Image-Based Localization of Vehicle Parts Guided by Visual Attention

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

***Abstract*—The automated servicing of vehicles is becoming more and more a reality in today's world. While certain operations, such as car washing, require only a rough model of the surface of a vehicle, other operations, such as changing of a wheel or filling the gas tank, require correct localization of the different parts of the vehicle on which operations are to be performed. The paper describes an image-based approach to roughly localize vehicle parts over the surface of a vehicle with a bounding box approach based on a model of human visual attention. The proposed method is automatically adapted for different views of a vehicle and obtains average localization rates for different vehicle parts of over 95% for a dataset of 120 vehicles belonging to three categories, namely sedan, SUV and wagon.**

Keywords-vehicles, vehicle parts localization, visual attention, image processing, automation.

I. INTRODUCTION

With the increase in the number of vehicles on the roads, there is also an increasing need for dealers to offer efficient and fast service. Complicated operations within the car are still and will remain in the realm of human technicians. However many operations such as unscrewing or changing wheels or filling the gas tank could become automated and be executed with the help of servicing robots. The successful execution of such operations requires an as-accurate-as-possible localization of vehicle parts to avoid excessive movement of the robotic equipment that is usually time-consuming, and leads to safety concerns.

The work in this paper proposes a novel solution to the problem of localization of vehicle parts such as wheels, windows, lateral mirrors, head lights and rear lamps, front and rear bumpers, and gas tank trap in a set of images of vehicles. It proposes an original bounding-box approach to roughly locate the vehicle parts based on biological visual attention. Since humans show a significantly superior performance in interpreting visual scenes and extracting regions of interest, human visual capabilities are a rich source of inspiration for the improvement of computational vision algorithms. A model of human visual attention, that plays the role of extractor of relevant regions of interest in a visual scene, is used in the context of this paper to identify areas of interest over the surface of the vehicle. Visual attention models build what is called a saliency map (SM) that is a map in which areas of high interest are highlighted. The projection curves on the two axes of the binary converted SM contain important information on the location of different parts of a vehicle and allow for the identification of a set of bounding boxes that contain vehicle parts. The set of bounding boxes is spatially adjusted over the surface of the vehicle according to its category.

The paper is organized as follows: Section II presents the related work on the topic and section III describes the proposed method for localization of vehicle parts. Section IV discusses the experimental results and compares them with a state-of-the-art solution. Finally, the conclusions and the future work are presented in Section V.

II. RELATED WORK

While there has been a lot of interest in the detection of vehicles in images (a survey is available in [1]), there are relatively few papers dedicated to the localization of vehicle parts in the literature. Most of the existing work is concerned with the location of the vehicle license plates and logos [2-7]. In [2], an adaptive segmentation technique called Sliding Concentric Windows is employed to locate the license plate. Chacon and Zimmermann [3] apply Pulse Coupled Neural Networks (PCNN) to generate candidate regions that may contain a license plate. To solve the same problem, Guo *et al.* [4] propose a hybrid method based on PCNN and wavelet analysis. The method proposed in [5] to extract the vehicle plate region uses specific knowledge, such as the higher density of the plate region due to the presence of characters. A three-layer neural network, trained with the texture descriptors computed from the front image of vehicles is used for car plate and logo recognition in [6], while in [7] the vehicle license plate location is followed by coarse-to-fine methods to identify the logo of the vehicle using a method based on phase congruency feature map. A summary of other techniques for plate localization and recognition is presented in [8].

A simplistic solution for side-view car fitting based on a sketch vehicle template is proposed in [9]. Brehar *et al.* [10] identify the pillars in side views of vehicles, based on a rough selection of objects that have one or two wheels using circular symmetry, followed by an adaptive boosting classifier built using histograms of oriented gradient features. Lam *et al.* [11] identify different vehicle components such as: roof, windshield, bonnet, side windows, lower front of car (grille,

978-1-4577-1772-7/12/$26.00 ©2012 IEEE

headlight and front bumper) and lower side of car (wheels and door panels) in a monocular traffic image sequence using a topological structure of the vehicle based on multi-scale textural couriers. More recently, Chávez-Aragón *et al.* [12] proposed a method for vision-based detection of vehicle parts in lateral views that include bumpers, door handles, windows, wheels, lateral mirror, windshield, center, roof, head light and rear lamp. The approach is based on a geometrical model to determine feasible search areas for parts and on a Cascade of Boosted Classifiers based on Haar-like features to detect the parts, in a fixed sized bounding box style, within each feasible search zone.

The current paper goes beyond state-of-the-art solutions by proposing a novel bounding box approach to identify the position of different vehicle parts that is adjusted to fit different categories of vehicles. An advantage of the proposed solution is that, unlike other approaches, it can be easily expanded to localize parts in different views of a vehicle. At this stage, the proposed method identifies areas where a part could be situated and does not attempt to precisely locate the part within its corresponding bounding box.

III. LOCALIZATION OF VEHICLE PARTS BASED ON VISUAL ATTENTION

The proposed system for localization of vehicle parts starts with an initialization step in which all the images in the dataset are aligned to a reference image and their SMs are built based on a model of human visual attention. The category of each vehicle is determined based on the SM using the solution proposed in [13].

For each category of vehicles, an average SM, $SM_{avg\_cat\_view}$, is built for each view of a vehicle by summing the individual SM models viewed from a given direction and dividing the resulting model by the number of vehicles within the category. This average model serves as a basis for the identification of category-specific bounding boxes. The bounding boxes are determined by adding the intensity values from the average SM model along every column (X) and row (Y) respectively and extracting the local minima and maxima over the resulting projection curves. These local extrema contain important information on the position of different parts of interest, as will be detailed in section III.A. The coordinates of the local minima and maxima serve as coordinates for the bounding boxes. The average model of a sedan, $SM_{avg\_sed\_view}$, is considered as a reference vehicle model and is the only one for which the user selects among the boxes, the ones that contain parts of interest. This model is then automatically adjusted to fit the position of the bounding boxes, BB_{cat_view}, for other vehicle categories, as further detailed in section III.B. By computing the difference between the local minima and maxima on the X axis obtained on the average SM of any other category and the local minima and maxima on the X axis of the sedan, the boxes can be displaced automatically to the left or right to better fit the shape of the new category. In a similar way, the boxes can be moved up or down based on the comparison of local minima and maxima on Y axis. Each time an image of a vehicle is presented to the system, the SM is computed and this information is first used to categorize the vehicle. The same information is also used to align the image with the reference image. The corresponding bounding box model is selected according to the category to which the vehicle in the image belongs and the corresponding boxes are displayed overlapped on the vehicle image. The overview of the proposed solution is illustrated in Fig. 1.

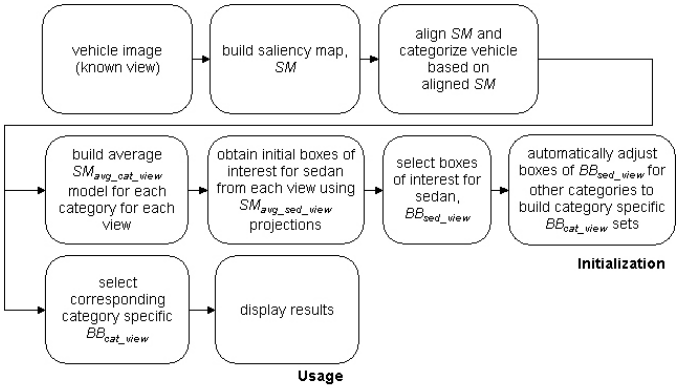


Figure 1. Overview of the solution.

A. Bounding box Localization of Vehicle Parts Based on a Model of Human Visual Attention

The main idea behind the bottom-up visual attention computational model of Itti *et al.*'s [14] is to compute several features derived from a color image provided as input and fuse their saliencies into a representation called saliency map. Several features such as the intensity ($I = (R+G+B)/3$ where R, G and B are the red, green and blue color channels), 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) are computed in parallel and feature-dependent saliencies are computed for each of the three channels. Center-surround operations modeled as a difference between fine and coarse scales are applied on all features. Each feature set is stored in feature dependent saliency 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, SM. The full implementation details are available in [14]. In this work, a simplified computational attention model, using the RGBY color space, 4 orientation channels, one image pyramid, one center scale and a small blur radius of the order 0.01 is used to identify the areas of interest in the images representing front, back and lateral views of vehicles belonging to three categories: sedan, SUV and wagon.

The saliency map plays a triple role in the context of this work. It first allows the identification of the category of vehicle. A method that was proposed for image-based vehicle classification based on SM and support vector machines is used for this purpose. Full details are available in [13]. The method achieves classification rates between 86.9% and 96.2% for the three categories of vehicles used this paper when the views from front, back and lateral sides are considered separately and a rate of 94.3% when the decision from all four views is considered.

Secondly, the SM provides the necessary information for the alignment of vehicles with the reference image. Images are aligned such that the center of all vehicles corresponds with the center of the vehicle in the reference image. The center is computed based on the width and height projection curves of the corresponding SM. In order to compute the width and height, the image representing the SM is initially converted to black and white, using the Otsu thresholding method to obtain SM_{bw}. The vertical and horizontal projection curves are built by summing all the columns of SM_{bw} to obtain width information, *w,* and all the rows to obtain height information, *h.* The horizontal projection curve is then searched starting from the left until a value different from 0 is identified. Each time a value of 0 (empty background) is encountered, the value of the width, *w,* is decreased by 1 pixel. When the first value different from 0 is encountered, the search from the left direction is stopped. The same procedure is used from the right direction by decreasing the remaining width value, *w,* until a value different from 0 is found. A similar top and down search is performed on the vertical projection curve to compute the height, *h,* of the vehicle in the image. The center of the vehicle is computed based on the width and height information, as having the X coordinate equal to the coordinate where the first non-zero value appeared in the horizontal projection curve (where the body of the car starts) plus half of the computed width, and the Y coordinate as the coordinate where the first non-zero value appeared in the vertical projection plus half of the computed height. The reference center value is computed for the reference image. A translation is performed to move the center of any other image to the reference center and therefore align all vehicles.

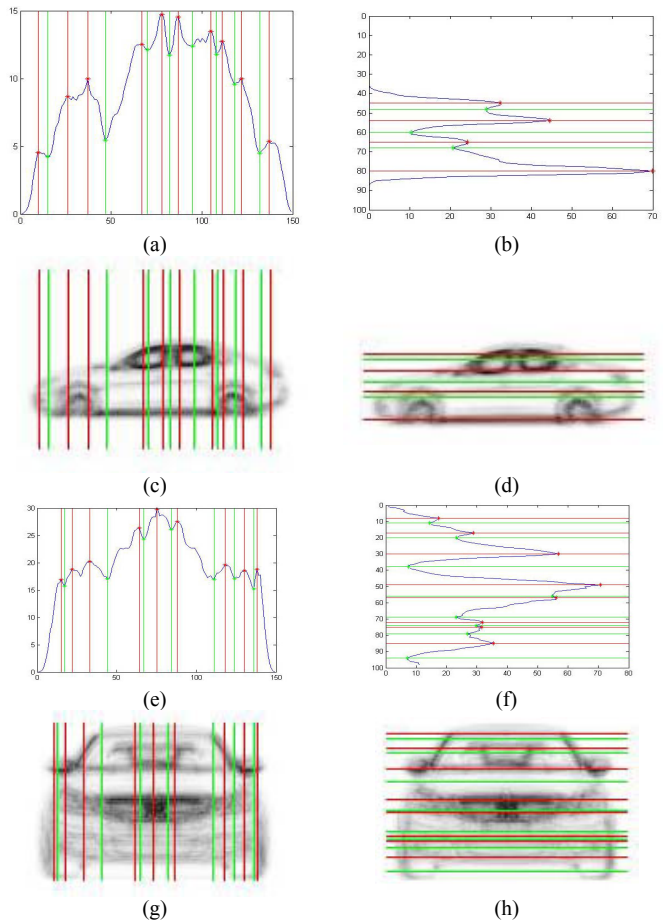
Thirdly, the SM provides the necessary information for identification of vehicle parts. Fig. 2 shows the average SM model, $SM_{avg\_sed\_view}$, for the sedan class viewed from the lateral and from the front views. The SMs are presented as negatives (1-SM) for better visualization of the results. In the image the areas of higher interest are marked by darker shades.

(a) (b)

Figure 2. Average saliency map for sedan class: a) lateral view and b) front view.

The projections on the X and Y axes of these average models are illustrated in Fig. 3a and 3b for the lateral view and in Fig. 3e and 3f for the front view, respectively. The local maxima and minima are computed using the solution proposed in [15]. These extrema are not local maxima and minima from a strict mathematical sense, but rather peaks and valleys of the curves, which is what is required in the current application. Local maxima are displayed with red stars/bars and local minima with green in Fig. 3. Local maxima and minima are used together to achieve better localization along the profiles.

Fig. 3c and 3g show the correspondence between the projection curves on the X axis and different parts of the vehicle viewed from lateral and front, with red vertical lines representing the values of local maxima and the green lines those of the local minima. One can notice, for example, in Fig. 3c that the first maximum corresponds roughly to the beginning of the head light, while the second and third maxima frame roughly the location of the wheel. Similarly, local maxima along the Y axis, illustrated in Fig. 3d and 3h for the two views, provide an estimate of the location of parts. For example, in the lateral view in Fig. 3d, the front and back bumper and the wheels are situated between the first and second local maxima (shown in red), starting from the bottom of the figure.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 3. Projections on X axis (a) for lateral view and (e) for front view, projections on Y axis (b) for lateral view and (f) for front view with local maxima shown in red and local minima in green, and correspondence between projections on (c) and (g) X axis and (d) and (h) Y axis and the location of parts for the lateral and front views, respectively.

By using jointly the information on the two axes, as illustrated in Fig. 3a and 3c for the lateral and front view respectively, a grid is obtained that provides a rough localization of the different parts of interest. Fig. 4b and Fig. 4d show the selected bounding boxes, BB_{sed_view}, corresponding to the parts of interest in this work. For example, in the lateral view illustrated in Fig. 4a, the front bumper is situated in the box that ends with the first local minima (green) on the X axis and between the two local maxima (in red) on the Y axis from the bottom of the figure.

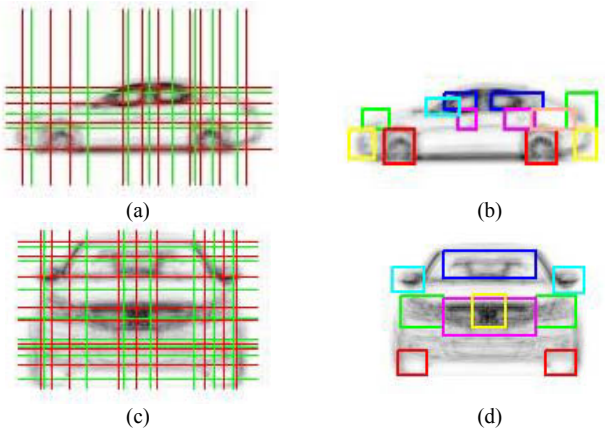


Figure 4. Grid for definition of bounding boxes for (a) lateral view and (c) front view, and bounding boxes corresponding to parts of interest in (b) lateral view and (d) front view.

The corresponding rectangle is shown in yellow in Fig. 4b, which also displays all the parts of interest visible in the lateral view: the bumpers are located within yellow rectangles, the head and back lamps in green, the wheels in red, the handles in magenta, the windows in blue, the mirror in cyan and the gas trap within the orange rectangle. For the front view, the same colors are used to localize the wheels, mirrors, windshield and front lamp. In this case, the grille is marked by a magenta box and the vehicle logo by a yellow box. In a similar way, average models can be built for other vehicle categories.

B. *Adjustments of the Location of Bounding Boxes for Other Vehicle Categories*

The knowledge about the local maxima and minima of the projection curves on the two axes allows for the automated adjustment of the position of bounding boxes to better fit the average models of other categories of vehicles. Fig. 5 shows the difference between the projection curves of the SUV average model (in red) and those of sedan (in blue) for the two views. Only the maxima are shown not to overload the figures.

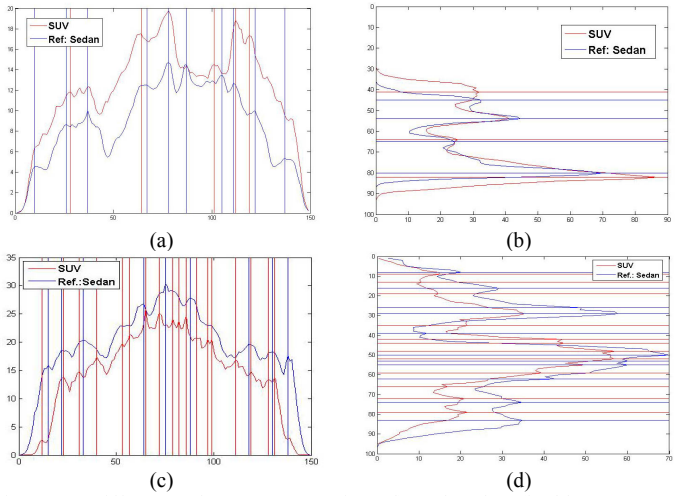


Figure 5. Differences between SUV (in red) and sedan (in blue), (a) X profile side view and (b) Y profile side view, (c) X profile front view and (d) Y profile front view.

The first, third and last maxima on the X axis in Fig. 5a as well as the third maxima from the bottom in Fig. 5b correspond for the two vehicles, but are not visible in the figure due to their overlap. One can notice that some small adjustments can be made to fit the bounding boxes of the sedan over the SUV model. For example, to better fit for the SUV category, the front wheel bounding box, situated between the second and third maxima on the X axis in Fig. 5a and between the first and second maxima from the bottom on the Y axis in Fig. 5b, has to be moved slightly to the right (proportional to the difference between the second local maxima for sedan and second local maxima for SUV) and slightly lower (proportional to the difference between the first local maxima for sedan on the Y axis and the first local maxima for SUV on the same axis). In those cases where more peaks are present in the profile of another category, the average of their value is computed and mapped to the closest value in the sedan profile.

Fig. 6b and Fig. 6d show the better fit of bounding boxes BB_{SUV_view} when such adjustments are performed, as opposed to the situation where the same boxes are used as in the case of the sedan BB_{sed_view}, as illustrated in Fig. 6a and Fig. 6c respectively. Adjustments for any other category of vehicles can be obtained in a similar fashion.

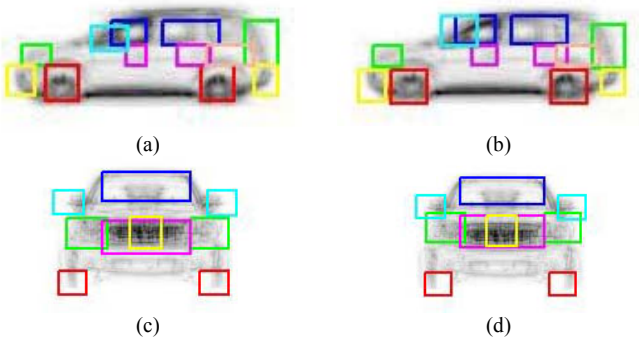


Figure 6. Reference bounding boxes of sedan superimposed over the average SUV model for (a) lateral view and (c) front view, and adjusted position of bounding boxes for SUV for (b) lateral and (d) front view.

In terms of computation time, it takes on average 0.02s from the moment of reading an input image to the moment that the bounding boxes are displayed using a Matlab platform running on a Pentium 1.3GHz machine with 512MB memory. The computation of average models, including the computation of local minima and maxima and the adjustments, takes on average about 0.9s per vehicle, but is performed offline.

IV. EVALUATION OF VEHICLE PARTS LOCALIZATION

A set of 120 vehicles [16] belonging to the sedan, SUV and wagon categories, is used for experimentation. For each of the vehicles, 4 views are available: the front and back views and the two lateral views: the driver and the passenger side. The vehicles in the dataset are presented against simple white background. This does not pose any problems for the purpose of this work, in which the background is likely known (e.g. a garage) and can be subtracted prior to the application of the algorithm. In fact, the solution has been tested on outdoor images and works well if the background is relatively uniform

in color, therefore not capturing the attention on parts that do not belong to the vehicle. The performance remains stable as long as there is no significant scaling between the vehicle in the test image and the ones in the datasets. Each image in the dataset has 99×155 pixels.

Fig. 7 to Fig. 10 show several examples of bounding boxes obtained using the proposed method for the three categories of vehicles in the dataset and on some outdoor images. One can notice that the bounding boxes are well localized over the surface of the vehicles in all cases. It is again worth mentioning that the boxes indicate candidate regions where parts would be situated without trying to detect them in the image. In order to quantify the accuracy of the proposed bounding box approach, a quantitative measure is computed as the percentage of the surface of the actual vehicle part covered by its corresponding bounding box. The surface of the actual vehicle part is obtained by manual segmentation. Table I summarizes in the first 4 columns the corresponding percentages for the driver lateral view for each of the three categories and as an average over the three categories.

Figure 7. Bounding boxes for sedans for all four views.

Figure 8. Bounding boxes for SUVs.

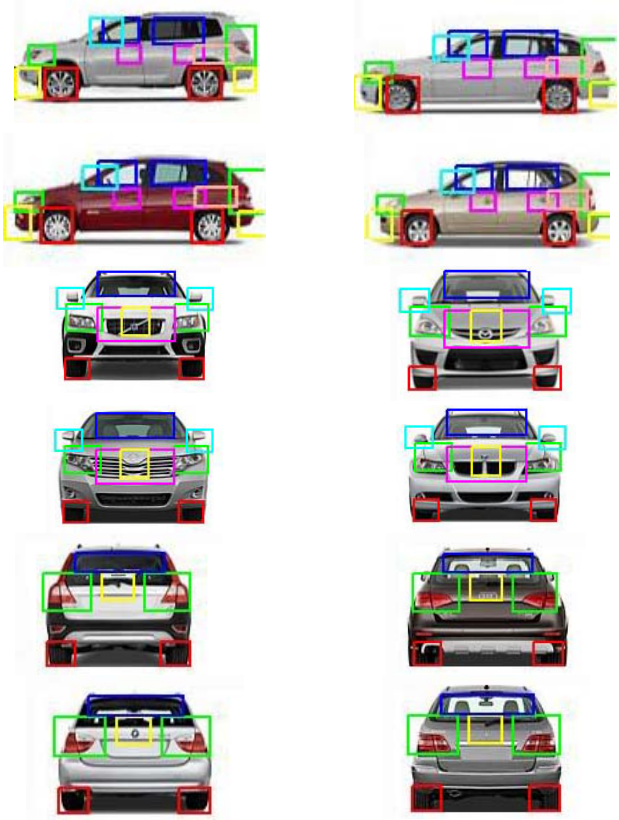


Figure 9. Bounding boxes for wagons.

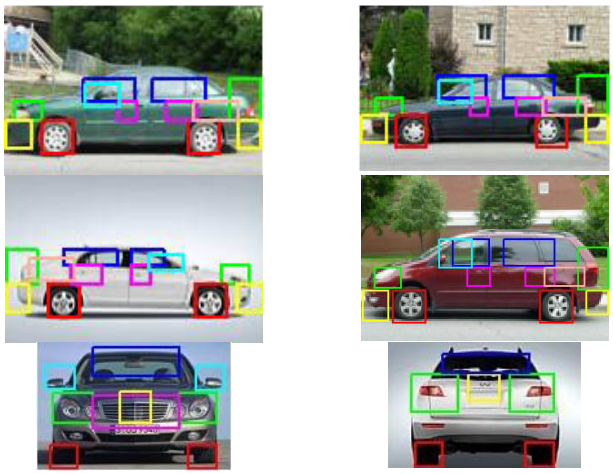


Figure 10. Examples of bounding boxes for outdoor data.

TABLE I. ACCURACY OF BOUNDING BOX LOCALIZATION (LATERAL VIEW)

Vehicle Part	Percentage in Bounding Box				Parts found	Parts found [12]
	Sedan	*SUV*	*Wagon*	*Avg.*		
Front wheel	92.4%	90.8%	89.7%	91%	100%	100%
Rear wheel	95.1%	90.6%	89.6%	92%	100%	100%
Front window	93.8%	94.3%	92.4%	93.5%	100%	N/A
Back window	94.9%	94.5%	95.7%	95%	100%	N/A
Front bumper	100%	100%	100%	100%	100%	96%
Rear bumper	100%	100%	100%	100%	100%	86%
Front handle	90.3%	82.8%	97.1%	90.1%	95.8%	90%
Rear handle	98.3%	89.1%	91.2%	93%	97.6%	85%
Head light	94.2%	99.1%	100%	97.8%	100%	66%
Rear light	98.9%	97.6%	96.7%	97.7%	100%	80%
Mirror	99.8%	99.3%	98.8%	99.3%	100%	53%
Gas trap	100%	89.6%	100%	96.5%	100%	N/A

It can be noticed that overall more than 83% of the different parts are properly located within their corresponding boxes. Similar results were obtained for all four views. In order to enable the comparison with another solution proposed in the literature that reports results on the lateral view only [12], and where the percentage of cases when a part is found are reported, the fifth column in Table I reports the percentage of cases when a part is found within its bounding box (if the part is really present in the image). A vehicle part is considered found when at least 45% of its surface is within its corresponding bounding box. When comparing with the solution in [12], that is also a bounding box style solution, and reported in the last column of Table I for the parts available, it can be seen that the proposed method achieves better detection and localization rates, especially for parts that are visually more challenging to detect such as lights, handles and mirrors.

V. CONCLUSION

The paper presents an improved bounding box approach to localize different vehicle parts in images based on a model of human visual attention and capitalizing on the correspondence between the location of parts and the projections on the axes of the saliency model. The bounding boxes are adjusted for different views over various categories of vehicles and were shown to better localize vehicle parts than other solutions proposed in the literature. As future work, local refinements will be performed within the bounding boxes to allow for a more accurate description of the contours of different parts.

REFERENCES

[1] Z. Sun, G. Bebis, and R. Miller, "On-road Vehicle Detection: A Review", *IEEE Trans. Pattern Analysis and Machine Intel.*, vol. 28, no. 5, 2006.

[2] C.N.E Anagnostopoulos, I.E. Anagnostopoulos, V. Loumos, and E. Kayafas, "A License Plate Recognition Algorithm for Intelligent Transportation System Applications", *IEEE Trans. Intel. Transportation Systems*, vol. 7, no. 3, pp. 377-392, 2006.

[3] M. Chacon and A. Zimmerman, "License Plate Location Based on A Dynamic PCNN Scheme". *Proc. Int. Joint Conf. Neural Networks*, vol. 2, pp. 1195–1200, 2003.

[4] M. Guo, L. Wang, and X. Yuan, "Car Plate Localization Using Pulse Coupled Neural Network in Complicated Environment", Q. Yang and G. Webb (Eds.): PRICAI 2006, *Lecture Notes in Artificial Intelligence*, vol. 4099, pp. 1206–1210, 2006.

[5] K.-B. Kim, S.-W. Jang, and C.-K. Kim, **"**Recognition of Car License Plate by Using Dynamical Thresholding Method and Enhanced Neural Networks"**,** N. Petkov and M.A. Westenberg (Eds.): CAIP 2003, *Lecture Notes in Computer Science*, vol. 2756, pp. 309–319, 2003.

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

[7] A. P. Psyllos, C.-N. E. Anagnostopoulos, and E. Kayafas, "Vehicle Logo Recognition Using a SIFT-Based Enhanced Matching Scheme", *IEEE Trans. Intel. Transportation Systems*, vol. 11, no. 2, pp. 322-328, 2010.

[8] C.-N.E. Anagnostopoulos, I.E. Anagnostopoulos, I.D Psoroulas, V. Loumos, and E. Kayafas, "License Plate Recognition From Still Images and Video Sequences: A Survey", *IEEE Trans. Intelligent Transportation Systems,* vol. 9, no. 3, pp. 377 - 391, 2008.

[9] F. Tan , L. Li, B. Cai, and D. Zhang, "Shape Template Based Side-View Car Detection Algorithm", *Proc. Int. Workshop Intelligent Systems and Applications*, pp. 1-4, China, 2011.

[10] R. Brehar, S. Nedevschi, and L. Daian, "Pillars Detection for Side Viewed Vehicles", *IEEE Conf. Intelligent Computer Communication and Processing*, pp. 247-250, 2010.

[11] W.W.L. Lam, C.C.C. Pang, and N.H.C. Yung, "Vehicle-Component Identification Based on Multiscale Textural Couriers", *IEEE Trans. Intel. Transportation Systems*, vol. 8, no. 4, pp. 681-694, 2007.

[12] A. Chávez-Aragón, R. Laganière and P. Payeur, "Vision-Based Detection and Labelling of Multiple Vehicle Parts", *IEEE Conf. Intel. Transportation Systems*, pp. 1273-1278, Washington, DC, 2011.

[13] A.-M. Cretu and P. Payeur, "Biologically-Inspired Visual Attention Features for a Vehicle Classification Task", *Int. Journal Smart Sensing and Intelligent Systems*, vol. 3, no. 4, pp. 402-423, 2011.

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

[15] E. Billauer, "Peakdet: Peak Detection Using Matlab". Available online: billauer.co.il/peakdet.html.

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