Applied Soft Computing 31 (2015) 369–380

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

An application of a bio-inspired visual attention model for the localization of vehicle parts

Ana-Maria Cretu [a,b,*], Pierre Payeur [b], Robert Laganière [b]

[a] Université du Québec en Outaouais, 101 Saint-Jean-Bosco, Gatineau, QC J8X 3X7, Canada
[b] University of Ottawa, 800 King Edward, Ottawa, ON K1N 6N5, Canada

ARTICLE INFO

Article history:
Received 5 November 2013
Received in revised form 27 January 2015
Accepted 20 February 2015
Available online 16 March 2015

Keywords:
Bio-inspired systems
Visual attention
Automotive vehicles
Vehicle part identification
Image classification
Active shape models

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 a correct localization of the different parts of the vehicle on which operations are to be performed. The paper describes a two-step approach to localize vehicle parts over the surface of a vehicle in front, rear and lateral views capitalizing on a novel approach based on bio-inspired visual attention. First, bounding-boxes are determined based on a model of human visual attention to roughly locate parts of interest. Second, the bounding-boxes are further searched to finely tune and better capture the boundaries of each vehicle part by means of active contour models. The proposed method obtains average bounding-box localization rates over 99.8% for different vehicle parts on a dataset of 120 vehicles belonging to sedan, SUV and wagon categories. Moreover, it allows, with the addition of the active contour models, for a more complete and accurate description of vehicle parts contours than other state-of-the-art solutions. This research work is contributing to the implementation of an automated industrial system for vehicle inspection.

1. Introduction

The increase in the number of vehicles on the roads generates new requirements for car dealers and garages to offer fast and efficient service. While more complicated operations within the car will continue to require the high expertise of human technicians, many simpler operations such as filling the gas tank or changing wheels could become automated and executed with the help of servicing robots in the near future. The successful execution of such operations requires an as-accurate-as-possible localization of specific 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 addresses this issue by proposing a novel solution to the problem of localization of vehicle parts such as wheels, windows, headlights and rear lamps, front and rear bumpers, lateral mirrors and gas tank trap in a set of images representing multiple views of vehicles. It initially proposes an original bounding-box approach to roughly locate vehicle parts based on biological visual attention. Human visual capabilities are a rich source of inspiration for the improvement of computational vision algorithms, since human beings show a significantly superior performance in interpreting visual scenes and extracting regions of interest than most of the current machine vision technologies. Mimicking the role of human visual attention that extracts relevant regions of interest within a visual scene, a computational visual attention model is used in the context of this work to identify areas of interest over the surface of a vehicle. Visual attention models output a representation, called a saliency map (SM), in which areas of high visual interest are highlighted [1]. The projection curves on the two axes of the binary converted saliency map, which appears as a features-of-interest map, 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 those vehicle parts. The set of bounding boxes is spatially adjusted over the vehicle surface according to the category of the vehicle. Active contour models (ACMs) [2] are then applied within the boxes to obtain a finer description of the contour of each part of interest. The proposed method goes therefore beyond state-of-the-art work on vehicle parts localization that is in general limited to fixed sized bounding boxes. It also allows the identification of a significantly larger number of parts in different views, compared to the current

* Corresponding author at: Université du Québec en Outaouais, 101 Saint-Jean-Bosco, Gatineau, QC J8X 3X7, Canada. Tel.: +1 819 595 3900x1527.
E-mail addresses: ana-maria.cretu@uqo.ca (A.-M. Cretu), ppayeur@uottawa.ca (P. Payeur), laganier@uottawa.ca (R. Laganière).

http://dx.doi.org/10.1016/j.asoc.2015.02.032
1568-4946/

literature that is generally restricted to a limited number of vehicle parts and to a single view of a vehicle.

The paper is organized as follows: Section 2 presents the related work on the topic and Section 3 describes the proposed method for localization and fine tuning of vehicle parts. Section 4 discusses the experimental results and compares them with state-of-the-art solutions. Finally, the conclusions and the future work are presented in Section 5.

2. Literature review

While there has been a lot of interest in the detection of vehicles in images (a survey is available in [3]), 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 [4–9]. In [4], an adaptive segmentation technique called Sliding Concentric Windows is employed to locate the license plate. Chacon and Zimmerman [5] apply Pulse Coupled Neural Networks (PCNN) to generate candidate regions that may contain a license plate. To solve the same problem, Guo et al. [6] propose a hybrid method based on PCNN and wavelet analysis. The method proposed in [7] 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 texture descriptors computed from the front image of vehicles is used for car plate and logo recognition in [8], while in [9] the vehicle license plate location is followed by a coarse-to-fine method to identify the logo based on a phase congruency feature map. A summary of various other techniques for plate localization and recognition is presented in [10].

A simplistic solution for side-view car fitting based on a sketch vehicle template is proposed in [11]. Brehar et al. [12], identify the pillars in side views of vehicles, based on a rough selection of objects that are likely to have one or two wheels based on circular symmetry, followed by an adaptive boosting classifier built using histograms of oriented gradient features. Because the solution is based on the detection of wheels, the approach cannot be directly applied for frontal or rear views. Lam et al. [13] identify different vehicle components such as: roof, windshield, bonnet, side windows, lower front of car (grille, 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. The vehicles are divided into multiscale regions based on the center of gravity of the foreground vehicle mask and the calibrated-camera parameters. A series of three key feature points, selected based on the assumption that cars have generally a windshield and headlights, allows for the identification of parts. The method is not directly extendable to rear views as the windshield is not visible and there is an uncertainty that the same key feature points would be useful in this case. Chang and Cho [14] detect in real-time the bumpers and wheels of a moving vehicle, viewed from the lateral side only, using Haar-like features. The algorithm capitalizes on temporal correspondence to reduce the search zones for parts, by verifying when a vehicle enters, and respectively exits the video frame. For this reason this approach cannot be directly applied for frontal or rear views. More recently, Chávez-Aragón et al. [15] proposed a method for vision-based detection of vehicle parts such as bumpers, windows, door handles, wheels, lateral mirror, windshield, center, roof, headlights and rear lamps in lateral views. The approach is using a geometrical model to determine feasible search areas for parts and a cascade of boosted classifiers based on Haar-like features to detect the parts within each feasible zone, in a fixed sized bounding box style. The algorithm first identifies the two wheels in a side view of a vehicle, and their relative position in order to determine the location of other parts with respect to it. Therefore this solution is not suitable neither for the identification of parts in frontal or rear views. In [16,17], the license plate and rear lamps (only red ones), are localized in rear images of cars using their distinctive color, texture, and expected regions in the context of an urban traffic surveillance application. This sort of approach does not work on the detection of parts that are not clearly identified as having a different color or texture or on different views of a vehicle.

The current paper builds on previous work of the authors on the topic [18,19] and goes beyond state-of-the-art solutions by initially proposing a novel, better performing bounding box approach based on visual attention to roughly identify the position of vehicle parts in a first phase. Following a simple initialization stage, in which the user selects from a series of bounding boxes over the surface of a single vehicle the ones of interest for his/her application, the novelty of the approach consists in adapting a visual attention model to automatically adjust the bounding boxes to better fit these parts of interest for any other vehicle category. Therefore this initialization step allows for a smooth and simple adaptation to any specific application that requires the identification of vehicle parts and contributes to ameliorate the performance by only processing relevant information. A significant advantage of the proposed solution is that, unlike other approaches available in the state-of-the-art literature, it can localize a larger number of parts and operates from different views (e.g. side, frontal or rear) of a vehicle. The solution is also further improved in order to obtain a finer description of the contour of each part discovered in a given view by using ACMs in each bounding box. A thorough comparison performed with similar work shows the superior performance of the proposed approach.

3. Localization of vehicle parts based on visual attention and active contour models

The proposed system for localization of vehicle parts is composed of two major steps: an initialization step and a refinement step and its main blocks are illustrated in Fig. 1.

In the initialization step, all the images in the dataset are aligned to a reference image (usually the one belonging to the first vehicle in the dataset) and their saliency maps (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 [19]. For each category of vehicles (here sedan, SUV and wagon categories are considered), an average SM model, $SM_{avg\_cat\_view}$, $cat \in \{sed, SUV, wag\}$, $view \in \{1,2,3,4\}$ 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, or in other words, by the number of individual SM models. The views for each vehicle are provided by four distinct cameras situated around the vehicle, one in front (view 1), one in the back (view 2) and the other two on the lateral sides (views 3 and 4). Due to this specific setup that is compatible with the proposed application, the view from which the vehicle is seen in an image is known because it comes from a given camera.

This average SM model serves as a basis for the identification of category-specific bounding boxes BB_{cat_view}. The bounding boxes are determined by projecting the average SM model on its *X* and *Y* axes and extracting the local minima and maxima over the projection curves. These local extrema contain important information on the position of different parts of interest, as will be detailed in Section 3.1. The coordinates of the local minima and maxima serve as coordinates for the bounding boxes. The average model of