DETECTION AND CLASSIFICATION OF 3D MOVING OBJECTS

André Lourenço*, Pedro Freitas*, M. Isabel Ribeiro†, Jorge S. Marques†

Instituto Superior Técnico/Instituto de Sistemas e Robótica
Av. Rovisco Pais, 1
1049-001 Lisboa, PORTUGAL
fax: +351-21 8418291
e-mail: * arlpjaf@isr.ist.utl.pt
† {mir,jsm}@isr.ist.utl.pt

Keywords: Detection, Feature Extraction, Classification, Moving Objects, Laser.

Abstract

The detection and classification of moving objects is a key task in many control systems. This paper presents a solution for this problem based on a laser range scanner. The method can be summarized as follows. A set of 3D data points on the object surface is obtained using the scanner. A small number of features is then computed to represent the object boundary. In fact, three alternative features are considered. Classification algorithms are then designed and evaluated using real data. The experimental tests show that the proposed techniques allow a robust classification of moving objects based on range information.

1 Introduction

This paper addresses the detection and classification of 3D moving objects. This problem has a variety of applications e.g., in the context of quality control in industry or for the classification of vehicles in highways. For example, no verification of the vehicle shape and volume is currently performed by the automatic tax collector system "Via Verde" used in the Portuguese highways.

The solution proposed in this paper is based on a laser range scanner which is able to measure the location of 3D data points on the object surface. In a first step the output of the laser scanner is preprocessed to separate the object measurements from the background and to compute the 3D coordinates of the points on the object surface. Then, a small set of features is computed to characterize the object shape. Three sets of features are considered in this paper: volume/length, the object profile and the approximation of the object surface by a set of planes whose parameters are used for classification. A modified EM algorithm is used on the planar surface extraction, this being the main novelty of this paper. The feature vector is then used to classify the object using statistical and syntactic classification methods.

The paper is organized as follows. Section 2 presents the experimental setup and discusses the object detection procedure. In Section 3 the three types of features are described. Object classification and the experimental results with 3D moving objects are shown in Section 4. Section 5 concludes the paper and presents directions for further work.

2 Data Acquisition and Object Detection

This section describes the experimental setup used for the experiments, the object detection procedure from the acquired range data and the data correction to account for the object velocity.

The primary source of data is a laser range scanner from SICK OPTICS (LMS-200 indoor version) configured with a resolution of 0.5^o over 180^o, as represented in Figure 1. For each scan, the acquired

data is the set $\{(r_i, \theta_i), 0^o \leq \theta_i \leq 180^o\}$ where r_i is the range measurement corresponding to the scanning angle θ_i. With a high speed serial port (MOXA Card - 500kbaud), and using the protocol RS-422, acquisiton is done at a rate of 38 scans per second, with each scan containing 361 range measurements. This corresponds to a time duration of 26 msec for a scan.

The laser scanner is mounted in a fixed structure, pointing down and oriented in such a way that the plane spanned by the scan is vertical. Two frames are used along the work: the laser frame $\{X', Y', Z'\}$ and the world frame $\{X, Y, Z\}$ as represented in Figure 2.

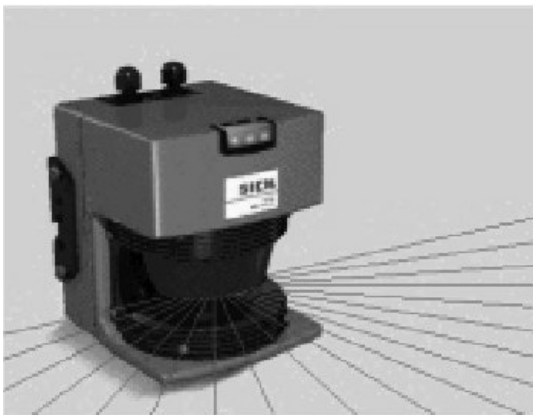

Figure 1: Sick Laser Range Scanner - LMS 200

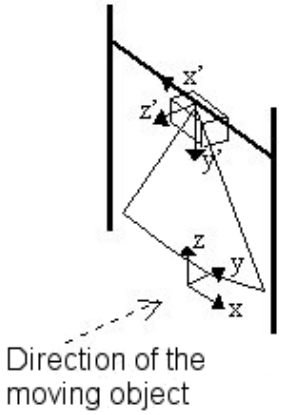

Figure 2: Laser frame and world frame

The shape of the acquired range data results from the crossing of the laser scanning plane by a moving 3D object and is a function of the object velocity. To relate successive range profiles from a 3D moving object it is necessary to have an estimate of the object velocity, $\hat{v}$. This is achieved by two infra-red sensors (S_1, S_2) and the corresponding reflector mirrors (R_1, R_2) installed in the geometric configuration represented in Figure 3 where d is the known distance between them. The velocity estimation is based on a time measurement. A PIC microcontroller (PIC16F84) measures the time elapsed between the obstruction of the two beams, uses this interval and the value of d to estimate the object's velocity and transmits its value to a PC via the parallel port. Furthermore, with the measurement of the time of obstruction of each beam ($\Delta t_i, i = 1, 2$) it is possible to have two estimates ($\hat{c}_i$) of the object's length:

$$\hat{c}_i = \hat{v}\,\Delta t_i, \qquad i = 1, 2. \tag{1}$$

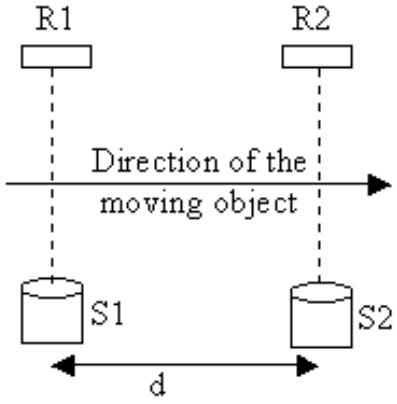

Figure 3: Velocity estimation

In the absence of objects, the laser scanner profile corresponds to the background environment and, apart from statistical variations, does not change. When a vehicle moves along, the range image is modified according to the object characteristics (e.g., shape). The object detection results from the comparison of the successively acquired range profiles with that of the background, previously recorded in the absence of any object. The comparison criteria is based on a statistical characterization of the background range data using, for each laser scanning an-

gle, θ_i, the mean μ_i and the standard deviation σ_i of background measurements experimentally evaluated. This statistical characterization takes place only once, before the objects move through.

A range measurement r_i, obtained for the scanning angle θ_i, is considered to belong to the background when it is within a $6\sigma_i$ neighborhood of μ_i. This relatively high threshold was chosen because the points in the limits of the range image have a very large variation. On the contrary, a point measurement is considered to belong to an object if $r_i < \mu_i - 6\sigma_i$, i.e., it does not belong to the background and lies between the sensor device and the background. If none of these conditions is satisfied, the point is considered to be an outlier and discarded. This object detection procedure is repeated for each data profile.

Figure 4 displays the result of the object detection procedure for a single scan. Data was converted from the polar coordinated frame associated with the sensing device to the frame $\{X^{'}, Y^{'}\}$. The raw data corresponding to the backgroung is represented by the dot points while the points of the object are displayed with diamond marks. The almost rectangular shape of the background results from the lateral columns of the metalic support where the laser is mounted (see Figure 2).

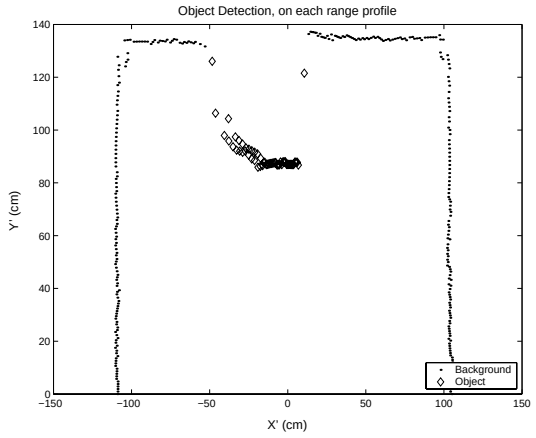

Figure 4: Detection of object and background for a single scan

The range data collected in consecutive laser scans has a spatial distribution that corresponds to the shape of the object that moves below the scanner. This shape, in particular along the $Z^{'}$ axis, can be recovered by applying the relationship between the duration and the number of points in a scan and the object velocity.

Let t_{scan} be the time duration of each scan, t_{scan0_j} the initial time instant of scan j, N_{points} the number of points in each scan and $\hat{v}$ the estimate of the object velocity. Then, the relative translation along the $Z^{'}$ axis of point i of scan j, $\Delta Z'(i)_j$, is obtained as:

$$\Delta Z'(i)_j = \hat{v}\left(\frac{t_{scan}}{N_{points}-1}i + t_{scan0_j}\right), i = 0, ..., 360 \quad (2)$$

Figure 5 displays the raw range data after background removal and velocity correction for the case when the object represented in Figure 6 moves below the laser scanner with a velocity of $25cm/sec$. Data was converted from the acquisition frame $\{r, \theta\}$ to the laser frame $\{X^{'}, Y^{'}\}$ with $Z^{'} = 0$ followed by the velocity correction given by (2). A final conversion from $\{X^{'}, Y^{'}, Z^{'}\}$ to $\{X, Y, Z\}$ is performed. The motion of the paper boxes in 6 is produced by a mobile robot Nomad scout disguised with the boxes.

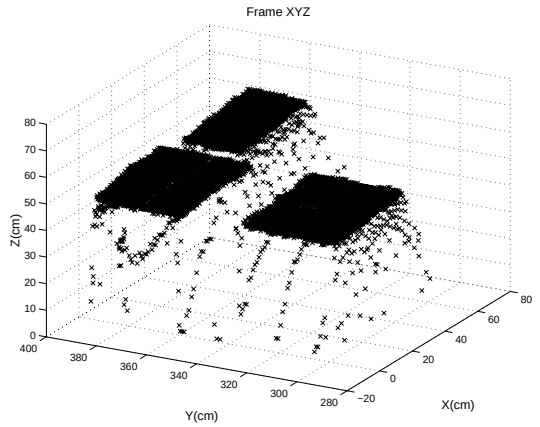

Figure 5: Raw data representation in the world frame after velocity correction and background extraction

The comparison of Figures 5 and 6 leads to the conclusion that the object detection process was successful, since the shape obtained in the 3D reconstruction is similar to the real one. Furthermore, the estimated dimensions of the object are approximately equal to the real ones as estimated in Section 3. The top surfaces of the moving object are clearly recognized in Figure 5 while its lateral planar surfaces are not displayed. Instead, at the weight of these surfaces a cloud of sparsely distributed points

Figure 6: Moving object

are present. These points may result from the Mixed Point Problem, [6], that occurs when the laser footprint (resulting from the intersection of the cylinder like laser beam with a surface) does not lie in a single surface. When the laser beam is directed towards the object edge, the footprint is partially in the top surface and partially in the floor leading to a range measurement at a random height.

3 Feature Extraction

The classification of moving objects is based on features extracted from the raw range data. This section discusses the extraction of three different feature vectors: Longitudinal View, Volume/Length and Planar Surfaces.

The availability of different types of features will support the choice of the best classifier as described in Section 4 as well as leading to an easier visualization of the results.

3.1 Longitudinal view

This feature is extracted directly from data as it corresponds to the longitudinal profile of the moving object. Fixing a emission angle of the laser, usually $\theta = 90^o$ that corresponds to $X^{'} = 0$, the longitudinal view is the vector that contains the distances acquired over the object, in X'Y'Z' frame, along that angle. The number of range points in this feature is a function of the object length and also of its velocity. Aiming at using this feature for object classification, it will be useful to render it independent of the object length and thus F equally spaced points are chosen along the referred raw data.

Figure 7 shows the feature extraction over the object in Figure 6 for the case of $F = 11$. In this last figure, the real longitudinal view is marked, on top of the real photograph, by a black solid line.

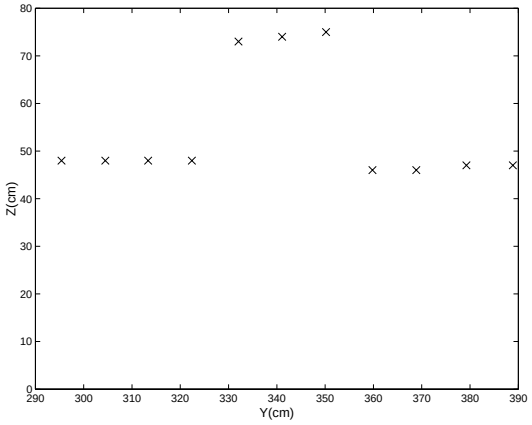

Figure 7: F point feature vector of the object in the world frame

Figure 7 displays three different distances to the sensor corresponding to the three top surfaces of the object.

3.2 Volume/Length

Volume and length characteristics emerge as natural features since they have a physical interpretation.

The length estimate, $\widehat{c}$, is the mean of the two estimates given by (1), presented in Section 2.

The volume, $\widehat{V}$, is iteratively computed using a trapezoidal approximation of the area in each scan. The area associated to each scan is estimated by

$$A = \sum_{i=0}^{359} A_{i+1} \tag{3}$$

with

$$A_{i+1} = (x_{i+1} - x_i)\frac{z_{i+1} + z_i}{2}, i = 0, ..., 359$$

where the index i defines the emission angle of the laser, θ_i. The volume estimation is then obtained, assuming a constant height between two consecutive scans, j and $j + 1$:

$$
\begin{aligned}
V_j &= A_j(y_{j+1} - y_j), j = 1, ..., N_{scan} - 1 \\
\widehat{V} &= \sum_{j=1}^{N_{scan}-1} V_j
\end{aligned} \tag{4}
$$

where N_{scan} is the total number of scans over the object.

The object in Figure 6 has length $\widehat{c} = 1.0234m$ and volume $\widehat{V} = 0.327m^3$, which are close to the real values. The number of scans over the object is $N_{scan} = 143$.

3.3 Planes

The reconstruction of 3D objects has been the subject of recent research, [6], and different methodologies have been used. Some methods are based on the representation of the object boundary e.g., using splines, polygonal meshes or planes. Surface approximation by a set of planes is used in this paper. This choice is appropriate for representing several classes of objects (e.g., vehicles). The planes coefficients are considered as features for object classification.

Consider a set of 3D points $x_i \in R^3$ belonging to a plane. They all verify the equation

$$x_i^T n = d \tag{5}$$

where $n \in R^3$ is the unit normal and $d \in R$ is the distance from the plane to the origin.

To estimate the plane parameters $\theta = (n, d)$ it is assumed that each x_i is a realization of a random variable x with normal distribution $N(\mu, \Sigma)$, where μ is the mean vector and Σ the covariance matrix. Furthermore, since the data belongs to a 2D subspace (plane) the covariance matrix has rank 2 and the unit normal n is defined by the null subspace of Σ.

Using the mean square error method, the plane parameters are obtained by minimizing

$$\varepsilon^2 = E\{\|d - x^T n\|^2\} \tag{6}$$

Assuming that $\|n\| = 1$, ε^2 is minimum when n is the eigenvector associated with the smallest eigenvalue of Σ and $d = E\{x\}^T n$.

This method allows to approximate a set of data points $\{x_i\}$ by a plane. In practice a single plane is not enough to describe the object boundary. Therefore, multiple planes must be considered. In this case, the estimation problem becomes harder.

The estimation of multiple planes is related to the estimation of a mixture of M probability density functions, each of them being represented by a normal distribution $N(\mu_i, \Sigma_i), i = 1, ..., M$. The Expectation Maximization (EM) algorithm has been used to estimate the mixture parameters [1], [4].

The EM algorithm is an iterative method based on two steps. The E-step computes the probability of each of the components being active, assuming that a data point x_i is observed and an initial estimate of the mixture parameters Θ' is available

$$w_i(l) = \frac{\alpha_l' p_l(x_i \mid \theta_l')}{p(x_i \mid \Theta')} \tag{7}$$

where $w_i(l) = P(y_i = l \mid x_i, \Theta')$ and yi denotes the active component associated to xi.

The M-Step updates the mixture parameters as follows

$$
\begin{aligned}
\mu_l &= \frac{\sum_{i=1}^N w_i(l) x_i}{\sum_{i=1}^N w_i(l)}, \\
\Sigma_l &= \frac{\sum_{i=1}^N w_i(l)(x_i - \mu_l)(x_i - \mu_l)^T}{\sum_{i=1}^N w_i(l)}
\end{aligned} \tag{8}
$$

The plane parameters are easily computed using μ_l, Σ_l as before.

The application of the EM algorithm in surface approximation raises several difficulties. First, the number of planes is unknown and has to be estimated from the data. Second, the EM algorithm often converges towards local minima i.e., some of the components approximate data from two or more planes or sometimes a single plane is represented by more than one mixture component. To overcome these difficulties a more complex procedure is adopted in this paper based on 3 steps: i) modified EM algorithm; ii) model separation and iii) model fusion.

The first step provides a first estimate of the plane parameters, avoiding the representation of multiple

planes by a single component. This step is briefly described bellow adopting an initial number of components M=2.

Modified EM algoritm

1. Mixture parameters are initialized by randomly selecting a set of data points for each class;
2. Update the mixture parameters using equations (7)-(8);
3. Test if the smallest eigenvector of one of the covariance matrices is smaller than a threshold:
   - YES: Go to Step 4;
   - NO: Go to Step 5;
4. Save the mixture components which meet the criterion. Eliminate from the data set the points which are well represented by the saved planes. Go to Step 6;
5. The method has converged and the smallest eigenvector is greater that the threshold?
   - YES: Return to Step 1, creating one more model: $M = M + 1$;
   - NO: Go to Step 2;
6. Does the model describe 95% of the initial data points?
   - YES: Terminate the algorithm;
   - NO: Go back to Step 1;

Figure 8 shows the results obtained with the modified EM algorithm, applied to the object in Figure 6. In this Figure several mixture components are used to represent the data points on the highest plane, each one corresponding to one grey level.

Model separation

Step 1 approximates the data points by a set of planes. However, it does not consider the spatial distribution of the data. For example, two separate regions may be represented by a single model if all the data points belong to the same plane. However, this is not convenient for classification purposes. We

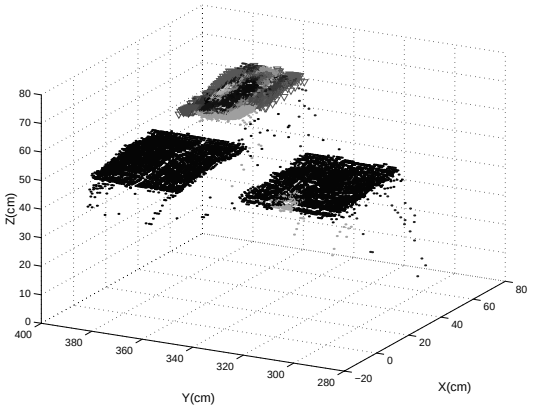

Figure 8: Result of EM algorithm

wish to have disconnected regions modeled by different components. A plane separation algorithm was used as described bellow (Step 2):

1. Find the plane with the highest eigenvalue (highest variance);
2. Sort the original data in the direction of highest variance and sort the data by projecting it in this direction;
3. Check for each model computed in the EM algorithm if the data represented by the model can be considered as a connected set;
4. Separate the data in different models if the data is not connected;
5. Repeat this algorithm, finding another vector in the plane orthogonal to the vector previously found.

Figure 9 illustrates the result obtained with the proposed separation technique. In Figure 8, two separate regions with the same height had been considered as belonging to a single plane. The separation algorithm splits them as shown in Figure 9 where different grey levels represent different planes.

Model fusion

Some of the mixture components estimated in Steps 1, 2 represent data from the same region and have similar parameters. The third step performs component merging as follows:

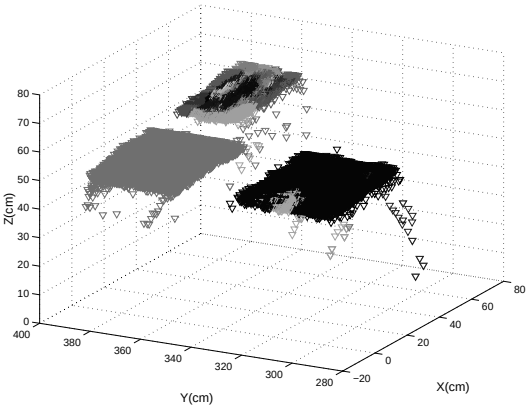

Figure 9: Result of the plane separation algorithm

For each two models m_1, m_2, test until $m_1 = N_{models} - 1$:

1. Check if the m_1, m_2 models are similar using the Euclidean distance
   - if the models are distinct, go to Step 3.
   - if the models are similar, go to Step 2.
2. Check if the m_1 and m_2 models were created by the separation algorithm:
   - YES: do not merge the models and go to Step 3.
   - NO: replace the two models by a single one, with parameters:

$$d = \frac{N_{m_1} d_{m_1} + N_{m_2} d_{m_2}}{N_{m_1} + N_{m_2}} \quad (9)$$

$$n = \frac{N_{m_1} n_{m_1} + N_{m_2} n_{m_2}}{N_{m_1} + N_{m_2}} \quad (10)$$

   where N_{m_i} is the number of points described by the i^{th} model, d_{m_i} and n_{m_i} are the distance to the origin and the normal vector of the plane, respectively, with $i = 1, 2$. Go to Step 3.
3. $m_2 = N_{models}$?
   - YES: $m_1 = m_1 + 1$, $m_2 = m_1 + 1$ and go back to Step 1.
   - NO: $m_2 = m_2 + 1$ and go back to Step 1.

Figure 10 illustrates the output of the model fusion. It is clear that all the very similar planar models represented in Figure 9 and corresponding to the highest surface of the object were merged in a single plane.

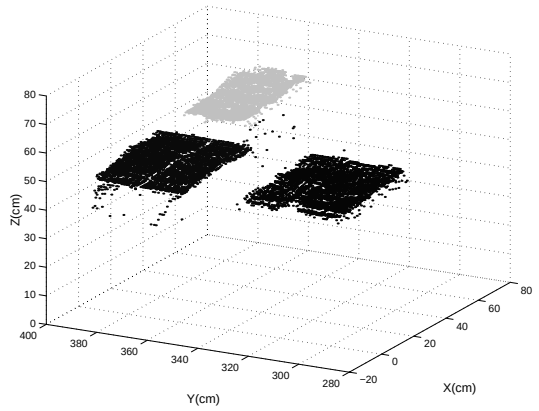

Figure 10: Results after mode merging

4 Results

The proposed features were used to classify 3D moving objects from measurements obtained using the laser range scanner. Six types of objects were considered in these experiments. The objects were obtained by covering a Nomad scout mobile robot with boxes of different sizes. The data set used in these experiments consists of 30 range sequences of each object to be classified.

Several classification algorithms were considered [2], [5]. In the case of the volume/length parameters and longitudinal view, the following methods were used: Bayes classifier, k-nearest neighbor (KNN) and Parzen method. The Bayes classifier was used assuming that the feature vector is a random variable with normal distribution. For this classifier, two different parameter estimation were used: one, named gauss1, estimates the mean vectors and covariance matrices for each class of the training set, using the maximum likelihood method; another, named gauss2, estimates the mean vectors for each class and one covariance matrix for all the training set. The KNN and Parzen classifiers are non parametric methods that make no assumption about the distribution of the data. The KNN classifier was implemented with k=7, for volume/length parameters and k=1 for longitudinal view, and Gaussian win-

dows were used in Parzen method.

The classification of 3D objects using the plane parameters was performed adopting a syntactic classifier, [3], based on a set of grammars obtained using the Crespi-Reghizzi algorithm and an error corrector parser for classification. This method assumes that the patterns are symbolically described. This was achieved using the k-means clustering algorithm which converts the planes parameters into a sequence of symbols.

Feature / Classifier	volume/ length	longitudinal view	planar surfaces
Gauss1	0.0857	0.1620	-
Gauss2	0,0714	0.007	-
KNN	0.0571	0	-
Parzen	0.1643	0	-
Syntactic	-	-	0.100

Table 1: Classification Errors

Table 1 shows the classification errors achieved by each method in these experiments. The classification error was estimated using the leave one out method. The best results (0%) were obtained with the KNN classifier and by the Parzen classifier using longitudinal views. Comparable results were also obtained by one of the Bayes classifiers. The planar description of the objects and the syntatic classifier lead to acceptable results, similar to those obtained with volume/length description, but the process is far more complex than the others.

These results show that the proposed system is able to accurately classify moving objects using 3D shape information. The best features were the longitudinal view.

5 Conclusions

This paper describes methods for the detection and classification of 3D moving objects using a laser range scanner. Three types of features were considered (volume/length, longitudinal view and surface approximation by a set of planes). Different types of classification strategies were used namely the Bayes classifier, the K-nearest neighbor method, the Parzen method and a syntactic classifier based on grammars. The best results were achieved with longitudinal view using the KNN and Parzen classifiers.

The methods described in this paper can be used in several applications namely for the automatic classification of vehicles in highways. This is an important issue since no verification of the vehicle size and volume is currently performed in Portuguese highways namely in Via Verde system.

The work presented in this paper can be extended by the use of two laser sensors aiming at a complete 3D reconstruction of moving objects.

6 Acknowledgements

We would like to thank Prof. Ana Fred, who allowed us to use the CSA application for grammar inference and syntactic recognition purposes.

We cannot forget the support provided by Mr. Christian Brenneke of the Institut für Steuerungstechnik & Zentrum der Didaktik der Technik - Hannover who helped with the Sick Laser driver and Moxa Card configuration.

References

[1] Dempster, A. P., Laird, N. M., Rubin, D. B., *Maximum Likelihood from incomplete Data via EM Algorithm*, Journal of the Royal Statistical Society,Series B, Volume 39, Issue 1, pp 1-38, 1977.

[2] R. Duda, P. Hart, Stork, *Pattern Classification*, 2nd ed., Wiley, 2000.

[3] Fred, Ana, *Structural Pattern Recognition: Applications in Automatic Sleep Analysis*, PhD Thesis, Instituto Superior Técnico, November 1994.

[4] Bilmes, J. A., *A Gentle Tutorial of the EM Algorithm and its Applications to Parameter Estimation for Gaussian Mixtures and Hidden Markov Models*, International Computer Science Institute Berkeley, CA 94704, and Computer Science Division, Department of Electrical Engineering and Computer Science, U.C.Berkeley TR-97-021, April 1998.

[5] Marques, Jorge Salvador, *Reconhecimento de Padrões: métodos estatísticos e neuronais*, IST Press, September 1999.

[6] Sequeira,V., *Active Range Sensing for Three-Dimensional Environment Reconstruction*, Ph.D. Thesis, Instituto Superior Técnico, December 1996.