

# Tracking and Classification of Dynamic Obstacles Using Laser Range Finder and Vision

Gonçalo Monteiro, Cristiano Premebida, Paulo Peixoto and Urbano Nunes

Institute for System and Robotics

University of Coimbra - Polo II

3030-290 Coimbra, Portugal

{ggmonteiro;cpremebida;peixoto;urbano}@isr.uc.pt

**Abstract**—A multi-module architecture to detect, track and classify objects in semi-structured outdoor scenarios for intelligent vehicles is proposed in this paper. In order to fulfill this task it was used the information provided by a laser range finder (LRF) and a monocular camera. The detection and tracking phases are performed in the LRF space, and the object classification methods work both in laser (with a Majority Voting scheme and a Gaussian Mixture Model (GMM) classifier) and in vision spaces (AdaBoost classifier). A sum decision rule based on the *Bayes* approach is used in order to combine the results of each classification technique, and hence a more reliable object classification is achieved. Experiments using real data confirm the robustness of the proposed architecture.

## I. INTRODUCTION

Intelligent vehicles may have its perception capabilities improved if multiple sensory information is fused (combined) in such a way that more relevant information is available as the result of a proper combination of individual sensor's measurements. This paper describes a multi-module architecture with the purpose of processing timely the information about the vehicle's surrounding environment for providing a collision avoidance behavior of low speed vehicles moving in cybercars scenarios [1]. In the proposed architecture, three different classifiers, working in distinct sensor spaces, are combined by means of a practical Bayesian approach in order to provide a higher level of inference and meaningful information to achieve a more reliable object classification. Moreover a cooperation strategy was used to establish the coordinate correspondence between a LRF and a monocular camera in order to facilitate the segmentation process and the object detection.

Detecting multiple objects using LRF [2],[3] and vision [4], or both sensors [5] for purpose of collision avoidance, navigation, SLAM and classification is a well reported subject. In this work, object detection and tracking are performed in a hierarchical structure, having twofold motivations: the segmentation process is simpler and easier to implement in the laser space, and the information provided by the laser sensor is directly related to the positions of the objects. Therefore, dynamic state space models are an intuitive and easy framework to be used in the tracking routines.

The architecture of the system contains some modules, subdivided in four systems: lidar-based and vision-based systems, coordinate transformation, tracking and classification

system. The lidar-based system detects the objects in the laser space, estimates its position and size, and classifies them. The position of the objects is then converted to the camera coordinates system in order to define a region of interest (ROI) in the image space. The vision-based system receives the information from the coordinate transformation module, extracts a window from each of the detected objects and performs the AdaBoost classification. The tracking and final classification module process the information from the others systems and outputs the class of the objects and its dynamic behavior (see Fig. 1).

Three different classifiers are used to distinguish two object categories: vehicles and pedestrians. A *Voting Scheme* [2] and a GMM *Bayes* classifier [6] are used to deal with the information provided by the LRF, and an *AdaBoost* classifier, using *Haar-Like* features, is applied to classify the detected objects in the image frames [7]. There are several categories of classifiers combinations which could be used. However, due to its feasibility, a Maximum A-Posteriori (MAP) practical Bayesian technique [8] was chosen to combine the classifiers presented in our system.

## II. RELATED WORK

Several works have been carried on using laser-scanners in multiple target tracking and object classification, generally related to the field of detecting and tracking moving objects (DATMO), including applications on localization and navigation [9], SLAM [10], collision avoidance, warning systems and others [11],[3]. In order to classify objects, some authors have proposed a variety of approaches: i) voting schemes [2], ii) a method based on "heuristic" rules [12], and iii) multi-hypotheses [13] are some examples of classification approaches using laser scanners. The first two methods have the disadvantage of not having a self-consistent mathematical framework in order to support its stability and consistency; nevertheless the results presented by the authors show its feasibility. The multi-hypotheses method presented in [13] is based on features tracking by means of a Kalman Filter, whose main drawback is a high computational cost and the inconvenient of managing many hypotheses.

Vision systems are widely used for object detection [14],[7] and constitute a feasible option to be implemented separately

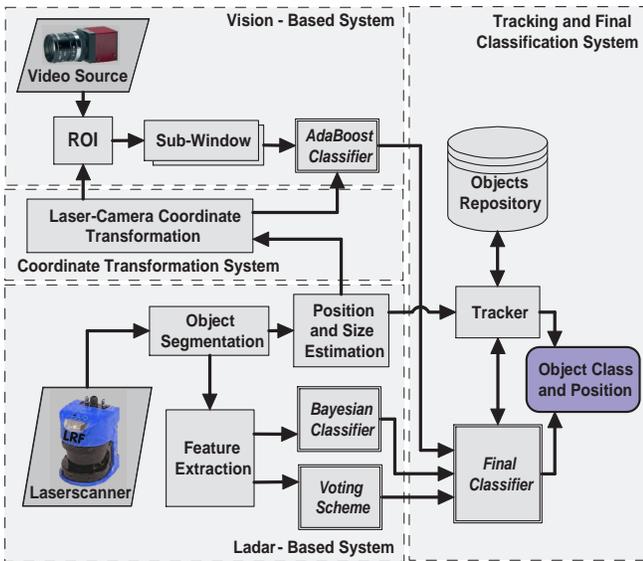


Fig. 1. Multi-module architecture using laser and vision information for detection and classification of moving objects in cybercar-scenarios.

or along with LRF [5]. In outdoor environments the performance of the vision based systems is negatively affected by weather conditions, and the range information, when available, is in general not appropriate for long distances. Therefore laser-scanners in cooperation with cameras are a useful solution to be exploited.

Most of the object detection and tracking systems apply a simple segmentation procedure like background subtraction or temporal differencing to detect objects [15]. But these approaches have a weakness, which is the background changes due to the camera motion.

Other techniques like the disparities discontinuity or optical flow can be used to make the segmentation of the objects when the camera is moving, and then the objects are classified using discriminative methods such as support vector machines, neural networks, etc. In [16] the optical flow was used to discriminate moving targets from the background in the presence of the camera motion. This technique is not appropriate for nonrigid objects extraction, such as pedestrians, since the movements of the body parts are not consistent over time. The disparities discontinuity of a pair of images taken by stereo cameras can be used to detect the foreground objects [17]. This is a robust technique for the segmentation of the objects with a moving camera, although it cannot discriminate when the objects are too close to each other. Another methodology that can be used to classify the objects without performing the segmentation of the frame image, is called *sliding window classification*. The detection of the object class is done by sliding a search window through the frame image and checking whether an image region at a certain location and scale matches with the classifier function. This way, the problem is limited to finding a robust classifier associated to some good features extracted from the frame sub-window to discriminate the class of the object.

In [18] the Distance Transforms is used to extract features from the pedestrians and then the online matching involves a simultaneous coarse-to-fine approach over the shape and over the transformation parameters. A pedestrian feature representation approach is proposed in [19], based on Sparse Gabor Filter. In the detection is used the *Support Vector Machines* classifier. An exhaustive research has been done by Broggi on the detection of pedestrians/obstacles on the road, whether under normal conditions, or night vision. In [20] he uses some algorithms based on edge density and symmetry maps to recognize pedestrians in various poses and with different kinds of clothing. Papageorgiou [21] introduced a trainable object detection architecture based on a novel idea of the wavelet template that defines the shape of an object in terms of a subset of the wavelet coefficients of the image. Motivated in part by this last architecture, [22] presented a machine learning approach for object detection which is capable of processing images extremely rapidly and archiving high detection rates. This is then achieved by the use of a new image representation called the *integral image*, and by a learning algorithm based on AdaBoost. This method combines increasingly more complex classifiers in a *cascade* which allows background regions of the image to be quickly discarded while spending more computation on promising object-like regions. The object detection system based in computer vision described in this paper is based in the Viola et. al. methodology [22].

### III. LADAR-BASED SYSTEM

In this section, it is discussed the tasks performed in the ladar-based system which are related to the processes of segmentation, feature extraction, object detection, classification and tracking.

#### A. Detection and Tracking Objects in LRF Space

To detect the objects, the surrounding environment is segmented using range information provided by the LRF. Among several possible segmentation methods to be used on 2D laser range images [23], a linear KF-based method [11] has been used to perform the segmentation stage. Within the measured segments, a tracking and data association technique is performed in the laser referential system, where the objects under tracking are considered to evolve in time according to a dynamic state equation driven by process noises. The state and measurement noises are considered zero-mean, mutually independent and white Gaussian sequences with known covariances matrices. For object tracking, the measurements are converted from original polar coordinates to Cartesian coordinates and then a linear filter (e.g. a KF) is used [24].

A multi-independent EKF was used to carry out the objects tracking. Although the measurement system is described by a non-linear model, the object motion was modelled by a linear function in the cartesian coordinates. The segments which define an object are a cloud of laser-points. The centroid (center of mass) of the point is calculated and used as the characteristic-point, i.e., the dynamic behavior of the object is described in respect to its centroid. Intrinsically, connected

with tracking, data association is performed considering two situations:

- 1) Segment to segment: the process of associating detected segments with other segments (non-classified objects) in the current scan;
- 2) Segment to object-tracker: the maintenance process, i.e. the association of observed segments with existing objects.

The first situation occurs when one, or more, current segments are “probably” related to an existing segment under tracking, i.e. the observed segment is a valid one, in the sense that its observed state appears in the validation region. Therefore a decision has to be made in order to separate or merge the valid segments. To deal with this situation, a combination of rectangular and ellipsoidal gates are used [24]. The second data association problem, i.e., observation-to-tracker association, is solved in a specific manner which accounts for the result of object classification. Hence, when a segment is finally classified, the data association technique is adjusted in accordance with the size and the dynamic behavior of the class into which the object has been classified.

### B. Ladar-Based Classifiers

This section describes the two implemented classifiers based on data from the LRF. The first is a GMM classifier, where each class is modelled by a finite-GMM whose parameters were estimated during a supervised training. A maximum a-posteriori (MAP) decision rule is used to ultimately classify an object. The second classifier, a majority voting (MV) scheme works with a five-dimensional feature-vector and weighted functions (voting actors) that characterizes each category. By adding the “votes” calculated by the voting actors, an object is classified when a reasonable value is achieved on the type of that category.

### C. GMM Classifier

Assuming that the feature vectors are conditionally and statistically independent given the classes, a GMM classifier (GMMC) [6], using the *Bayes* decision rule, is described in this section. The object classes are modelled by a weighted combination of Gaussian probability density functions (pdf) which are referred to in this context as Gaussian components of the mixture model describing a class (object category). In a GMM model, the probability distribution of a multi-dimensional random vector  $x$  is a mixture of  $M$  gaussian probability density function (GPDF) (1), defined as follows:

$$p(x|\Theta) = \sum_{m=1}^M \alpha_m p(x|\theta_m) \quad (1)$$

where  $\theta_1, \dots, \theta_M$  are the parameters of the Gaussian distributions and  $\alpha = [\alpha_1, \dots, \alpha_M]$  is the weighted vector, such that  $\sum_{m=1}^M \alpha_m = 1$ . The complete set of parameters that specify the mixture model is  $\Theta = (\alpha; \theta_1, \dots, \theta_M)$ , with each parameter  $\theta_m = (\mu_m, \Sigma_m)$  consisting of a mean vector  $\mu$  and a covariance matrix  $\Sigma$ . Considering a  $d$ -dimensional feature-vector  $\Omega$ , the

likelihoods of each class ( $q_i$ ) are described as linear combinations of Gaussian mixture probability density functions:

$$p(\Omega|q_i, \Theta^i) = \sum_{m=1}^M \alpha_m^i p(\Omega|\theta_m^i) \quad (2)$$

where each GPDF component is given by

$$p(\Omega|\theta_m^i) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_m^i|}} \exp\left[-\frac{1}{2}(\Omega - \mu_m^i)^T (\Sigma_m^i)^{-1} (\Omega - \mu_m^i)\right] \quad (3)$$

The GMM parameters for each object class were estimated using the expectation-maximization (EM) algorithm, i.e., for each set of  $N$  labelled feature-vectors ( $\Omega^N = \Omega_1, \Omega_2, \dots, \Omega_N$ ) the EM algorithm calculates  $M$  Gaussian parameters-vector that maximizes the joint likelihood (4) among the GPDF-components:

$$p(\Omega^N|q_i, \Theta^i) = \prod_{j=1}^N p(\Omega_j|q_i, \Theta^i) \quad (4)$$

A recent book devoted entirely to EM and its applications is [25], whereas [26] is another interesting reference which provides a public toolbox.

To select which of the categories ( $q_i$ ), modelled by the GMM parameters  $\Theta^i$ , fits the current observation feature-vector  $\Omega_k$ , i.e at current time interval  $k$ , a practical maximum a posteriori (MAP) decision rule based on the likelihood 2 and on the prior probability ( $P(q_i|\Theta^i)$ ) is used. Considering the attributes equiprobable, the posterior probability  $P(q_i|\Omega, \Theta^i)$  for all categories is proportional to the multiplication of the likelihood with the prior probability:

$$P(q_i|\Omega, \Theta^i) \approx p(\Omega|q_i, \Theta^i) P(q_i|\Theta^i) \quad (5)$$

By knowing the “initial” prior probability for each class, our classification method outputs the maximum a posteriori probability for each observed vector. To decide which is the most “likely” class  $q_i$ , the MAP decision rule 6 is used as follows:

$$\text{Object} \in q_i \text{ if } P(q_i|\Omega_k, \Theta^i) = \max(P(q_u|\Omega_k, \Theta^u)) \quad (6)$$

where  $u = 1, 2, \dots, N_{\mathcal{C}}$ , and  $N_{\mathcal{C}}$  is the total number of classes.

### D. Majority Voting (MV) Scheme Classifier

Ideally, the object classification should give every object a trustful classification on every scan. However, it does not happen because of the LRF sensor characteristics. For instance, a vehicle way go from a small line to two big perpendicular ones when performing the laser scan. One way to overcome this problem is to take into account as many features as possible from previous object detections, in particular dimensions and dynamics. Since we cannot classify an object with a high confidence immediately at the first scan when it appears, our MV method [2] is based on a voting scheme considering every hypotheses over time, until there is a high classification confidence. Each feature that characterizes the object represents a voter actor, where the weight of the vote depends on the

influence of the related feature to characterize the object and on the value  $v$  of that feature, given by the expression 7.

$$V(v) = \begin{cases} V_0 & v \leq L_0 \\ m_1 v + b_1 & L_0 < v < L_1 \\ V_1 & L_1 \leq v \leq L_2 \\ m_2 v + b_2 & L_2 < v < L_3 \\ V_2 & v \geq L_3 \end{cases} \quad \text{with } v \in \mathfrak{R}^+ \quad (7)$$

The confidence level is achieved by adding the votes of every actor (7). When some of the hypothesis reach a reasonable value, we assume that the object is classified as belonging to the type of that hypothesis.

#### E. The Posterior Probability

The expression used to estimate the posterior probability in the case of MV approach is a combination of the vote actor responses normalized by the total number of votes, yielding:

$$P(q_i | \Omega_k) \triangleq \frac{NV_i}{\sum_n V(n)_i} = \frac{NV_i}{NV_T} \quad (8)$$

where  $NV_i$  is the number of current votes, and  $NV_T$  is the total number of possible votes (8) for the category  $q_i$ .

Dividing the equation (5) by the marginal likelihood (normalized factor), the posterior probability for the GMM classifier is directly obtained.

## IV. VISION-BASED SYSTEM

The vision-based system uses *Haar-Like* features to extract the information from the image frames. The detection of the objects is performed using these features as an input to an AdaBoost classifier. Each detected object is then tracked in order to decrease the false positive rate and to maintain the objects labelled in the scene.

#### A. Haar-Like Features

Each *Haar-Like* feature is represented by a template, its coordinate relative to the search window origin and the size of the feature. A subset of the features prototypes used is shown in Fig. 2. The *Haar-Like* features values are calculated as a weighted sum of two components: the pixel gray level values sum over the black rectangle and the sum over the whole feature area. Hundreds of features are used in a real and robust classifier. To reduce the computational time it was used the concept of Integral Image introduced in [22].

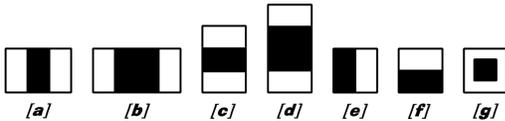


Fig. 2. Subset of the Haar-like features prototypes used in the object detection.  $a$ ,  $b$ ,  $c$  and  $d$  are the line features,  $e$  and  $f$  the edge features and  $g$  is the center-surround feature.

#### B. AdaBoost Classifier

A variant of AdaBoost [27] is used for selecting a small set of features and training each object classifier. The computation of the complete set of the features is unfeasible. The main goal here is to find a very small number of these features that can be combined to form an effective classifier. To support this purpose, a weak learning algorithm is designed to select the single feature which best separates the positive and negative examples. For each feature, the weak learner determines the optimal threshold classification function, so that the minimum number of examples are misclassified. A weak classifier  $h_j(x)$  consists of a feature  $f_j$ , a threshold  $\theta_j$  and a parity  $p_j$  indicating the direction of the inequality sign.

$$h_j(x) = \begin{cases} 1 & p_j f_j(x) < p_j \theta_j \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

In (9) the value 1 represents the detection of the object class and 0 represents a non-object. Each of these classifiers *per se* is not able to detect the object category. Rather, it reacts to some simple feature in the image that may be related to the object. An extended presentation of the boosting process is described in [22].

Another approach to reduce the computational time of the detection task consists in using a cascade of classifiers. With the cascade of classifiers is possible to achieve an increased detection performance, while radically reducing the computational time. The key objective is to construct smaller, and therefore more efficient, boosted classifiers which reject many of the negative sub-windows, while detecting almost all positive instances. Simpler classifiers are used to reject the majority of sub-windows before more complex classifiers are called upon, in order to achieve low false positive rates [7].

#### C. Detecting and Tracking the Objects

The detection of the objects is done by sliding a search window through the frame image and checking whether an image region at a certain location is classified as the object. Initially, the detecting windows is of the same size of the classifier ( $15 \times 37$  for pedestrians and  $30 \times 30$  for cars), then the window's size is increased by  $\alpha$  until the size of the window is equal to the image size. This process is repeated for each of the object class trained.

The objects detected by the vision system are then tracked with a Kalman Filter using a simple second order model. We do not need to have a robust tracker, because it is used just to reduce the false positive rate and to maintain the objects in the scene labelled. Since, most of the times, false positives are completely independent from the time sequence, the false positive rate decreases significantly. Moreover, the tracking of the global system is done in the laser space.

#### D. The Posterior Probability

To obtain the a posteriori probability of the AdaBoost classifier it was used the formulation called *Logistic Correction* [28]. Initially, in AdaBoost, each of the examples in a training set  $(x_i, y_i)$  have the same weight. At each step  $i$  a weak learner

$h_i$  is trained on the weighted training set. The error of  $h_i$  determines the model weight  $\alpha_i$  and the future weight of each training example. Assuming  $y_i \in \{-1, 1\}$  and  $h_i \in \{-1, 1\}$ , the output of the boosted model is given by (10), where  $T$  is the number of weak classifiers.

$$F(x) = \sum_{i=1}^T \alpha_i h_i(x) \quad (10)$$

It is shown that AdaBoost builds an additive logistic regression model for minimizing  $E(\exp(-yF(x)))$ . They show that  $E(\exp(-yF(x)))$  is minimized by:

$$F(x) = \frac{1}{2} \log \frac{P(y=1|x)}{P(y=-1|x)} \quad (11)$$

This suggests applying a logistic correction in order to get back the posterior probability:

$$P(q_i|\Omega_k) \triangleq P(y=1|x) = \frac{1}{1 + \exp(-2F(x))} \quad (12)$$

## V. COORDINATE TRANSFORMATION SYSTEM

The task of multi-sensor fusion leads to establishing a correspondence between the measurements gathered by distinct sensors. In this case, it is necessary to find a correspondence between the color camera reference and the LRF reference system. The coordinate transformation system in the proposed architecture (see Fig. 1) is used to calculate this correspondence. Thereby the result of the object's position and size estimation, in the LRF space, is used to construct a ROI in the image frame by means of a set of coordinate transformations. The ROI is formed in the image plane as the result of a correspondence between the objects under tracking (in the laser reference) and the objects in the image frame, in order to facilitate the process of segmentation and detection in the vision-based system and to decrease the computational time of the AdaBoost classifier.

The advantages of using a LRF in cooperation with the vision system are: simpler segmentation process and data processing; the laser measurements are not very sensitive under weather changes and consequently the whole system become more robust; the laser sensor has a good accuracy in the distance/depth measurements; vision systems are cheap and have a very detailed information of the surrounding.

### A. Calibration

The LRF and the camera are mounted in a common base, where their axes are parallel and "ideally" aligned. The calibration procedure is necessary to obtain a mapping expression to transform points in the laser reference system  $\{L\}$  to the camera reference system  $\{C\}$  and then to the image plane. Considering that the laser and the camera are parallel and aligned with one another, and using a flat target ("checkerboard") positioned at different distances to the laser-camera common base, the transformation between  $\{L\}$  and  $\{C\}$  was obtained under a quadratic error minimization criteria. Various images and laser measurements taken at different

positions of the target were used to estimate this coordinate transformation. The camera's intrinsic and extrinsic parameters were calculated using a Matlab Toolbox [29], therefore given the camera calibration parameters it is easy to map the points in the camera reference  $\{C\}$  to the image frame.

Once the range points, that define the segments detected by the LRF, are transformed to the camera reference frame and hence are projected in the image plane, a ROI is constructed in the image plane corresponding to each object that rely in camera's FOV. With the LRF information it is only possible to obtain the horizontal limit of the object in the image. If it is assumed that the vehicle/robot moves on a "flat" surface, it is geometrically easy to obtain the bottom limit of the ROI (resorting to the distance of the object). The top limit of the ROI can be estimated using the distance to the object and the maximum height possible in all categories.

## VI. COMBINING CLASSIFIERS

The classification module proposed is illustrated in Fig. 1. Input data from the camera and the laser sensor is used by different classifiers whose results are fused by means of another classifier in order to produce a unique decision rule. A classifier algorithm does not usually work suitably with raw data, then a vector of structured data (feature array) is used as aforementioned.

Inspired in [8] it was used a sum rule to combine (fuse) the classifiers outputs discussed previously. The decision rule uses the Bayesian framework to ultimate classify an object based on the outputs of each classifier.

Let us consider the number of classifiers as  $NC$ , and the feature vector used by the  $i$ th classifier denoted by  $\Omega_i$ . Let us assume that each class  $q_i$  is represented by a class-conditional probability density function  $p(\Omega_i|q_i)$  and its a priori probability of detection  $P(q_i)$ . Given the pdf and the a priori probability, the classical decision rule can be stated as (13):

$$\begin{aligned} \text{assign Object} &\longrightarrow q_j \text{ if} \\ P(q_j|\Omega_1, \dots, \Omega_{NC}) &= \max_k P(q_k|\Omega_1, \dots, \Omega_{NC}) \end{aligned} \quad (13)$$

Assuming that the features vectors are conditionally statistically independent, and that the posterior probability of each classifier do not deviate dramatically from the prior probability, after some mathematical formulations [8], a "practical" combinational Bayesian decision rule is stated as:

$$\begin{aligned} \text{assign Object} &\longrightarrow q_j \text{ if} \\ (1 - NC)P(q_j) + \sum_{i=1}^{NC} P(q_j|\Omega_i) &= \\ \max_{k=1}^{NC} [(1 - NC)P(q_k) + \sum_{i=1}^{NC} P(q_k|\Omega_i)] \end{aligned} \quad (14)$$

This "sum" decision rule depends on the prior probability of occurrence each class  $q_i$  and the posterior probabilities yielded by the respective classifiers. These last terms are critical, and its computation is subjected to some constraints depending of each classifier under consideration.



Fig. 3. Vehicle used in the experiments conducted in a outdoor scenario. The LRF and the camera are mounted on the front of the vehicle.

## VII. EXPERIMENTAL RESULTS

Some preliminary tests were conducted in an outdoor environment. Fig. 3 shows one of the vehicles used in the experiments. The sensory devices were mounted in the front of the vehicle approximately  $65\text{cm}$  of the ground. The system was tested by using data sequences (image frames and laser vectors) of pedestrians and cars in different size, pose and lighting conditions. This set consists of 962 frames with 753 labelled pedestrians and 337 cars. The Hit Rate (HR) and the number of False Positives (FP) of each classifier are shown in the table I. Two different forms of the AdaBoost classifier were implemented, namely AdaBoost 1 and AdaBoost 2. The first one was implemented in a way that a global search is done in all the image in order to detect the objects (i.e. without the laser information), and the later one integrates the laser information so as to construct a ROI around the object. As expected, the number of false positives in the second case decreased significantly with this cooperative approach. Figs. 4 and 5 show some detected objects, in the laser and in the image space, during the experimental trial.

In the vision based system, the classification performance of cars is better than that of the pedestrians. This happens because pedestrians are non-rigid bodies. In other words, the shape and size of a pedestrian varies greatly, and therefore the model of pedestrians is much more complex than that of rigid objects. For the laser based system, the classification results for car like category are, in general, less accurate than for pedestrian class. This is more evident when the detected cars appear to be static (parked) and far from the laser sensor.

## VIII. CONCLUSIONS AND FUTURE WORK

A multi-modal system for detecting, tracking and classifying objects in outdoor environment was presented. The objects of interest were restricted to pedestrians and cars. A cooperative technique was implemented in order to “fuse” the information from a ladar-based and a vision-based system. Results were obtained for each classifier separately and also for a classifier combination scheme, employed to improve the performance

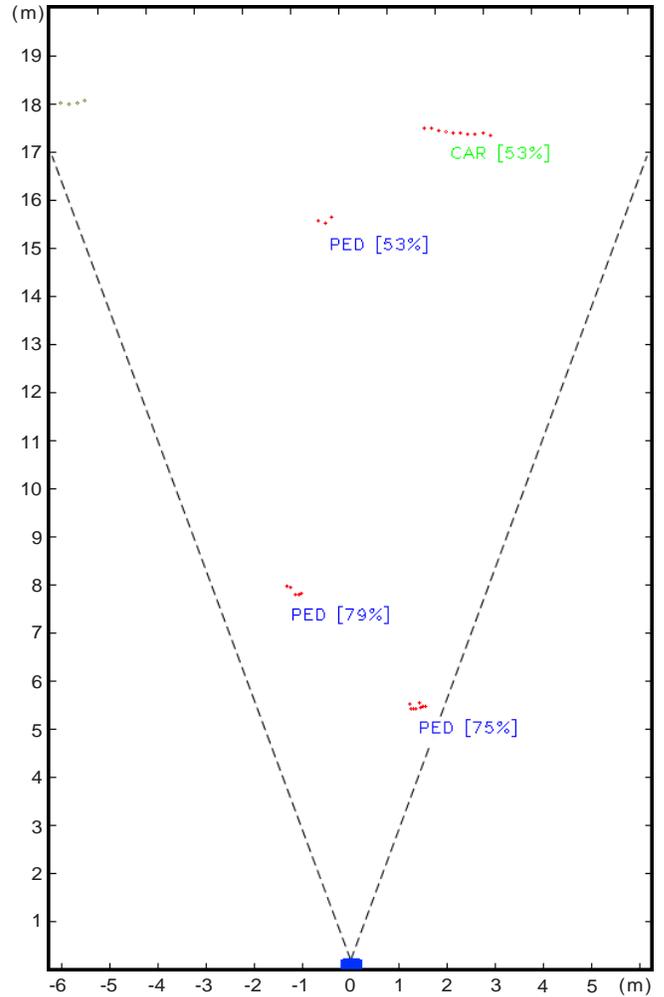


Fig. 4. Segmented objects in the laser space. Each detected object has its class and a posteriori probability given by the GMM classifier. The dashed lines represents the camera FOV.



Fig. 5. Output of the vision-based system in the same sample time, see Fig. 4. The vertical lines represent the ROI defined by the laser.

TABLE I  
CLASSIFICATION RATES

Classifier	Pedestrian		Vehicles	
	HR	FP	HR	FP
AdaBoost classifier 1	86.1	102	91.1	68
AdaBoost classifier 2	89.8	47	93.4	31
GMM classifier	90.3	17	89.9	28
Majority voting scheme (MV)	82.5	23	84.0	30
Combining classifier	92.4	9	93.8	11

of the system. This compound classification technique can be summarized as a practical Bayesian framework that combines, in a probabilistic sense, the a posteriori probability of each classifier by means of a sum rule. Some preliminary experiments, conducted on a large number of urban street scenes, demonstrate that the system is able to detect and classify pedestrians and cars in various positions, shapes, sizes and colors with a good accuracy. It was also demonstrated the feasibility of the proposed approach in a vehicle navigation in cybercar-like scenarios.

In a future work, and to improve the system, we intend to integrate more cameras in the system to increase the field of view and consequently the perception of the surrounding environment of the autonomous vehicle. Moreover, new methodologies/classifiers will be studied to improve the accuracy of the vision-system and of the ladar-system. It is equally planned to investigate new combinations of the classifiers to improve the global system performance.

#### ACKNOWLEDGMENT

This work was supported in part by Portuguese Science and Technology Foundation (FCT), under Grant POSC/EEA-SRI/58279/2004, and by CyberC3 project (European Asia IT&C Programme).

C. Premebida is supported by the Programme AlBan, the European Union Programme of High Level Scholarships for Latin America, scholarship n<sup>o</sup> E04M029876BR.

The authors gratefully acknowledge the suggestions, comments, and contributions from the following colleagues: Luciano Rebouças de Oliveira and João Luís da Silva Carreira.

#### REFERENCES

- [1] Cybercars. Cybernetic technologies for the cars in the city. [online], www.cybercars.org, 2001.
- [2] A. Mendes, L. C. Bento and U. Nunes, "Multi-target detection and tracking with laser range finder", in *IEEE Intelligent Vehicles Symposium*, Parma, Italy, 2004.
- [3] K. Dietmayer, J. Sparbert and D. Streller, "Model based object classification and object tracking in traffic scenes from range images", in *Proceedings of IV IEEE Intelligent Vehicles Symposium*, Tokyo, Japan, 2001.
- [4] A. J. Lipton, H. Fugiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video", *IEEE Image Understanding Workshop*, pp. 129-136, 1998.
- [5] J. Neira, J.D. Tardis, J. Horn and G. Schmidt, "Fusing range and intensity images for mobile robot localization", in *IEEE Trans. Robotics and Automation*, Vol. 15, No. 1, pp 76- 84, 1999.

- [6] C. Premebida, and U. Nunes, "A multi-target tracking and GMM-classifier for intelligent vehicles", in *9th International IEEE Conference on Intelligent Transportation Systems*, Toronto, Canada, 2006.
- [7] G. Monteiro, P. Peixoto and U. Nunes, "Vision-based pedestrian detection using Haar-like features", in *Proc. 6th National Festival of Robotics, Scientific Meeting (ROBOTICA)*, Guimaraes, Portugal, 2006.
- [8] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers", in *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 20, no. 3, pp. 226-239, 1998.
- [9] J. Vandorpe, H.V. Brussel and H. Xu, "Exact dynamic map building for a mobile robot using geometrical primitives produced by a 2D range finder", in *IEEE International Conference on Robotics and Automation*, pp901-908, Minneapolis, U.S.A., 1996.
- [10] J. Guivant, E. Nebot and H.F. Durrant-Whyte, "Simultaneous localization and map building using natural features in outdoor environments", in *Intelligent Autonomous Systems VI*, pp. 581-588, 2000.
- [11] G. A. Borges and M. J. Aldon, "Line extraction in 2D range images for mobile robotics", *Journal of Intelligent & Robotic Systems*, v. 40, n. 3, pp. 267-297, 2004.
- [12] D. Castro, U. Nunes, and A. Ruano, "Feature extraction for moving objects tracking system in indoor environments", in *Proc. 5th IFAC/euron Symposium on Intelligent Autonomous Vehicles*, Lisboa, Portugal, July 2004.
- [13] D. Streller and K. Dietmayer, "Object tracking and classification using a multiple hypothesis approach", *IEEE Intelligent Vehicles Symposium*, Parma, Italy, June 2004.
- [14] U. Franke and S. Heinrich, "Fast obstacle detection for urban traffic situations", *IEEE Trans. Intell. Transport. Syst.*, vol. 3, pp. 173-181, Sept. 2002.
- [15] D. Toyh and T. Aach. "Detection and recognition of moving objects using statistical motiondetection and Fourier descriptors." in *ICAP 03*. 2003.
- [16] P.Burt, J. Bergen et. al. "Object tracking with a moving camera: An application of dynamic motion analysis." in *Proc. IEEE Workshop Visual Motion*. 1989.
- [17] L. Zhao and C. Thorpe. "Stereo- and Neural Network-Based Pedestrian Detection." in *Trans. ITSC 2000*. 2000.
- [18] D. M. Gavrilu and V. Philomin. "Real-Time object detection for smart vehicles." in *Proc IEEE CV 99*. 1999.
- [19] H. Cheng, N. Zheng and J. Qin. "Pedestrian detection using sparse gabor filter and support vector machine." in *Proc. IEEE Intelligent Vehicles Symposium 2005*. 2005.
- [20] M. Bertozzi, A. Broggi et. al. "Pedestrian localization and tracking system with Kalman filtering." in *IEEE Intelligent Vehicles Symposium 2004*. 2004.
- [21] M. Oren, C. Papageorgiou and P. Sinha. "Pedestrian detection using wavelet templates." in *Proc. IEEE CVPR 1997*. 1997.
- [22] P.Viola and M. Jones. "Rapid object detection using a boosted cascade of simple features." in *IEEE CVPR 2001*, 2001.
- [23] C. Premebida, and U. Nunes, "Segmentation and geometric primitives extraction from 2D laser range data for mobile robot applications", in *Proc. 5th National Festival of Robotics, Scientific Meeting (ROBOTICA)*, Coimbra, Portugal, 2005.
- [24] Y. Bar-Shalom and X.R. Li, *Multitarget-Multisensor Tracking: Principles & Techniques*, YBS Publishing; 1995.
- [25] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley and Sons, New York, NY; 1997.
- [26] P. Paalanen, J. K. Kamarainen, J. Ilonen, and H. Klviinen, "Feature representation and discrimination based on Gaussian mixture model probability densities - practices and algorithms", *Research Report-95*, Lappeenranta University of Technology, Dep. of Information, 2005. (www.it.lut.fi/project/gmmbayes/).
- [27] Y.Freund and R. E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting." in *European Conference on Computational Learning Theory*, 1995.
- [28] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, 38(2), 2000.
- [29] Camera Calibration Toolbox for Matlab. [online], www.vision.caltech.edu/bouguetj/calibdoc/, 2006.