Proceedings of the
2007 IEEE Intelligent Vehicles Symposium
Istanbul, Turkey, June 13-15, 2007

A Tracking System for Automated Inventory of Road Signs

S. Lafuente-Arroyo, S. Maldonado-Bascón, P. Gil-Jiménez, J. Acevedo-Rodríguez and R.J. López-Sastre
Dpto. Teoría de la Señal y Comunicaciones
Universidad de Alcalá.
Alcalá de Henares, Madrid, Spain
{sergio.lafuente,saturnino.maldonado,pedro.gil,javier.acevedo,robertoj.lopez}@uah.es

Abstract— **This paper describes the tracking stage of a road sign identification system with the use of a Kalman filter estimator. Although traditionally traffic sign recognition systems work in a single-frame way, tracking using inter-frame information is crucial in order to reduce false alarms. In applications that involve maintenance systems, the purpose of tracking is to give a single stroke of each road sign although it is presented in several frames of a sequence.**

The tracking module uses the output of the recognition stage, which processes each frame independently. Each traffic sign is followed for a specific distance which is computed in each case based on the system parameters. This setup has permitted improve the accuracy of the performance with respect to the obtained at the output of the recognition system. Additionally most false alarms are eliminated because their characteristic are temporally incoherent.

I. INTRODUCTION

The average of people who die each year in the developing countries as a result of traffic accident is considerable. As an example, according to the USDOT IVI program, in the US, more of than 42,000 Americans die each year as a result of 6.8 millions accidents (http://www.itsdocs.fhwa.dot.gov/index.htm). For this reason, automatic road sign detection systems has been an important issue for research recently and suppose a substantial investment of public money in road infrastructure. In recent years, some of highway agencies have been conducting sign inventory as part of their management program. Particularly, when traffic sign data is coupled with location information through the use of Distance Measurement Instrument (DMI) and differential Global Positioning System (GPS) receiver, a comprehensive sign inventory can be established relative to sign content, sign condition, and sign positioning with resolution scale of one to several meters. However, the current level of automation in sign detection and recognition, size dimensioning of sign, and location identification is not satisfactory due to a process of elaborate manual intervention is required. As sign recognition is a critical task of automated sign inventory, this paper covers a tracking state of road signs in images whose purpose is to improve the automation level of sign inventory.

Most efforts during the past few years have focused on the development of computational algorithms to detect road signs in each independent frame. Many works divide the algorithms into two stages, detection and recognition. There are two main approximations to detect traffic signs: those based on color criteria [1]–[3] and those that employ a border detection [4], [5]. At recognition stage there are many solutions, such as techniques based on different Neuronal Networks [6], [7], Support Vector Machines (SVM) [8], and genetic algorithms [9], where no tracking technique is used to improve the performance.

The detection of road signs using only a single image has two problems: 1) the correctness of road signs is hard to verify; and 2) it is difficult to detect correctly a traffic sign when temporary occlusion occurs. By using a video sequence, we preserve information from the preceding images, such as the number of detections of road signs and their corresponding sizes and positions projected in the image. This information increases the accuracy of road sign detection in subsequent images. Moreover, information supplied by later images is used to assist in the verification of detection of traffic signs, so that detected and tracked objects that are not road signs can be eliminated as soon as possible. Thus, using inter-frame information the amount of false alarms will be reduced and we will manage more valuable information for road sign detection than using single images. However, through the tracking algorithm, the candidate region of a road sign can be predicted based on the previous image frame. The methodology described in this paper allows us to track multiple signs separately in a sequence of image frames.

Recently, a complete system to detect traffic signs including a tracking algorithm was presented in [6], with two important limitations: a) the speed of the vehicle must be constant and known, and b) the algorithm assumes straight trajectories. In the mentioned work, both the prediction of the radius of the road sign in the image and the position of the sign (vertical and horizontal coordinates) in the image plane are estimated.

In this paper we focus our work in the tracking process where the visibility distance of each sign is estimated. We also consider the movement of the centroid of the objects in order to perform the tracking.

II. SYSTEM OVERVIEW

Results of our whole system for detection and recognition can be found at http://roadanalysis.uah.es/Documentos/Results and an exhaustive description of our system can be found in [10]. It allows us to extract possible signs from the image, and classify the candidate objects into specific type categories

1-4244-1068-1/07/$25.00 ©2007 IEEE.

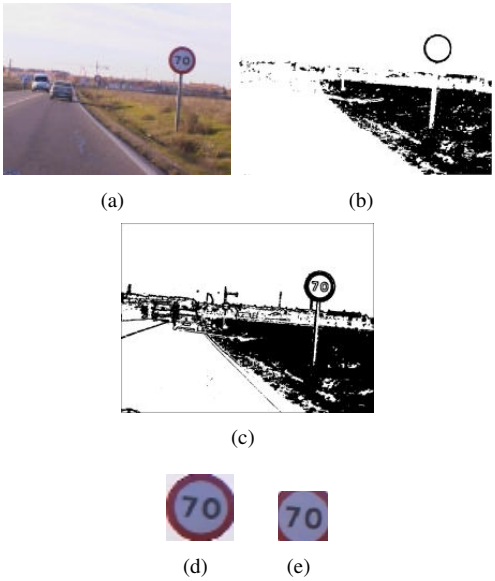


Fig. 1. Example of segmentation process. (a) Original image; (b) segmentation mask by red; (c) segmentation mask by achromatic decomposition; (d) traffic sign detected by red contour; (e) traffic sign detected by white inner area.

such as speed limit or stop. The system consists of the following steps:

- Segmentation
- Shape classification
- Recognition

A. Segmentation

Color information is considered to extract candidate objects from the input image by thresholding. For this purpose, HSI color space is employed for chromatic signs. The advantage of this color model is that two components (Hue and Saturation) encode the color information, being strongly robust against lighting conditions variations. At the same time, white signs are detected with the help of an achromatic decomposition in a similar way that in [11]. Then, a connected-component operator is applied to label the interesting objects. Most common road signs present a red rim and an inner white region. This characteristic leds us to consider the sign as a possible sum of two contributions corresponding to their chromatic and achromatic segmentation masks (see Fig.1) where both parts are processed independently in the complete system. The advantage of this idea is that a sign can be detected by different colors. Obviously, some limits can be imposed respect to the size or aspect ratio. In other words, small regions or big regions are eliminated.

B. Shape classification

All objects from the segmentation are classified in this stage using linear SVM. For extraction feature vectors two methods have been implemented based on distance to borders vectors and the signature of objects. According to the color which has been used in the segmentation, only some given shapes can be expected. For example, objects segmented using red clues will be circular, triangular or octagonal. Before classification step, all objects are oriented in a reference position and thus, the evolution of the geometric shape-vectors is so similar in most cases.

C. Recognition

This step is implemented by SVM with gaussian kernel where the input vector is a normalized size block of 31 x 31 pixels in gray-scale image for each candidate blob. In order to reduce the dimensions of feature vectors, only those pixels that must be part of the sign (pixels of interest) are computed. Different one-vs-all SVM classifiers with a gaussian kernel are used so that the system can recognize every sign. Both the training and test are done according to the color and shape of each candidate region. Thus, every object is only compared with those signs that have the same color and geometric properties than the blob to identify.

The amount of training samples per class varies between 20 and 100. We use an average of 50 training patterns for each class, but only some of them define the decision hyperplane as support vectors. Of course, in the training set are included samples of noisy objects that could be confused with traffic signs by this recognition module. Searching for the decision region, all feature vectors of a specific class are grouped together against all vectors corresponding to the rest of classes (including here noisy objects). Due to the size normalization of each blob, the method is invariant to scale changes and, on the other hand, since all interesting objects are oriented in a reference position we can conclude that the system is strongly robust to habitual rotations of road signs. The results for two video sequences using a single-frame detection are illustrated in Fig.2-3, where all candidate objects are drawn over the image with their respective geometrical shape. In Fig. 4 we can see a sequence in which the sign is not detected in some frames due to difficult segmentation to isolate the sign from the sky.

III. Road Sign Tracking

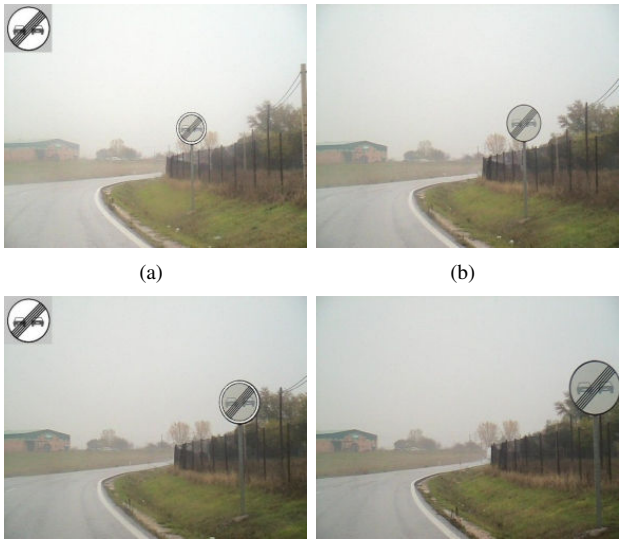
Recognition and tracking objects with a CCD camera is a non trivial problem because there exist relative motion between the camera, the static objects (in our case, traffic signs) and the environment. On the other hand, the system has to be able to track many targets simultaneously. As we approach to traffic signs, these can be considered, from a camera point view, as a dynamic system whose behaviour can be estimated by a cinematic model. Using smoothing and prediction filters we can estimate sign parameters between successive frames of a sequence. To this end, the Kalman filter [12] and the Alpha-Beta-Gamma ($\alpha-\beta-\gamma$) filter [13], as a special case of the previous one, are a fundamental tool.

It is assumed that the optical axis of the camera is roughly horizontal and the motion of the camera is moving along its optical axis. This assumption is often true in real world settings. Particularly, a camera is mounted on the vehicle and its optical axis is calibrated to be parallel to the horizontal plane of the vehicle.

(a) (b)

(c) (d)

Fig. 2. Results of a video sequence S1 at recognition output.



(a) (b)

(c) (d)

Fig. 3. Experimental results of a video sequence S2. Examples of recognition process including miss detection in (b,c).

A. Methodology Description

Once the system recognizes a traffic sign in the incoming frame, two options are possible. If there is no correspondence between the new object and the previous ones, a new track process is initiated. Otherwise, the track data structure which contains the objects to track is updated regarding to the new information. This ensures that sequential detections are processed together to estimate the parameters of the object. It is important to point out that, at least, N detections are required to consider the object as a traffic sign.

Information extracted by the global recognition system related to each object, such as position coordinates and size in the image plane, color and type category, constitutes the components of the track process and are stored as the object signature. The tracking system must decide whether each candidate blob from the current frame is a new object or an object previously detected in earlier frames of the sequence. In order to perform this task, the tracking module uses association algorithms based on the features of signature. In other words, in this process each new detection is compared with all previous objects to prevent the system from redundant tracks according to:

- Segmentation color: red, blue, yellow and white. To establish a correspondence between two objects, both candidate objects must be detected by the same color. In addition, according to Spanish road signs, a correspondence can be fixed for objects detected by red contour and those ones detected by their white inner area.
- Geometric shape: triangular, circular and rectangular. To associate a new object with an old tracked object, both objects must be detected with the same shape.
- Position in the image plane: the location of a new traffic sign must be closed to the predicted position of the centroid of a previous detected object. In this case, the tracking stage will consider the new object as a child of the parent object.
- Dimensions of the blob: the dimensions of each object in the image plane increase as the capture system approaches it.

All observations whose position error, defined as the difference between the actual position and the predicted position of a given track, are lower than a maximum threshold, are associated to that track. For each new object a new track process is initiated. Since new objects are compared with all existing track processes, a matrix defines the association between the new detections and all existing track processes. In cases where a new observation is associated with more than one track process, a criterion based on the euclidean distance between the new object position and the predicted position of all existing track objects is used to establish a single tracking.

The goal of this work is to study how the proposed algorithm based on probabilistic prediction techniques is able to track traffic signs. The steps to use a predictive filter for vision tracking are:

1) **Initialization**. In this step objects of interest are searched in the whole image since we do not know the objects position. It is important to point out that some researchers reduce the research window in order to reduce the computation time. Nevertheless, traffic signs do not appear always in the usual position (i.e. lateral margin of the road) and our system must be invariant to possible shifts.
2) **Prediction**. In this stage the predictive filter estimates the position of the object at time t+1, which in our case is given by the two points that define the bounding-box rectangle of the object.

600 mm	900 mm	1200 mm
21.93 m	32.90 m	43.87 m

TABLE I
MAXIMUM DISTANCE OF TRACKING FOR NORMALIZED SIZE SPANISH TRAFFIC SIGNS

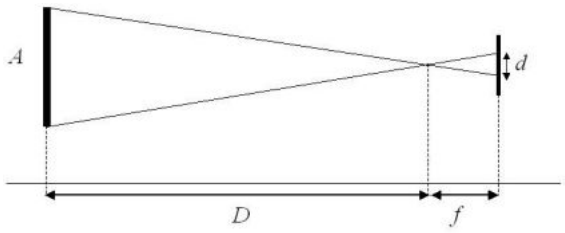


Fig. 4. Pin-hole model of the camera for evaluation of maximum tracking distance.

3) **Correction**. In this part the detection and recognition system detects the candidate objects, which should be in the neighborhood of predicted points in the previous stage, and we use their actual position (measurement) to carry out the state correction.

Several cases may happen while the system processes a video sequence. When a new object is detected while no corresponding object exists in the past, it can alert a new traffic sign is present in the field of view. If this new region can be tracked successfully for several frames, it will then be considered as a new sign with only one label assigned and a predictive filter is initialized to track this new road sign in the ensuing frames.

The steps of prediction and correction are carried out for a given sign while the sign is supposed not to be overcome and at this point, the information offered by GPS is considered. Using a pin-hole model of the camera, as we can see in Fig.4, the tracking system computes for each new sign the maximum distance for which it will be tracked, D, according to the following expression:

$$D = \frac{Af}{d} \tag{1}$$

where A is the actual diameter of the sign, f is the focal distance and d is the diameter of the sign in the plane image, which is computed as the product between the size in pixels and the unit cell size of CCD camera. According to Spanish normative [14] three sizes are possible for road signs: 1200 mm, 900 mm and 600 mm. Thus, if we consider that the minimum size for a candidate traffic sign to be detected has been fixed in our global system to 50 pixels, the unit cell size for our camera is 4.65 μm x 4.65 μm and the focal distance is 8.5 mm, the maximum distances of tracking are summarized in Table 1.

Finally, since a interesting object can be identified with different class types by the recognition module in a video sequence, we define a criterion to assign the final type to each sign using a weighting function. Thus, the weight associated to each possible *i*-class is given by:

$$K_i = \frac{N_i}{1 + \sum_k D_k} \tag{2}$$

where, N_i defines how many times the object is identified with class *i* in the sequence and D_k represents the distance between the index of each frame of the sequence in which the object is assigned to class *i* and the index of the last frame of the sequence. The summatory function of distances in the denominator considers the fact that the recognition process is normally more accurate in the last images than in the first ones of a sequence, except in cases of partial occlusion. The class type assigned is the corresponding to the highest weight.

B. Object Tracking

The $\alpha - \beta - \gamma$ filter is used in this work to model the relative motion between the camera and the road-sign in the scene. The Kalman filter as an optimal stochastic filter is used to estimate the motion parameters in two stages: prediction and correction and in our case, the state vector include position, velocity, and acceleration. Suppose that the centroid of the object is described as $\mathbf{p}(t) = [x(t), y(t)]^T$. The state vector for each point is $\mathbf{x}(t) = [\mathbf{p}(t), \dot{\mathbf{p}}(t), \ddot{\mathbf{p}}(t)]^T$ where $\mathbf{p}(t)$ represents the position, $\dot{\mathbf{p}}(t)$ the velocity and $\ddot{\mathbf{p}}(t)$ the acceleration of the centroid that defines the tracked object. By using Taylor series expansion, the state prediction equation is:

$$\mathbf{x}(t^-) = \mathbf{A}\ \mathbf{x}(t - \Delta t) + \mathbf{w}(t) \tag{3}$$

$$\begin{bmatrix} \mathbf{p}(t^-) \\ \dot{\mathbf{p}}(t^-) \\ \ddot{\mathbf{p}}(t^-) \end{bmatrix} = \begin{bmatrix} \mathbf{I}_2 & \Delta t\mathbf{I}_2 & \frac{\Delta t^2}{2}\mathbf{I}_2 \\ \mathbf{O}_2 & \mathbf{I}_2 & \Delta t\mathbf{I}_2 \\ \mathbf{O}_2 & \mathbf{O}_2 & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{p}(t - \Delta t) \\ \dot{\mathbf{p}}(t - \Delta t) \\ \ddot{\mathbf{p}}(t - \Delta t) \end{bmatrix} + \mathbf{w}(t)$$

where $\mathbf{I}_2$ and $\mathbf{O}_2$ represent 2 x 2 identity and null matrices, respectively. The relation between the state vector in times: t^- and $(t - \Delta t)$ is given by a 3 x 3 matrix known as the transition matrix, $\mathbf{A}$. The vector $\mathbf{w}(t)$ represents the model prediction uncertainty, and it is assumed to be a random variable with a zero mean and a covariance matrix $\mathbf{Q}(t) = E[\mathbf{w}(t)\mathbf{w}^T(t)]$.

In the prediction stage, the sign position is computed by extrapolating the state of the Kalman filter from the previous frame to the current frame. Then the correspondence between the predicted object position and the object position detected by our recognition system is computed. We estimate the sign's position, and velocity and acceleration of the relative motion by evaluating the difference from the corresponding trajectory:

$$\mathbf{z}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{v}(t) \tag{4}$$

where $\mathbf{z}(t)$ represents the external observation (the detected position), and $\mathbf{v}(t)$ represents the uncertainty in such an

observation. Here we assume $\mathbf{v}(t)$ is zero mean with a covariance matrix $\mathbf{R}(t) = E[\mathbf{v}(t)\mathbf{v}^T(t)]$. In correction stage, the new measurement (object's position computed by automatic detection and recognition system) is incorporated to update the state of Kalman filter model by the following process:

$$\mathbf{x}(t) = \mathbf{x}(t^-) + \mathbf{K}(t)(\mathbf{z}(t) - \mathbf{H}\mathbf{x}(t^-)) \quad (5)$$

The weighting factor $\mathbf{K}(t)$ comes from summarizing the following three equations:

$$\mathbf{K}(t) = \mathbf{E}(t^-)\mathbf{H}^T(t)[\mathbf{H}(t)\mathbf{E}(t^-)\mathbf{H}^T(t) + \mathbf{R}(t)]^{-1} \quad (6)$$

$$\mathbf{E}(t^-) = \mathbf{A}\mathbf{E}(t - \Delta t)\mathbf{A}^T + \mathbf{Q}(t) \quad (7)$$

$$\mathbf{E}(t^+) = [\mathbf{I} - \mathbf{K}(t)\mathbf{H}(t)]\,\mathbf{E}(t^-) \quad (8)$$

where $\mathbf{E}(t) = E[(\mathbf{x}(t) - \hat{\mathbf{x}}(t))(\mathbf{x}(t) - \hat{\mathbf{x}}(t))^T]$ is the error covariance matrix in the state estimation process and $\hat{\mathbf{x}}(t))$ is the expected value for $\mathbf{x}(t)$.

For the $\alpha\beta\gamma$ filter, the weighting matrix is given by:

$$K = \begin{bmatrix} \alpha \\ \beta/T \\ \gamma/T^2 \end{bmatrix} \quad (9)$$

where the optimum relationship between the parameters: α, β and γ that minimizes the mean-square error in the position, acceleration and velocity estimates, is given in [13] as follows:

$$\beta = 2(2 - \alpha) - 4\sqrt{1 - \alpha} \quad (10)$$

and

$$\gamma = \frac{\beta^2}{\alpha} \quad (11)$$

knowing that the parameter α takes values between 0 and 1 and is adjusted experimentally.

IV. Experimental results

In our experiments, test sequences have been recorded with a video Sony Firewire digital camera fixed onto the front windshield of a vehicle driving at usual speed. The resolution of each image is 800 x 600 pixels and the average number of frames captured per second is 15. Currently, in our research we are particularly concerned with automatic inventory control of road signs and for this reason, sequences are run as batch processes. We have implemented the multi-road-sign tracking system using a 2.2 GHz Pentium 4-M. Fig. 5 shows the acting of the prediction for the trajectory of a traffic sign using two models: constant acceleration and constant velocity. Experimentally, the optimum value for α, β and γ are: 0.8, 0.61 and 0.47, respectively.

The sequences we have tested have been captured over a stretch of, approximately, 4 kilometers. In Table 2 are reported the results for a sequence of 460 frames at the output of tracking system, which include for each sign the number of detections, the corresponding GPS position and the distance of tracking specified in meters. The results of

Type	Number of Detections	Latitude	Longitude	Distance of Tracking
	14	4031.0486N	320.9319W	40.6213
	11	4031.0312N	320.9462W	39.1705
	9	4031.0052N	320.9688W	43.0108
	5	4030.9880N	320.9837W	22.8495
	14	4030.9417N	321.0240W	37.8198
	5	4039.9119N	321.0511W	31.6728

TABLE II
Analysis of Traffic Signs at the output of Tracking Module

the example show that all signs are detected and tracked correctly. As we can observe in Table 3, it is important to point out that the tracking system rejects for this example three false alarms of individual detections because they do not present a temporal coherency in the sequence. We have to consider that the number of times that a sign appears in a sequence is not constant and depends on several factors; mainly, the velocity of the vehicle. In Fig.6 we can observe as the road-sign is tracked even when an occlusion occurs due to the effect of the windscreen wiper. It is important to note that in each frame of the sequences the bounding-box of the detected sign is depicted and since, as we explained in section II-A, a sign can be processed as a sum of two contributions (the rim and the inner area), the same sign can be detected once or twice in each frame. Finally, the identified sign, which is tracked, is represented by a synthetic template in the upper-left corner of the image.

V. CONCLUSIONS AND FUTURE WORK

This work describes the framework of a system for tracking road signs at the output of a global traffic sign detection and recognition system. The integration of tracking stage allows to dynamically predict the locations of road signs in the following frame even though there is no detection by possible failures, attributed to incorrect segmentation, classification or recognition. Tracking algorithm improves the reliability of the whole system due to false alarms are difficulty confused

Type	Number of Detections	Latitude	Longitude
	1	4030.9797N	320.9909W
	1	4030.9562N	321.0114W
	2	4030.9228N	321.0409W

TABLE III
FALSE ALARMS DISCARDED BY TRACKING MODULE

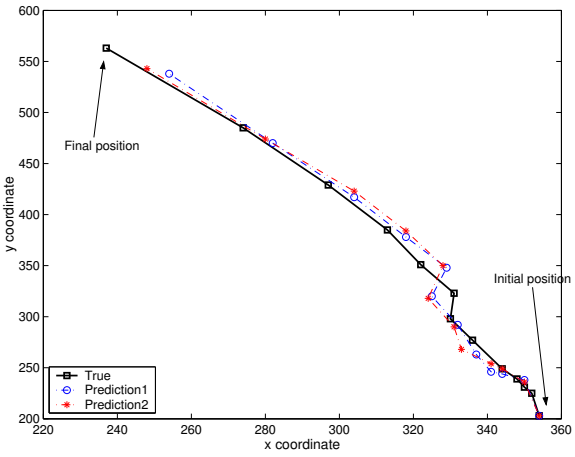
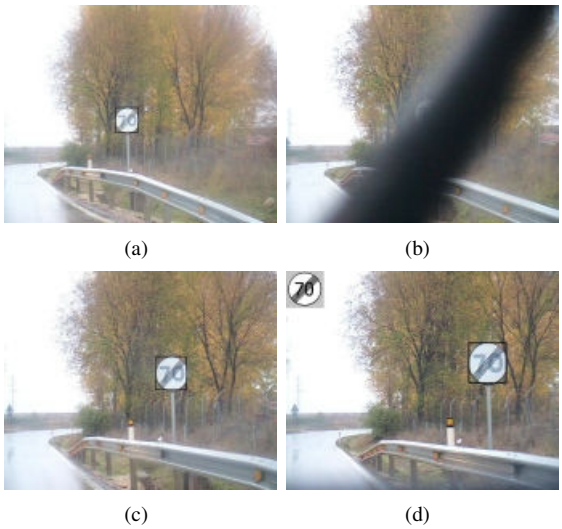


Fig. 5. Position prediction with a model of constant acceleration (Prediction 1) and with a model of constant velocity (Prediction 2).



(a) (b)

(c) (d)

Fig. 6. Tracking results of a sequence.

with traffic signs in consecutive frames and in most cases are discarded. On the other hand, consecutive detections of the same sign are grouped as a single one whose position information is provided by GPS. The GPS position system provides positioning data for road signs. Even though we are interested in automation roadway inventory, our future work will focus on a real time implementation of our system using lower resolution images. At the moment we can process 3.26 frames per second in a Pentium IV using 400x300 pixels images.

VI. ACKNOWLEDGMENTS

This work was supported by the project of the Ministerio de Educación y Ciencia de España with number TEC2004/03511/TCM.

REFERENCES

[1] H. Kamada, S. Naoi, and T. Gotoh, "A compact navigation system using image processing and fuzzy control," in *Proc. Southeastcon*, vol. 1, New Orleans, April 1990, pp. 337–342.

[2] A. de la Escalera, J. M. Armingol, J. M. Pastor, and F. J. Rodríguez, "Visual sign information extraction and identification by deformable models for intelligent vehicles," *IEEE Trans. on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 57–68, June 2004.

[3] S. Lafuente-Arroyo, P. García-Díaz, F. J. Acevedo-Rodríguez, P. Gil-Jiménez, and S. Maldonado-Bascón, "Traffic sign classification invariant to rotations using support vector machines," in *Proc. of Advanced Concepts for Intelligent Vision Systems*, Brussels, Belgium, August-September 2004.

[4] N. Barnes and A. Zelinsky, "Real-time radial symmetry for speed sign detection," in *Proc. of IEEE Intelligent Vehicles Symp.*, Parma, Italy, June 2004, pp. 566–571.

[5] G. Loy and N. Barnes, "Fast shape-based road sign detection for a driver assistance system," in *Proc. of International Conference on Intelligent Robots and Systems (IROS)*, Senday, Japan, Sept.-Oct. 2004, pp. 70–75.

[6] C. Fang, S. Chen, and F. C., "Road sign detection and tracking," *IEEE Trans. on Vehicular Technology*, vol. 52, no. 5, pp. 1329–1341, September 2003.

[7] A. Farag and A. E. Abdel-Hakim, "Detection, categorization and recognition of road signs for autonomous navigation," in *Proc. of Advanced Concepts for Intelligent Vision Systems*, Brussels, Belgium, August-September 2004.

[8] P. Gil-Jiménez, S. Lafuente-Arroyo, F., S. Maldonado-Bascón, and H. Gomez-Moreno, "Shape classification algorithm using support vector machines for traffic sign recognition," in *Computatonal Intelligence and Bioinspired Systems*, Barcelona, Spain, June 2005, pp. 873–880.

[9] Y. Aoyagi and T. Asakura, "A study on traffic sign recognition in scene image using genetic algorithms and neural networks," in *Proc. of the 22nd. IEEE Int. Conf. Industrial Electronics, Control and Instrumentation*, vol. 3, Taipeh, Taiwan, August 1996, pp. 1838–1843.

[10] S. Maldonado, S. Lafuente, P. Gil, H. Gómez, and F. López, "Road-sign detection and recognition based on support vector machines," *IEEE Trans. on Intelligent Transportation Systems*, Pending of publication.

[11] H. Liu, D. Liu, and J. Xin, "Real-time recognition of road traffic sign in motion image based on genetic algorithm," *Proc. of the 1st. Int. Conference on Machine Learning and Cybernetics*, pp. 83–86, November 2002.

[12] R. E. Kalman, "A new approach to linear filtering and prediction problems," *IEEE Trans. of the ASME, Journal of Basic Engineering*, vol. 82, 1960.

[13] P. R. Kalata, "The tracking index: A generalized parameter for $\alpha - \beta$ and $\alpha - \beta - \gamma$ target trackers," *IEEE Trans. on Aerospace and Electronics Systems*, vol. 20, no. 2, 1984.

[14] M. de Fomento, "Orden 1798, boletín oficial del estado," *BOE*, no. 25, pp. 4049–4106, January 2000.