% classification rate is obtained by Kazemi *et al.* [14] for the classification of five vehicle types, namely Peugeot 206, Peugeout 405, Pride, Renault 5 and Peykan, using fast curvelet transform features and a k-nearest-neighbor classifier. In [15], 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. Lee [16] employs a neural network trained with texture descriptors (contrast, homogeneity, entropy and momentum of the gray level co-occurrence matrix) derived from the front view images to classify 24 types of Korean vehicles and obtains 94% recognition rate. Yoshida *et al.* [17] obtain a limited 54% classification rate for the recognition of 4 vehicle types: sedan, wagon, minivan, and hatchback, when using computer generated images of vehicles viewed from the top and their local features obtained by a corner detector. In [18], a hierarchical classification technique is proposed to distinguish between seven vehicle types: sedan, van, pickup, truck, van truck, bus, and trailer, starting with an initial coarse classification (large or small vehicle) and followed by a fine classification (based on length, aspect ratio, and compactness ratio). An overall recognition rate of 91.35% is achieved. Petrovic and Cootes [19] classify vehicles based on make and model into 77 distinct classes by 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. Dalka and Czyzewski [20] evaluate a combination of several descriptors such as statistical moments, speeded-up robust features from luminance images and image descriptors based on filtering with a bank of Gabor filters and several classifiers, such as neural networks, decision trees, nearest neighbors and random forests for three-category vehicle classification (e.g. sedans, vans and trucks) and obtain a maximum of 95% correct classification rate. In [21] a multiclass vehicle type (make and model) identification is proposed based on oriented contour points obtained from several grayscale frontal vehicle images, and a nearest-neighbor process is used to determine the vehicle type. A classification rate of 90% is reported.

In the current work, visual attention features derived from a bottom-up computational attention model [22] are used as a basis to perform fast automated 5-category vehicle classification from 6 viewpoints. Different viewpoint combinations are evaluated to identify those that provide the best results while allowing the use of fewer cameras than the whole set of 6.

III. VEHICLE CLASSIFICATION BASED ON VISUAL-ATTENTION FEATURES

It is initially considered that images from 6 views of each vehicle are available as illustrated in Fig. 1a, namely straight front and rear views (camera 1 and 2), driver and passenger side profiles (camera 3 and 4) and front and rear three quarter views (camera 5 and 6). Fig. 1b to Fig. 1f show examples of images from the dataset used for experimentation that contains images of 155 vehicles from the following 5 categories [23]: sedan (Fig. 1b), sports car (Fig. 1c), SUV (Fig. 1d), pickup truck (Fig. 1e) and wagon (Fig. 1f). The size of each image is 99×150 pixels.



a)

b) c) d) e) f)

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

The computational model of visual attention of Itti *et al.* [22] is employed to identify a feature set, containing a predetermined number of features for each view and therefore for each camera, as will be detailed in section IV. The feature set is built based on the saliency map, *SM*, obtained by the attention model, as shown in Fig. 2.

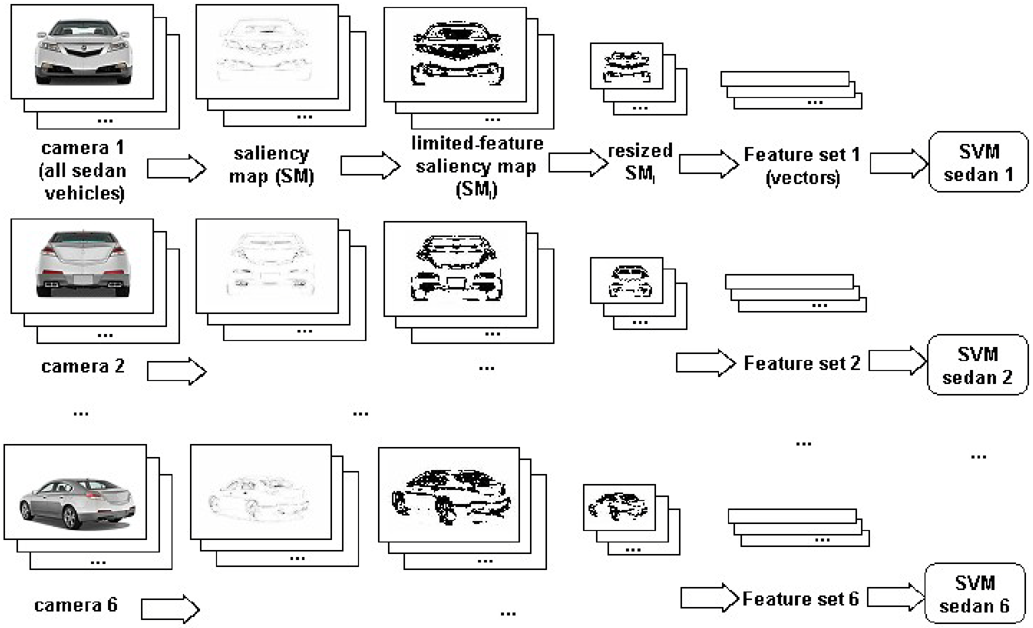


Figure 2. Multi-view vehicle classification: feature sets

The limited-feature saliency map is then downsampled and transformed into a feature set vector for each image by concatening its lines. Each image presented to the system needs to follow these preprocessing steps prior to its classification. A set of 6 support vector machine (SVM) classifiers, one per each view of a given category (e.g. sedan in Fig. 2), is trained to perform a binary classification of the feature vectors derived from images of vehicles coming from a given camera. The overall number of SVM classifiers equals 6 times the number of categories to be classified by the system (e.g. 30 for 5-class classification and 6 views). Binary classifiers are chosen instead of multi-class classification because our experimentation with both approaches revealed that binary classifiers obtain better performance on the dataset and for the task considered, as it will be discussed in section V.b. Each classifier outputs a 1 if it recognizes the vehicle in the image from a given viewpoint, represented as a feature vector built from the downsampled limited-feature saliency map, to be belonging to the category that the classifier has learnt and 0 otherwise. An example is illustrated in Fig. 3, which continues the information flow from Fig. 2.

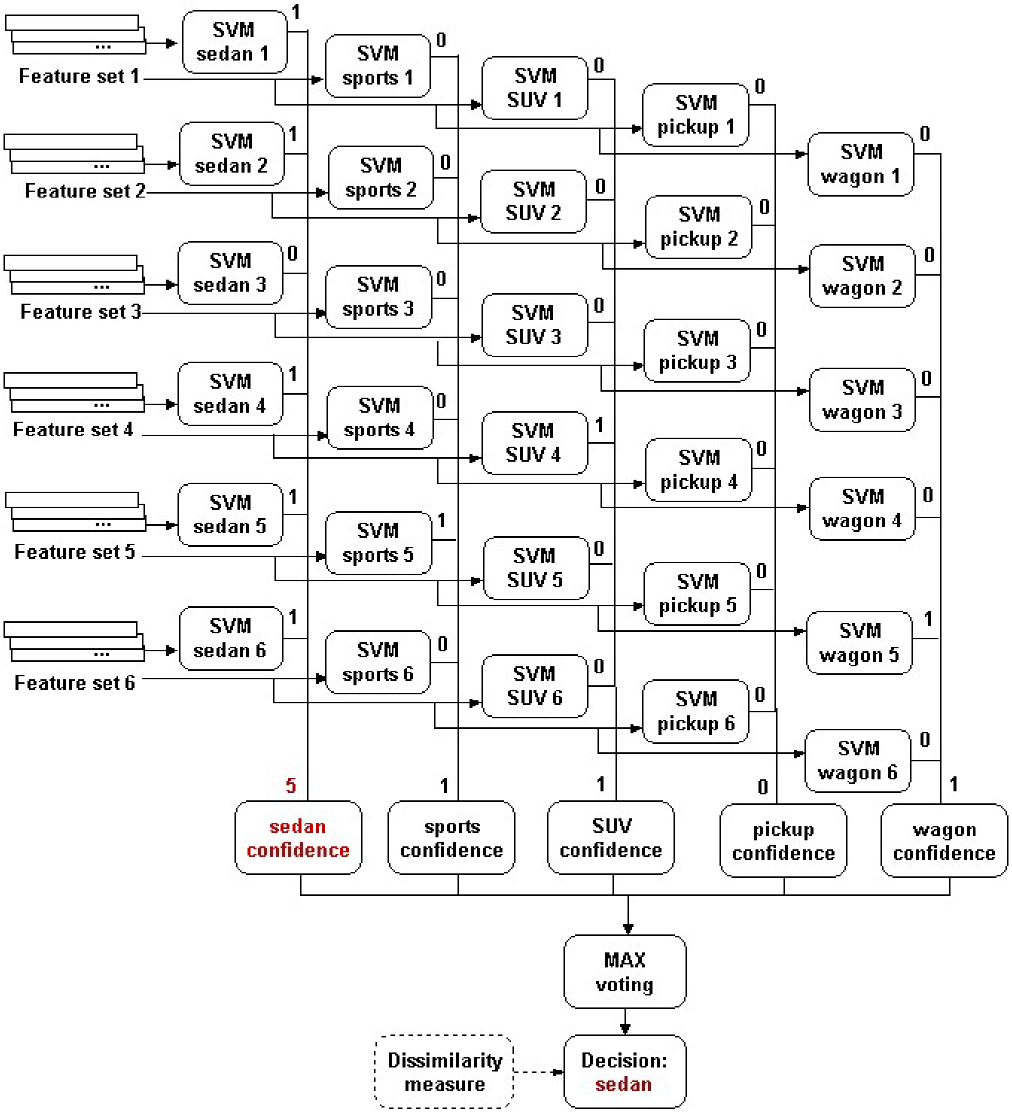


Figure 3. Multi-view vehicle classification: binary SVM classification

A SVM trained for the sedan class (e.g. SVM sedan 1 in Fig. 3) that recognizes a vehicle in a test image coming from a given camera (e.g. camera 1) as being a sedan, will output a 1. The results of the classifiers representing each viewpoint are composed into what is called a confidence measure by adding the decision of all classifiers for a certain category from the different views. When none of the classifiers identifies the vehicle in the image as belonging to the category that the respective classifier has been trained for, the confidence is 0. When all the classifiers (if all views are used) recognize the vehicle in the test image as belonging to a certain category, the confidence is 6. Such confidence measures are computed for every category of vehicles when a certain test image is presented at the input. For example in Fig. 3, most of the SUV classifiers do not recognize the vehicle to be an SUV and therefore the SUV confidence measure is 1, while most of the sedan classifiers recognize the vehicle as being a sedan and output a 1, leading to a confidence of 5. To compute the final decision, a MAX voting is performed on the resulting confidence measures. Therefore a vehicle in an image is recognized as belonging to the category that provided the highest confidence measure. In Fig. 3, the highest confidence comes from the sedan and therefore the vehicle is classified (correctly) as a sedan. When no decision can be produced by the system because two or more categories result in the same confidence measure, an additional procedure based on dissimilarity, illustrated in Fig. 4, is employed to help make the decision.

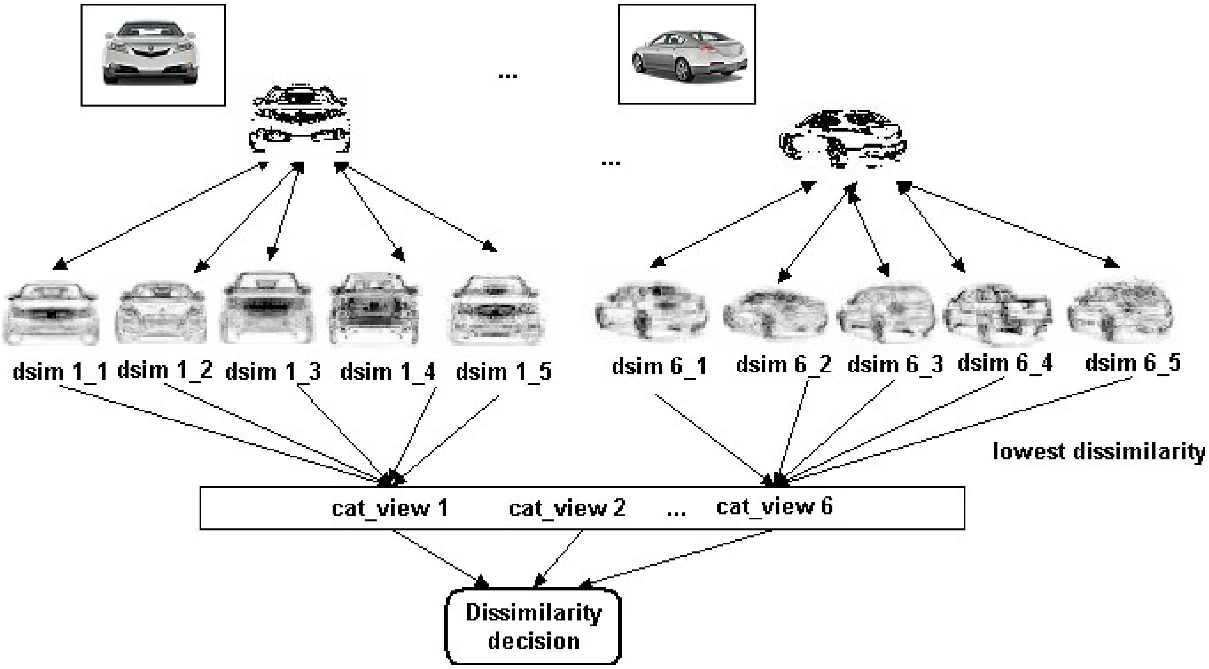


Figure 4. Multi-view vehicle classification: computation of dissimilarity

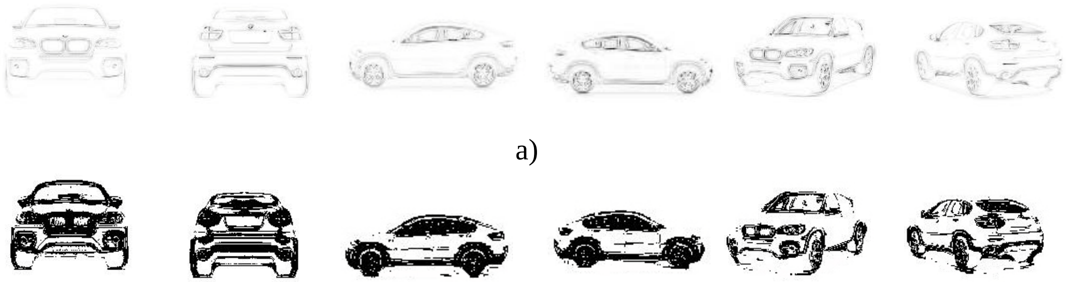
A score of image dissimilarity is computed between the features of a vehicle on which the decision cannot be produced (viewed from each viewpoint) and an average feature model for each vehicle category (also as viewed from each viewpoint) as further detailed in section V.c. These average models are denoted dsim 1_1 to dsim 6_5 in Fig. 4. From each viewpoint, the category that is associated with the lowest dissimilarity is considered the winner and is denoted cat_view_1 to cat_view_6. The elements in the cat_view vector can have a value of 1 (sedan), 2 (sports), 3 (SUV), 4 (pickup truck) or 5 (wagon). To compute the final decision, the category that occurs most often in the cat_view vector is considered the winner. For each cat_view element, the value of the lowest dissimilarity score that led to its selection as a winner (one of the dsim 1_x to dsim 6_x) is also saved. This value can be used to discriminate between categories when two or more categories occur in equal number in the cat_view vector. When no decision can be made based only on the category that occurs most often, the category that has the lowest dissimilarity score is considered the winner. Therefore it is guaranteed that a decision will always be produced, unlike when only binary SVMs are used. However, this dissimilarity score is only used when needed and as a complementary measure to the proposed series of SVMs because, on its own, it produces lower classification rates.

IV. SALIENT FEATURES EXTRACTION

a. Extraction of visual-attention inspired salient features

The computational model of attention of Itti *et al.* [22] computes several features derived from an image provided as input and fuses their saliencies into a representation called saliency map. One or several image pyramids (3 in the context of this work) are created from the input image in order to enable the computation at different scales. Several features are then computed in parallel, namely 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-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: 0, 45, 90, and 135 degrees in the context of this work). 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 [22]. This model is employed in the context of this work to detect the salient features in each of the images in the dataset. The saliency map, *SM*, obtained is shown for a sedan in Fig. 2 and for an SUV in Fig. 5a.

The images are presented as negatives (e.g. 1-*SM*) to better visualize the results by showing the areas of highest saliency with darker shades.



a)

b)

Figure 5. Saliency maps for the SUV in Fig. 1a obtained using Itti *et al.*'s visual attention computational model and b) selected saliency feature points used for classification

b. Selection of the number of features for classification

The number of salient feature points to be used as a basis for classification is identified separately for all 5 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 is possible in the context of this work because it is assumed that all the images from a given viewpoint are provided by a given static camera and therefore the number of features identified can be used for all images coming from that same camera. It allows for the proper selection of the number of salient feature points that ensures the best performance for a given view.

To select the number of feature points from the saliency map to be used as a basis for classification, a set of test runs of the computational attention model described in Section IV.a are performed on each view for an increasing number of saliency points between 1000 and 5500 with a step of 100. These salient points, *s*, are selected in order from *S*, a list in decreasing order of saliency of all the pixels in *SM* from the most salient to the least salient.

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

The upper bound of 5500 points represents the totality of salient points in the saliency map, *SM*, computed as an average over all images in the dataset. The average recognition rate is computed over the 5 categories of vehicles for every number of salient points between 1000 and 5500. The number of salient points selected, *m*, corresponds to that number which supported the highest average recognition rate and all these points are stored in a list S_m (S_m contains the *m* most salient points in *S*). All the feature points in S_m are replaced with 1s in the saliency map, *SM*, and all the other points are replaced with 0 to build a limited *m*-feature saliency map, SM_l:

$$SM_l(x,y) = \begin{cases} 1, & if \ SM(x,y) \in S_m \\ 0, & otherwise \end{cases} \tag{2}$$

The limited *m*-feature saliency map, SM_l, is downsampled to one third of its size (e.g. 30×50 pixels) to reduce the computational burden and transformed into a feature set vector that is used as input to the classifier.

V. TRAINING AND EVALUATION OF SVM CLASSIFICATION

a. SVM classification based on salient features

As explained in Section III, a binary SVM classifier is trained to recognize the category of a vehicle from a given viewpoint against all the other categories of vehicles viewed from the same direction. The target dataset is built by assigning 1 to all the vehicles representing that category (positive examples) and 0 to all vehicles belonging to other categories (negative examples). The input dataset, composed of feature set vectors obtained in Section IV.b, and the target dataset are split into training and testing data for the classifier using 5-fold cross validation. In the first fold, 80% of randomly selected input vectors built from all the images representing vehicles from a certain viewpoint and their corresponding targets are used for training and the rest of 20% for testing. In the next fold, another 20% is selected for testing and the old testing set is returned to training data. This process is repeated 5 times, that is until all input vectors have been considered in the testing set. The set of input vectors is classified using least-squares SVMs (LSSVM) [24] for each given viewpoint and the results are added to compute the confidence for a given category, as illustrated in Fig. 3. 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 155 vehicles from a given viewpoint takes about 0.09s per image. The testing per test image takes on average 0.03s on a Matlab platform.

The training–testing sequence is repeated for an increasing number of saliency feature points, to determine the appropriate number to be used for classification purposes, as described in Section IV.b. The identified number for each of the 6 views available over all 5 vehicle categories is reported along with their maximum and average classification rates in Table 1. These rates are computed as the number of test images correctly classified over the number of test images and averaged over the 5-folds. All cases where no decision is produced with the SVMs alone are considered at this stage classification errors. It can be observed that the totality of salient points in the saliency map (5500 for the dataset used during experimentation) is used for views 1 and 2, while a smaller number is sufficient to obtain the maximum classification rate for other viewpoints.

Table 1: Number of salient points, the maximum and the average classification rate per view over all categories

View	**1**	**2**	**3**	**4**	**5**	**6**
Number of salient points	5500	5500	4650	4650	4750	3750
Maximum classification rate	94.2%	94.5%	95.9%	95.1%	95.4%	94.9%
Average classification rate	93.0%	93.0%	94.4%	94.0%	95.2%	94.2%

The limited saliency maps, SM_l, obtained with the predetermined number of salient features in Table 1 and with black pixels representing 1s, are shown for a sedan in Fig. 2 and for an SUV in Fig. 5b. They produce some sort of a sketched shape of the vehicle which provides rich inputs to the SVMs for classification. The average classification rate for each viewpoint and the average over all viewpoints are reported in Table 2.

Table 2: Average classification rates per view and vehicle category (SVMs only)

View	**Sedan**	**Sports**	**SUV**	**Pickup**	**Wagon**
1	89.2%	97.8%	91.9%	97.8%	94.1%
2	86.5%	95.7%	96.2%	100%	94.1%
3	93.0%	96.8%	94.6%	98.9%	96.2%

4	93.0%	94.6%	93.5%	98.9%	96.2%
5	91.3%	96.2%	96.2%	97.8%	94.6%
6	94.1%	94.1%	94.1%	98.9%	93.5%
All views	**91.2%**	**95.8%**	**94.4%**	**98.7%**	**94.8%**

It can be observed that the average classification rate for each classifier for all views is over 91%. The proposed solution is compared with a classification based on several classical features such as SIFT key points [25], Harris corners [26], Difference-of-Gaussian (DoG) features (with σ_1=1 and σ_1=0.01) and Gabor features (with orientations 0, 45, 90, and 135 degrees). Results are reported in Table 3. The same dataset and the same set of binary SVMs are used for the classification of all the features and the computation of average classification rate for all views over the same 5-folds. The average classification rate achieved with the proposed solution using biologically-inspired visual attention saliency maps is reproduced from the last row of Table 2 as a basis for comparison.

Table 3: Comparison of the proposed solution with other feature detectors (average classification rate for all views)

	Sedan	**Sports**	**SUV**	**Pickup**	**Wagon**
SIFT key points	73.2%	79.1%	67.9%	80.0%	81.2%
Harris corners	76.5%	80.3%	67.9%	67.9%	81.0%
DoG features	88.2%	90.9%	89.6%	98.2%	93.5%
Gabor features	87.1%	91.5%	92.4%	98.5%	93.5%
Proposed solution	91.2%	95.8%	94.4%	98.7%	94.8%

It can be seen that SIFT key points and Harris corner features provide similar and relatively low average classification rates, and are therefore not discriminative enough for the task. The performance of DoG and Gabor features as a basis for classification is also similar, but overall the rates are lower than those obtained with the proposed solution.

b. Classification Based on Multiple Views and Comparison with Multi-Class Classification

The final decision of the SVM classification system is based on a MAX voting on the results provided by the confidence scores of the multiple view classifiers. The category that corresponds to the highest confidence classifier is the winner. When all 6 viewpoints are used, the system classifies correctly 95.7% of the vehicles on average over all categories.

To ensure that the choice of a set of binary classifiers is the appropriate alternative, additional experiments are performed to compare the performance against multi-class classifiers. A set of 6 multi-class classifiers, one per view, are trained to recognize the 5 categories of vehicles from a given viewpoint. In a similar manner to the set of binary classifiers, the dataset is split into training and testing data for each classifier using 5-fold cross validation. A trained SVM that recognizes a vehicle in a test image coming from a given camera as being a sedan, will output a 1. If it recognizes the vehicle as being a sports car, it will output a 2. The output will be 3 for a vehicle recognized as an SUV, 4 for a pickup and 5 for a wagon. LSSVM classifiers with a Gaussian RBF kernel, the regularization parameter $\gamma=10$ and the squared bandwidth $\sigma^2=0.4$ are used for the 5 category classification. To compute the final decision, a MAX voting is performed on the results obtained per viewpoint. The results are reported in Table 4. The average classification rate is computed for all 6 viewpoints available and averaged over the 5-folds. It can be observed that the performance of multi-class classifiers is lower than the one of the set of binary-classifiers, justifying the selection of the latter in this application.

Table 4: Comparison between binary classifiers and multi-class classifiers

	Average classification rate
Multi-class classifiers	88.1%
Binary classifiers	95.7%

c. Classification Based on Multiple Views and Improvement with a Dissimilarity Score

In the previous section, all non-decision cases are considered classification errors. The average number of non-decision cases computed as average over the 5 folds is 0.4 when all the 6 views are used, but can become larger when only a limited number of viewpoints is considered. To eliminate these non-decision cases, a dissimilarity score is added, as explained in Section III. Initially, an average saliency map model, SM_{avg}, is built for each category of vehicles by computing the pixel mean values of the saliency in all resized SM_l maps that belong to that category, from each view. Examples of such average models are presented in Fig. 4 for the front (dsim 1_x) and three-quarter back (dsim 6_x) views. Each such salient average model is compared with the resized limited *m*-feature saliency map , SM_{test}, computed from the input image corresponding to each viewpoint for every vehicle on which no decision is available when using only the binary SVMs. A sum-of-squared-differences cost (D_{SSD}) measures the intensity difference as a function of dissimilarity between the images corresponding to the salient average model for every category respectively and the saliency map of the uncategorized vehicle within a shifting window *W*:

$$D_{SSD} = \sum_{i,j \in W} \left(SM_{avg}(i,j) - SM_{test}(x+i, y+j) \right)^2 \tag{3}$$

For the experimentation, the size of the window *W* chosen is 9×9. The category that is associated with the lowest dissimilarity D_{SSD} from each viewpoint is considered the winner. To compute the final decision, the category that occurs as a winner the most often among the 6 viewpoints is selected, as detailed in Section III. When all the viewpoints are considered and the dissimilarity score D_{SSD} is applied to disambiguate the few cases where SVM classifiers alone are not sufficient, the system classifies correctly 96.8% of the vehicles, 1.1% more than in the situation where only SVMs are used. However this increase comes at an additional computation cost, as the testing per image takes on average 0.14s, an increase of 0.11s with respect to the case where only SVMs are used.

Table 5 details the evaluation of the classification performance based on true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). It also compares the results of categorization when using the joint confidence (MAX voting) of SVMs only, with the results obtained when the joint confidence produced by the SVMs is complemented by dissimilarity scores. The TP, FP, TN and FN values represent numbers of vehicles. They are computed for each vehicle category as an average over the 5-folds and using the confidence scores of all 6 views available. The precision (PRE), or the percentage of positive predictions that are correct, that is the number of correct results divided by the number of all returned results, and the recall

(REC), or the percentage of positive cases found, that is the number of correct results divided by the number of results that should have been returned, are computed as:

$$PRE = \frac{TP}{TP + FP} \times 100 \qquad (4)$$

$$REC = \frac{TP}{TP + FN} \times 100 \qquad (5)$$

the accuracy (ACC), or the percentage of the predictions that are correct, as:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \qquad (6)$$

and the F_1-score, another measure of accuracy that is a weighted average of precision and recall and that identifies a perfect categorization task with the value 1, as:

$$F_1 = 2 \times \frac{PRE \times REC}{PRE + REC} \times \frac{1}{100} \qquad (7)$$

Table 5: Average values for TP, FP, TN, FN, precision, recall, accuracy and F_1-score over 5 folds for all views

	SVM joint confidence							
	TP	**FP**	**TN**	**FN**	**PRE**	**REC**	**ACC**	**F_1-score**
Sedan	8	0.4	28.4	0.2	95.2%	97.6%	98.4%	0.96
Sports	6.2	0	30.6	0.2	100%	96.8%	99.4%	0.98
SUV	10	1	26.2	0	90.9%	100%	97.3%	0.95
Pickup	6.2	0	30.8	0	100%	100%	100%	1.00
Wagon	4.6	0	30.8	1.4	100%	76.6%	96.2%	0.86
	SVM joint confidence + dissimilarity score							
	TP	**FP**	**TN**	**FN**	**PRE**	**REC**	**ACC**	**F_1-score**
Sedan	8.2	0.2	28.6	0	97.6%	100%	99.5%	0.98
Sports	6.2	0	30.6	0.2	100%	96.8%	99.5%	0.98
SUV	10	1	26.2	0	90.9%	100%	97.3%	0.95

Pickup	6.2	0	30.8	0	100%	100%	100%	1.00
Wagon	5	0	31	1	100%	83.3%	97.3%	0.90

First of all, one can notice the high values for the precision, recall, accuracy and F_1-scores in general. Also, an improvement of these measures can be observed when using the dissimilarity score as a complement to SVMs, as denoted by the slightly higher percentages for ACC and F_1-scores for some of the categories, particularly the sedan and wagon. From Table 5 and also from Table 2, one can notice that the pickup truck classifier performs the best. This is likely due to the pickup truck characteristic shape that significantly differs from other categories. The most common error is the erroneous categorization of wagons and SUVs, as shown by the 1s in the FP for SUV and in the FN for wagon in Table 5. These 1s say that on average 1 wagon is not detected as a wagon (FN wagon = 1), but as an SUV (FP SUV = 1). This is not surprising because many exemplars of SUV and wagon look very similar from the front and back views and both the SUVs and wagons have rounded backs in the side and three-quarter views.

d. Viewpoint Evaluation

Experiments are further 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.

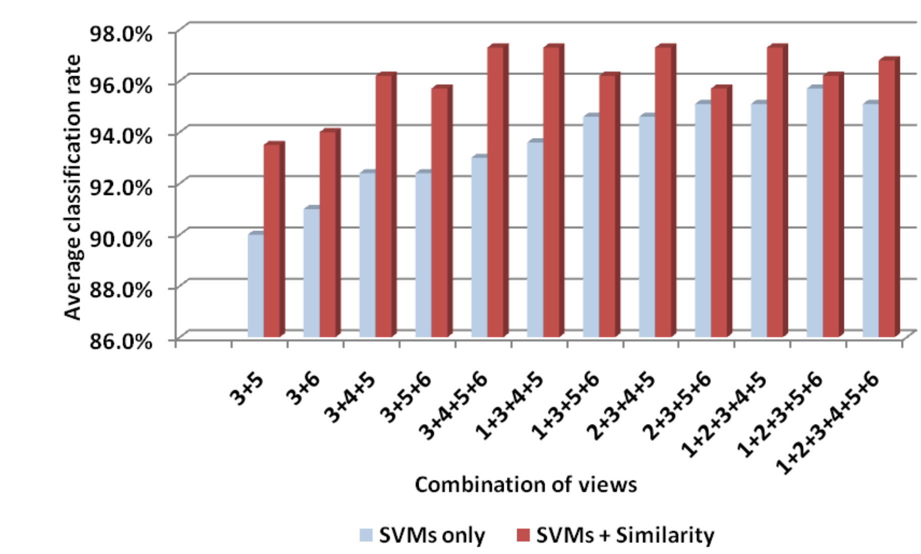


Figure 6. Average classification rates for combinations of views

Most combinations include views 3 to 6 that provided the highest maximum and average classification rate per view as illustrated in Table 1. A series of combinations of views are presented in Fig. 6 together with their average classification rate when using the series of SVMs and when complementing the binary SVMs results with the dissimilarity score.

It can be observed that the classification rates using dissimilarity scores are higher for all combinations of views when compared with the case where only SVMs are used. When only SVMs are used, the views seem to contribute more to the decision. This fact is reflected by a higher number of views required to provide better performance, with (1+2+3+5+6) representing the optimal configuration that achieves 95.7%. Other 3 combinations, namely (2+3+5+6), (1+2+3+4+5) and all 6 views (1+2+3+4+5+6) all obtain 95.1%.

In terms of combinations of views when dissimilarity scores are involved, one can notice that combinations of 4 views, namely lateral and three quarter views (3+4+5+6), front, lateral and front three-quarter views (1+3+4+5), and rear, lateral and front three-quarter views (2+3+4+5) all lead to the best performance of 97.3%, which is higher than the classification rate obtained when using all 6 viewpoints available (1+2+3+4+5+6). A maximum of 96.2% can be obtained for three views when dissimilarity scores are considered. The two lateral views and the front three quarter-view (3+4+5) provide better performance than any case where only SVMs are considered (independently from the number of views). Such a combination provides a less costly solution in terms of use of cameras and still provides better performance for 5-category vehicle classification, but is more costly in computation time. Therefore in choosing the SVM only solution or the improved SVM solution with dissimilarity score, one must consider the compromise between a fast solution (on average 0.03s per testing image) where some non-decision cases remain (between 0.4 cases per fold when all 6 views are used and 3.4 cases per fold when only 2 views are used) and a slower solution (on average 0.14s per testing image) that eliminates all the non-decision cases. The latter approach is also cheaper in terms of equipment and leads to higher classification performance when a minimum of 3 views are used.

e. Comparison with Existing Vehicle Classification Solutions

Even without the addition of the dissimilarity score, the proposed classification technique obtains better results than the case in which SIFT key points, Harris corners, DoG and Gabor features are used as features to support the classification, as shown in section V.a.

Table 6: Comparison with existing solutions from the literature for vehicle classification

Solution	**Categories considered**	**Best classification rate reported**
Ref. [11]	Motorcycle, car, bus, trailer 1, trailer 2	83%
Ref. [12]	Car, van, truck	93.51%
Ref. [13]	Sedan, van, hatchback, bus, truck	93% - 95%
Ref. [15]	Car- minivan, and car - taxi	96%, and 98%
Ref. [17]	Sedan, wagon, van, hatchback	54%
Ref. [18]	Sedan, van, pickup, van, bus, trailer	91.35%
Ref. [20]	Sedan, van, truck	95%
Proposed solution	Sedan, sports car, SUV, pickup truck, wagon	97.3%

Moreover, the proposed approach compares and even surpasses in performance the best solutions found in the literature for multi-category vehicle classification, as detailed in Table 6, while remaining computationally efficient for real-time applications.

VI. CONCLUSIONS

The paper discusses the use of human visual attention mechanisms to identify a set of features that allows for fast automated classification of vehicles based on images taken from 6 viewpoints, with possible application in many areas, such as optimizing available parking spaces, balancing ferry loads, planning infrastructure and managing traffic, or servicing vehicles. The experimental results demonstrate that biologically-inspired features derived from visual attention combined with series of binary support vector machines obtain better classification rates than the cases in which SIFT key points, Harris corner, DoG or Gabor features are used to support the classification. Two original approaches are presented and validated, that is the SVM only solution and the improved SVM solution with dissimilarity score. These alternatives provide the user with the possibility to make a compromise between a fast solution that will leave a low rate of non-decision cases, when these can be tolerated; and a slower solution that eliminates all non-decision cases, while having also reducing the cost of the required equipment and complexity of physical implementation and leading to higher classification performance when a minimum of 3 views are used. The classification rates obtained by the proposed solution compare and even surpass the best solutions reported in the literature for multiple category vehicle classification.

As future work, the proposed solution will be tested on other vehicle datasets and for additional vehicle categories for a more thorough evaluation of 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] E. Meyers and L. Wolf, “Using Biologically Inspired Features for Face Processing”, Int. Journal of Computer Vision, vol. 76, pp. 93-104, 2008.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] 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. Int. Conf. Technology and Aging, Toronto, Canada, 2007.

[11] 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, pp. 1119-1142, 2009.

[12] D. Ponsa, and A. Lopez, "Cascade of Classifiers for Vehicle Detection", J. Blanc-Talon et al. (Eds.): ACIVS 2007, Lecture Notes in Computer Science, LNCS 4678, pp. 980–989, 2007.

[13] 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.

[14] F. M. Kazemi, H. R. Pourreza, R. Moravejian and E. M. Kazemi, "Vehicle Recognition Using Curvelet Transform and Thresholding", T. Sobh (Ed.): Advances in Computer and Information Sciences and Engineering, Springer, pp. 142–146, 2008.

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

[16] H. J. Lee, "Neural Network Approach to Identify Model of Vehicles", J. Wang et al. (Eds.): ISNN 2006, Lecture Notes in Computer Science, vol. 3973, Springer, pp. 66–72, 2006.

[17] 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.

[18] C.-L. Huang, and W.-C. Liao, "A Vision-Based Vehicle Identification System", Proc. Int. Conf. Pattern Recognition, vol. 4, 2004, pp. 364-367.

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

[20] P. Dalka, and A. Czyzewski, "Vehicle Classification Based on Soft Computing Algorithms", M. Szczuka et al. (Eds.): RSCTC 2010, Lecture Notes in Artificial Intelligence, vol. 6086, Springer, pp. 70–79, 2010.

[21] X. Clady, P. Negri, M. Milgram, and R. Poulenard, “Multi-class Vehicle Type Recognition System”, L. Prevost, S. Marinai, and F. Schwenker (Eds.): ANNPR 2008, Lecture Notes in Artificial Intelligence, vol. 5064, Springer, pp. 228–239, 2008.

[22] 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.

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

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

[25] S. Ettinger, SIFT Point Detector Matlab implementation, available online, http://robots.stanford.edu/cs223b04/MatlabSIFT.zip.

[26] C. G. Harris, and M. J. Stephens, “A combined corner and edge detector”, Proc. Vision Conference, Manchester, pp 147-151, 1988.