Vision-Based Detection and Labelling of Multiple Vehicle Parts

Alberto Chávez-Aragón, Robert Laganière and Pierre Payeur

Abstract— This paper presents a method for the visual detection of parts of interest on the outer surface of vehicles. The proposed method combines computer vision techniques and machine learning algorithms to process images of lateral views of automobiles. The aim of this approach is to determine the location of a set of car parts in ordinary scenes. The approach can be used in the intelligent transportation industry to construct advanced monitoring and security applications. The key contributions of this work are the introduction of a methodology to locate multiple patterns in cluttered scenes of vehicles which makes use of a probabilistic technique to reduce false detection, and the proposal of a method for inferring the location of regions of interest using a priori knowledge. The results demonstrate excellent performance in the task of detecting up to fourteen different car parts over a vehicle.

I. INTRODUCTION

This paper describes an approach for visual detection of regions of interest on the outer surface of vehicles. The approach uses computer vision techniques and machine learning algorithms to process images of lateral views of vehicles in search of parts of interest that characterize automobiles such as bumpers, door handles, windows, wheels, lateral mirror, windshield, center, roof, head light and rear lamp. Also, using the location of detected features, it is able to infer the most likely position of those parts that were not located successfully. The proposed solution is scale and rotation invariant, which provides the ability to inspect vehicles from various distances and without specific constraints on the location and orientation of the camera. The approach discussed in this work determines with precision the location of vehicle parts in cluttered scenes which opens the door to a wide variety of applications in the intelligent transportation industry, including automotive assembly lines and traffic surveillance systems.

There are three main contributions in this work. The first contribution is a reliable method for implementing a machine vision system to detect multiple car parts in ordinary images. The second is the introduction of a statistical model to accomplish two tasks: elimination of false detections and inference of the most likely position of undetected parts. The third contribution is the definition of feasible search zones. The latter allows to narrow down the search area for each car part from the complete image to a small segment.

The rest of the paper is organized as follows. Section 2 describes related work. Section 3 introduces the proposed approach. Section 4 presents the evaluation of the method and experimental results. Finally, Section 5 provides some concluding remarks.

II. RELATED WORK

Different computer vision systems have been developed in the past that aimed at detecting regions of interest in images of vehicles. Applications of these systems are mainly in the area of Intelligent Transportation Systems (ITS). Many ITS applications involve the use of Artificial Intelligence techniques for pattern recognition. Recently, some authors have reported the use of a technique proposed by Viola and Jones [1] called Cascade of Boosted Classifiers (CBC). This technique has proven to be useful in detecting faces, wheels, back views of cars and license plates among others [2] [3] [4]. A Cascade of Boosted Classifiers combines a pool of weak classifiers to quickly discard areas where the features of interest are not presented. Then, complex classifiers focus their attention on promising regions of the image. Classifiers use very simple features that are fast to evaluate; as a result, this technique can run in real time. In this work, use is made of the CBC technique to detect not only one pattern in a scene but more than ten different patterns.

Automatic vision systems have also been developed to inspect final products on assembly lines. This kind of system requires techniques to simultaneously locate many reference points and many regions of interest [5] [6]. The use of templates is a popular solution. Templates contain semantic information about the problem domain [7]. Those techniques consist in the translation of a template over the image combined with semantic information. To determine the correct position of a template over the image, similarity measures are used. Those similarity measures are based on the Euclidean distance, Mean Square Error cross-correlation functions and probabilistic functions among others.

In this work, machine learning algorithms are used to detect car parts. Then a probabilistic function calculates the probability that each car part has been detected in the right place using information of pre-defined templates. We called those templates geometrical models. They store semantic information about the location of features of interest in a car image. Finally, if there are any missing features, geometrical models combined with a probabilistic technique permit to infer the position of those car parts.

III. PROPOSED APPROACH

Our main objective is to determine the locations of a set of key car parts over the outer surface of a vehicle. The system

works with color or gray scale images of lateral views of vehicles. Images are 800 pixels wide and 600 pixels high.

To detect a particular key vehicle part, the proposed method narrows the complete image down to a small region which contains that car part. The process involves three stages: 1) identification of the region containing the vehicle, 2) definition of a feasible zone most likely to contain a given key part and 3) localization of a small bounding box around the part.

A. Vehicle Location

The first stage of the approach establishes a frame of reference on the vehicle which will be used in the next stages. The frame of reference is a Cartesian coordinate system whose origin is the center of the rear wheel. The direction of the x-axis corresponds with the directed line that goes from the origin to the center of the front wheel. A coordinate system based on the position of the wheels allows the system to estimate the size of the car in the scene and the slope of the plane where the vehicle is located. To detect the wheels two methods are used. The primary method is a trained classifier and as a backing method the Hough transform algorithm for circles was selected. The classifier for wheels detection uses Haar-like features instead of working directly with image intensities [1].

In order to establish a bounding box around any vehicle in the database, it was necessary, in a previous stage, to extract knowledge from an expert user about the areas of interest in car images. To accomplish this task, a semantic annotation tool was built. The annotation tool allows an expert user to define a set of features of interest on the surface of vehicles from a set of training images. A detailed description of this tool is outside the scope of this paper, but the result of the annotation tool is a set of structures that are called geometrical models (GMs) that define the spatial relationship among parts of interest.

B. Geometrical Model and Feasible Search Zones

A geometrical model which defines the spatial relationship between the features of interest and a fixed point (frame of reference) is composed of a set of Euclidean vectors, where each vector points to a region of interest. Vectors are defined by (r, θ), where r is the norm and θ is the deviation angle with respect to a fixed axis (polar axis). To achieve scale invariance, each vector is normalized with respect to the length between the wheels' centers. The direction of the vector that points toward the center of the front wheel is used as the fixed axis, achieving rotation invariance. Table I illustrates a geometrical model of a car. The second row shows the normalized distance between the center of each area of interest and the fixed point. The third row shows the deviation angle with respect to the polar axis. The deviation angles, expressed in radians, are negative because the upper-left coordinate reference system is used. A patch size is assigned to each geometrical model. The size of the patch depends on the resolution of the image. Figure 1 shows the regions of interest recreated from a geometrical model.

TABLE I
GEOMETRICAL MODEL OF AN ANNOTATED IMAGE

tag	rear wheel	front wheel	rear bumper	front bumper	mirror	rear handle	front handle
d	0	1	0.41	1.26	0.83	0.27	0.55
θ	0	0	-2.78	-0.08	-0.35	-1.09	-0.42
tag	head light	rear lamp	roof	front window	back window	center of car	wind-shield
d	1.19	0.38	0.59	0.66	0.44	0.51	0.80
θ	-0.14	-2.47	-0.87	-0.54	-0.91	-0.57	-0.46

Fig. 1. Areas of interest of an annotated image

The geometrical models permit to define Feasible Search Zones (FSZ) for each part of interest. A FSZ is triangular and it is defined by two Euclidean vectors. A FSZ for a specific vehicle part is the area between those vectors with the highest and lowest deviation angle from the set of geometrical vectors constructed for that specific vehicle part. Fig. 2 shows a FSZ for lateral mirrors. It is expected that inside each defined zone one part car is detected.

Fig. 2. Example of a feasible search zone for mirrors

C. Detection of Vehicle Parts

The detection of the car parts in FSZs is performed using a Cascade of Boosted Classifier (CBC) based on Haar-like Features [1]. This machine learning approach can process images very rapidly with good detection rates. This technique uses a set of classifiers to detect desired patterns. First, simpler classifiers quickly reject false sub-windows and then

more complex classifiers are used to recognize the pattern we are looking for. In this work we experimented with the use of CBC to detect multiples patterns in a cluttered scene. To make viable the use of CBC to recognize more than ten different patterns in a single scene, it was necessary to semi-automatize the task of gathering positive and negative examples from the training database. An expert user defined geometrical vectors for each part in our training image set. These vectors were then used to train a separate classifier for each car part we wanted to detect. To train the set of classifiers, 67% of a collection of images containing 110 images was used.

D. Estimation of the Location of Car Parts

The CBC technique focuses on speed more than on accuracy, therefore two problems can arise: detection of multiple parts in a single FSZ, or failure to detect one or more parts anywhere in the image. To address these problems a probabilistic method for discarding extra parts within a FSZ and for inferring the location of missing parts was developed. The method calculates the most likely position of a vehicle part using the Geometrical Models calculated from the Training Set (GMTS). To calculate the similarity between models m_1 and m_2 we propose a measure d defined as follows:

$$d(m_1, m_2) = \sum P(a_i) \cdot G(f_i, a_i) \tag{1}$$

where $P(a_i)$ is the probability of the successful detection of the classifier for a particular feature a_i. This probability distribution was determined experimentally. $G(f_i, a_i)$ gives the probability that a detected feature was found in the correct place. This Gaussian distribution was calculated based on geometrical models from the training image set and it is defined as follows:

$$G(f_i, a_i) = \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{\delta}{\sigma}\right)^2} \right] \tag{2}$$

$$\delta = \sqrt{\sum (f_{ij} - a_{ij})^2} \tag{3}$$

where f_{ij} and a_{ij} are the geometrical model elements to be compared. Function δ is the euclidean distance between two vectors. Parameter $\sigma = 8$, which was estimated experimentally, allow us to tune the probabilistic model to be more or less restrictive.

The car parts detected are arranged in a new geometrical model where extra parts have been left out. Then, to complete the geometrical model, the probabilistic method estimates the position of the missing parts taking into account two aspects: the location of features of interest in the new model and the location of those features in a template that best fit the partial geometrical model.

Algorithm 1 details in pseudo code the proposed method for the visual detection of car parts. Algorithm 2 presents the proposed method to infer missing car parts.

Algorithm 1 Algorithm for detection of multiple car parts

Input: An image containing a side view of a vehicle
Output: An image where the vehicle features are highlighted

$wheels \leftarrow CBCwheelsDetector(image)$
if not$DetectedWheels(wheels)$ **then**
   $wheels \leftarrow HoughTWheelsDetector(image)$
   if not$DetectedWheels(wheels)$ **then**
     $Exit$
   end if
end if
$D \leftarrow Distance(wheels)$
$S \leftarrow SlopeBaseLine(wheels)$
set up frame of reference
for each car part C to be detected **do**
   for each Geometrical Model in GMTS **do**
     $v_1 \leftarrow leastDeviationAngle(C, GMTS)$
     $v_2 \leftarrow greatestDeviationAngle(C, GMTS)$
   end for
   $v_1 \leftarrow normalize(v_1, D)$
   $v_1 \leftarrow rotate(v_1, S)$
   $v_2 \leftarrow normalize(v_2, D)$
   $v_2 \leftarrow rotate(v_2, S)$
   $FSZ \leftarrow setFSZ(v_1, v_2)$
   $F \leftarrow CBCcarDetector(C, FSZ)$
   for each feature c in F **do**
     $P_c \leftarrow ProbabilityCorrectLocation(c)$
   end for
   $discartFalseDetections(F, P_c)$
   $PGM \leftarrow addDetectedFeature(F)$
   $HighlightAreaOfInterest(image, F)$
end for
$completedGM \leftarrow InferMissingParts($
$image, PGM, GMTS)$
$showResult(image, completedGM)$

Algorithm 2 Algorithm for inferring missing car parts

Routine: InferMissingParts
Input: An image IM, a partial geometrical model PGM of IM and $GMTS$
Output: A completed geometrical model

for each missing car part C in PGM **do**
   for each Geometrical Model GM_i in $GMTS$ **do**
     $P_i \leftarrow probabilisticSimilarityMeasure($
     $GM_i, PGM)$
     $Hp \leftarrow highestProbability(P_i)$
   end for
   $GMBestFit \leftarrow (GMTS, Hp)$
   $F \leftarrow PositionOfFeature(GMBestFit)$
   $PGM \leftarrow addInferredFeature(F)$
end for
$Return(PGM)$

IV. EXPERIMENTAL EVALUATION

A. Data Set and Metrics

A collection of one hundred and ten $800x600$ color images of vehicles was partitioned into two subsets: the training set and the testing set. The training set contains 67% of the collection. The remaining 33% of the images were used to evaluate the proposed method. The distribution of the testing set according to types of vehicle are: 70% 4-door sedan, 10% SUV, 7% pickup truck, 6% 4-door hatch back, 4% 2-door sedan and 3% 2-door hatch back. On the other hand, the training set has the following distribution: 57% 4-door sedan, 16% SUV, 13% 4-door hatch back, 4% 2-door hatch back, 3% 2-door sedan, 4% wagon and 3% 2-door hatch back.

Two standard metrics were used to measure the success of the proposed method : Precision and Recall. They are defined in terms of true positives (tp) (features correctly classified), false positives (fp) (features incorrectly classified) and false negatives (fn) (missing features). Precision is defined as $Precision = (tp)/(tp + fp)$ and Recall is defined as $Recall = (tp)/(tp + fn)$. Both metrics scores go from 0 to 1. In the context of this work, a precision of 1.0 means that every detected feature in the image was correctly labeled. A recall of 1.0 means that all the features presented in the image were successfully detected and labeled.

B. Experiments

The first evaluation consists in testing each classifier separately, the performance of the set of classifiers is reported in the ROC curves shown in figure 3. The ROC curves, which plot true positive rate vs. false positive rate, compare the accuracy and sensitivity of the classifiers for each part of interest. The analysis does not make a distinction between detectors for rear and front handles, and between detectors for front and back wheels. The ROC curves show that the classifiers for wheels present a perfect discrimination. The classifiers for Windshields and rear lamps present a discrimination value close to 85%. The set of classifiers for mirrors and rear bumpers have the worse ratio of accuracy to sensitivity. Rear bumpers and mirrors are difficult to detect due to the fact that in many modern vehicles these parts are integrated with the bodywork; moreover, rear bumpers are located in a region of poor contrast while mirrors frequently are very small.

The next experiments were conducted using the set of classifiers along with geometrical models. Table II shows the average successful classification rate for each part of interest over the entire testing set, taking into account the successful rate of each classifier working individually. In this case, classifiers for rear and front handles and rear and front wheels are evaluated individually.

In these experiments, although there are fourteen parts of interest to be detected in most of the images, the classifiers were trained for detecting twelve for each image in the testing set. The positions of the remaining features (windows) are inferred later on.
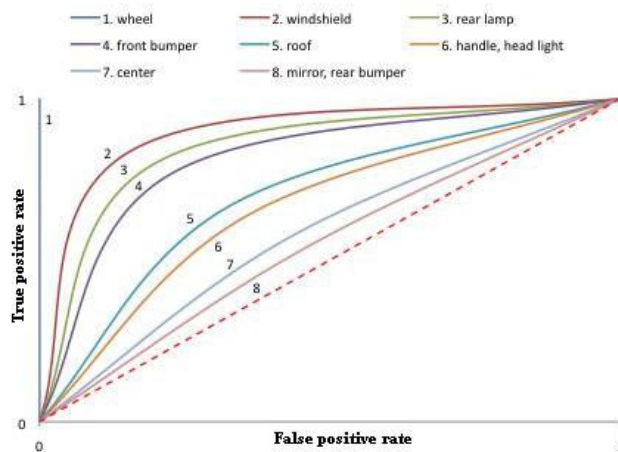

Fig. 3. ROC curves according to detected vehicle features

TABLE II

AUTOMOTIVE FEATURES SORTED ACCORDING TO THEIR CLASSIFICATION RATE

Classification rate			
Feature	Percentage	Feature	Percentage
rear wheel	100%	front wheel	100%
front bumper	96%	roof	93%
front handle	90%	rear bumper	86%
windshield	86%	rear handle	85%
rear lamp	80%	head light	66%
center of car	56%	mirror	53%

Then each detected object was categorized as either true positive, false positive or false negative. The true positive curve in figure 4 indicates that all the features were correctly detected and classified in 13% of the images, eleven features in 23%, ten features in 30% and from nine to seven features in the remaining 34%. The false positive curve indicates the number of features incorrectly classified. Three or two false detections occurred only once, the rest of the images have one or zero. The false negative curve is the complement of the true positive curve. For images 21 and 28 the number of features to be detected is 11, since those images correspond with a 2-door sedan and a 2-door hatch back.
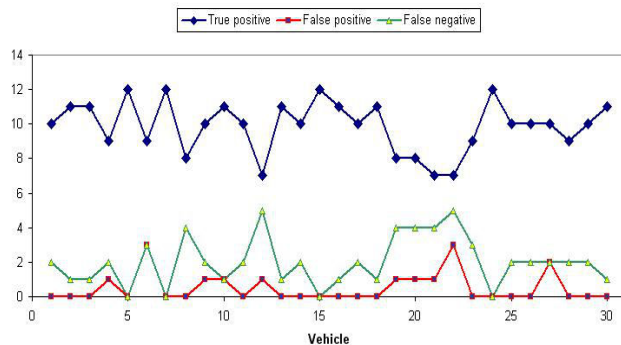

Fig. 4. True positive, false positive and false negative detections for 12 features of interest over the testing set of 30 vehicles

For these experiments the resulting Precision was in the interval $[0.7, 1]$ and averaged 0.95. The Recall was in the interval $[0.58, 1]$ and averaged 0.82. Figure 5 shows the results for a set of queries.
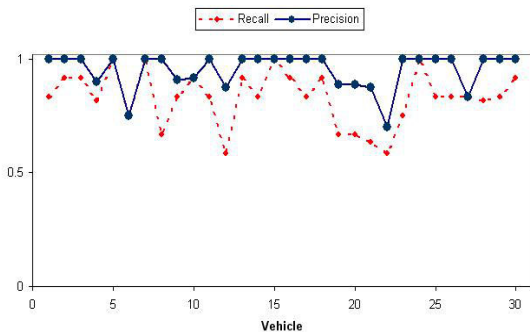

Fig. 5.   Precision and Recall metrics for the testing set

Next, the performance of the system when inferring the location of missing features was analyzed. In order to determine if a feature was correctly inferred we use the same criteria as for calculating precision and recall. In 63% of the images all the missing features were inferred correctly. The overall successful rate of the proposed method for inferring missing features is 79.66%. Finally, some resulting images sorted according to their precision, recall and inference values are presented. Features detected by the set of classifiers are circled while the features inferred by the probabilistic method for missing parts are marked by squares. Figure 6 shows two examples of images processed by the proposed method. Vehicles presented in those images differ in orientation, size, background and slope of the road; however, the method succeeded in detecting all the features. Windows are normally inferred after other parts are detected given the lack of visual features that they exhibit.

Figure 7 shows some results of the method in the presence of other car parts from different vehicles. The proposed method automatically focuses on the vehicle which has its full side view in the scene without a need for cropping the image before processing. Since the system uses the location of the wheels in the image to set up the frame of reference, it is not expected, at this time, to find more than one pair of full wheels in an image.

The proposed approach has shown a good performance working with different kinds of vehicles as shown in Figure 8. It is important to mention that no pick-up truck was used in the training stage and very few minivans. The system, however, was able to locate and label the two door handles of the minivan which are located very close to each other and did not try to infer a rear door handle in the case of the pick-up. These apparently simple decisions made for the system are very important because that indicates that the system can deal with very few information and classify correctly the type of vehicle. It can also generalize the knowledge it had about 2-door cars and apply this information to another type of vehicle such as pick-up trucks.
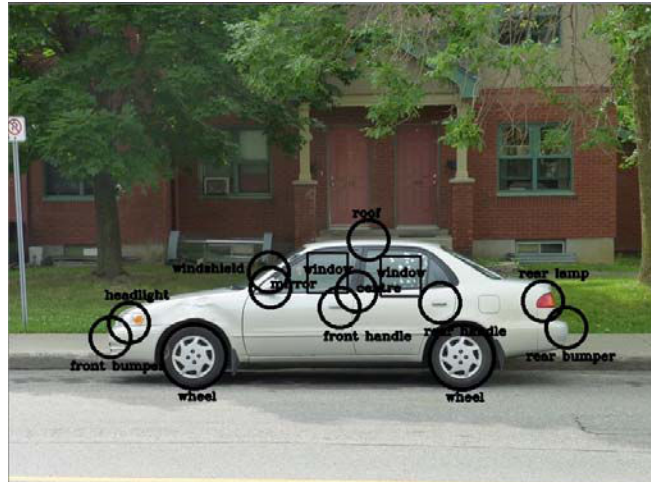

Fig. 6.   Successful detection of all car parts of interest for different types of vehicles in real outdoor images
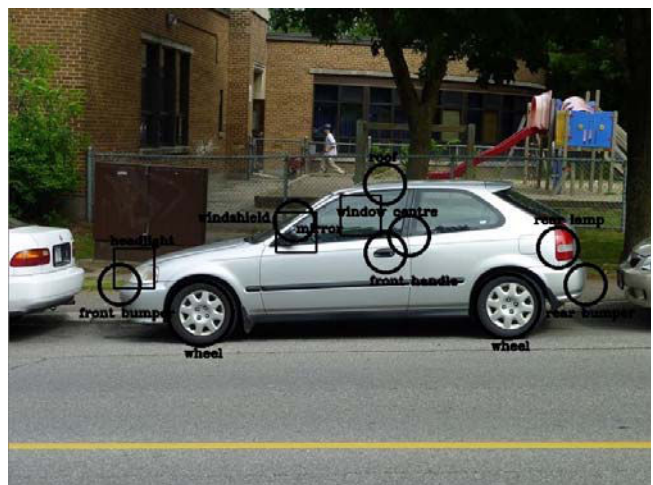

Fig. 7.   Successful detection of car features over the vehicle of interest even when parts of others vehicles are present in the scene
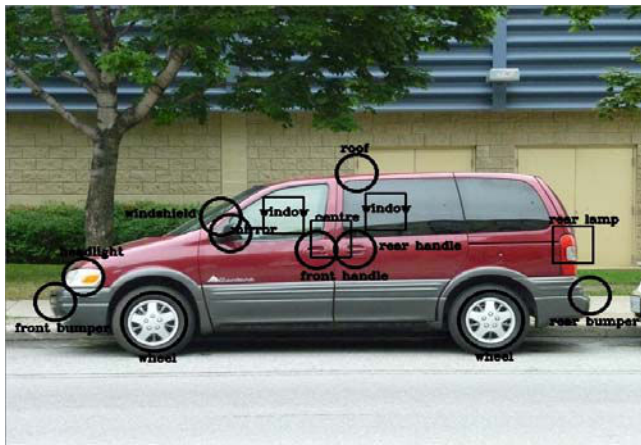
Fig. 8. Few minivans and no pick-up truck were used to train the system; in spite of this, the system can locate, label and infer the desired vehicle features
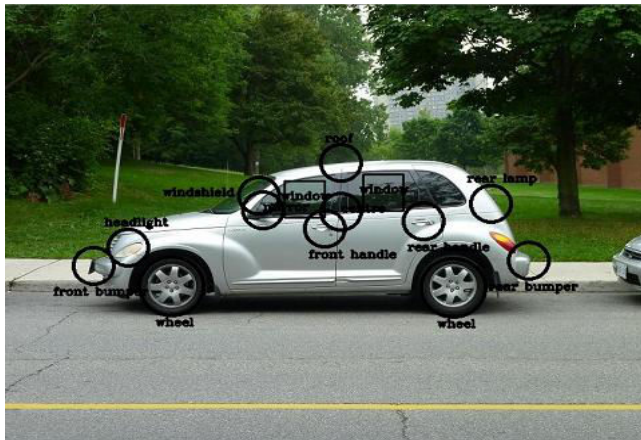

Fig. 9. Most of the features were detected successfully; however, the system mistook the rear lamp for a small part of the bodywork

The great variability in between types of vehicles and the physical characteristics of some parts such as: front lights, rear lamps, bumpers and door handles makes difficult to locate those features on vehicles with unusual shapes. Figure 9 shows a false detection of a rear lamp. Increasing the number of examples of unusual vehicles in the training set could help the system to succeed in locating atypical parts.

The system was implemented in C++ and run on a computer with an Intel Core 2 duo at 3.0 Ghz, 2.0 GB RAM and Windows XP. Table III summarizes the performance of the algorithms in the task of detecting vehicle parts.

TABLE III

PERFORMANCE OF THE SYSTEM

Performance		
Feature detection	minimum time per feature	55.00 ms
	maximum time per feature	170.00 ms
Overall time	Average time detecting plus inferring	3048.79 ms

V. CONCLUSION

In this paper, an approach for the automatic visual detection of car parts on the outer surface of vehicles was presented. The approach is able to determine, with a high level of precision, the location of up to fourteen regions of interest on vehicles such as bumpers, handles, windows, wheels, lateral mirror, windshield, center, roof, head light and rear lamp. The method has immediate applications in the intelligent transportation, security and automobile industries, but also it can be adapted easily for applications where multiple patterns have to be detected in cluttered scenes. The proposed approach contributes to the area of object recognition with the introduction of a probabilistic method to reject false detections and to infer the position of missing objects based on pre-defined semantic patterns. In this work two key concepts were introduced: Geometrical Models (GM) and Feasible Search Zones (FSZ). GMs are used to model the spatial relationship among regions of interest in training images. FSZs delimit search areas over images for each part of interest. GMs and FSZs along with the proposed probabilistic method permit to infer the location of missing objects and eliminate false detections. Experiments with the task of detecting car parts of interest in outdoor environments show an average precision of ninety five percent.

REFERENCES

[1] Viola Paul and Jones Michael, Rapid Object Detection using a Boosted Cascade of Simple Features. *Conference on Computer Vision and Pattern Recognition*, 1:1-9, 2001.
[2] Achler Ofer and Manubhai Trivedi Mohan, Camera Based Vehicle Detection, Tracking, and Wheel Baseline Estimation Approach. *The 7th International IEEE Conference on Intelligent Transportation Systems*, pp. 743-748, 2004.
[3] Fung Yu-fai, Lee Homan and Ercan M. Fikret, Image Processing Application in Toll Collection. *IAENG International Journal of Computer Science*, 32:473-478, 2006.
[4] Trivedi Mohan Manubhai, Gandhi Tarak and McCall Joel, Looking-In and Looking-Out of a Vehicle: Computer-Vision-Based Enhanced Vehicle Safety. *IEEE Transactions on Intelligent Transportation Systems*, 8:108-120, 2007.
[5] Kim Seong-Min, Lee Young-Choon and Lee Seong-Cheol, Vision Based Automatic Inspection System for Nuts Welded on the Support Hinge. *SICE-ICASE International Joint Conference*, 1:1508-1512, 2006.
[6] Agarwal Shivani, Awan Aatif and Roth Dan, Learning to Detect Objects in Images via a Sparse, Part-Based Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1475-1490, 2004.
[7] Kiryakov Atanas, Popov Borislav, Ognyanoff Damyan, Manov Dimitar and Goranov Kirilov Miroslav, Semantic Annotation Indexing and Retrieval. *Journal of Web Semantics*, 2:49-79, 2004.