

# A FIRST APPROACH TO LEARNING A MODEL OF TRAFFIC SIGNS USING CONNECTIONIST AND SYNTACTIC METHODS

Miguel SAINZ and Alberto SANFELIU  
Instituto de Cibernética, Universidad Politécnica de Catalunya - CSIC  
e-mail: sainz@ic.upc.es, sanfeliu@ic.upc.es

## Abstract

A system to learn and recognize traffic signs is described. The system uses neural network image processing and syntactic methods. The learning process is based on the representation of traffic signs by means of a grammar, which is inferred from a set of positive and negative samples. The recognition of traffic signs in a scene is done in two steps. First, the sign is located in the scene by using a connectionist segmentation method. Second, the sign is coded and analysed to determine which traffic sign it is. The system has been tested successfully only for the first step. The second step is currently under development.

## 1 Introduction

During the last few years much research effort has been devoted to autonomous vehicle navigation using digital image processing. Most has been aimed at road boundary detection and obstacle avoidance, and several very robust and reliable systems have been implemented. Several different techniques have been applied, as shown in [1], [2], but lately the use of neural networks has produced promising results because of their robustness and computational simplicity ([5], [4]).

Besides road boundary detection and obstacle avoidance, another aspect of autonomous vehicle navigation is traffic sign detection and recognition. Only a few systems have been designed for that purpose. Some of them are described in [3].

The main purpose of our research is to develop a system for learning and recognizing traffic signs using neural networks and syntactic methods. We have studied the use of automatic learning to see to what extent it is possible to use a learning process instead of having to develop a specific method for every problem. We use two levels of learning: segmentation learning based on neural networks, and model learning based on grammatical inference. The two levels of learning will be explained in the following section.

The recognition process uses the results of the learning process as inputs for identifying the traffic signs. In this work we will describe the learning and recognition process and we will discuss the results obtained so far.

## 2 The learning process

The learning process consists of two levels: the segmentation learning process and the model learning process.

We use a color image of 512 x 512 pixels obtained from the TV camera on the vehicle. The analog signal of the TV camera is digitized by an 8-bit A/D converter into three channels corresponding to the red, green and blue color components.

The segmentation learning process is used to learn the different classes of pixels that we will use to segment the scene in the recognition process. In order to reduce the amount of data of the 512x512 pixels images, we work with 4x4 pixels windows in this level.

A human operator decides how many different labels the system will consider and then he marks and labels some areas in a set of images ( these areas will be called segmentation areas). He then selects from those images several positive samples which will be used in the segmentation learning module. In the traffic sign recognition problem we have considered the following five classes: road, road lines, sky, grass and traffic signs. The last label considered is used to locate the traffic sign in the scene.

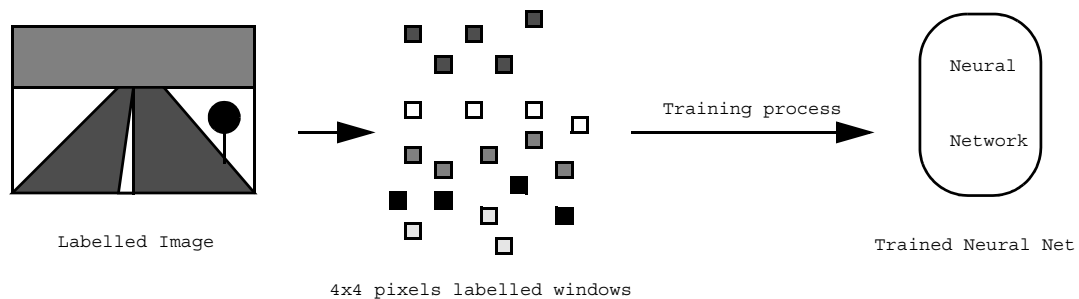


Figure 1 Segmentation learning process.

---

The segmentation module consists of a three-layered neural net. This net has 48 inputs corresponding to a 4x4 pixels x 3-color window of the image and one output for each considered label ( 5 in our case ). The number of neurons in the input, hidden and output layers are set by the operator. In our case and after several tests we have set the numbers of neurons to be 10 in the first layer and 10 in the hidden layer. The output layer has the same number of neurons as the number of considered output labels.

This net is trained by the back-propagation method using the set of samples from the segmentation areas selected by the human operator.

Once the net is trained, we perform a validation test over a set of test images to check the learning performances. At this point the human operator can modify the samples or the net parameters to improve the learning of the segmentation module.

When the segmentation learning level is completed the operator can proceed to the model learning level.

In this level the operator marks the areas on the scenes where the model to learn is located. These areas will be called model areas. These areas are the set of positive samples for model learning.

Before the learning process starts it is necessary to preprocess the image. This preprocessing has three parts: optimization of the areas, normalization of the sizes and coding into symbols the contents of the sample areas.

The normalization of the size of the model areas is performed in order to get the same information from any of the different images areas. We have set an arbitrary size of 50x50 pixels because this is the average size of the traffic sign samples.

We code each pixel of the model area into one of the following 4 symbols: red (R), white (W), black (B) and the remaining of colors (\$). These four symbols will be the four primitives of the grammar. We perform a linear transformation of the R,G,B channels in order to intensify the red and the white colors and we classify the pixels by a standard histogram based thresholding.

After the coding, a morphological process is applied to improve the shape of the traffic sign by removing holes and smoothing the contour. Also, the information (the speed limit,...) inside of the traffic sign is removed and is changed to a white pixel ( the information of the traffic sign will be taken into account once the system will be completely tested ). This is done because it is desired to learn only the shape of the sign (round, triangular or square).

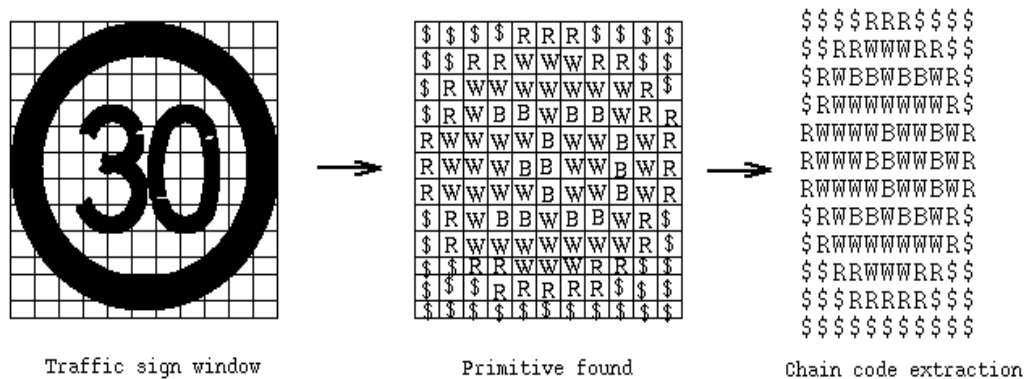


Figure 2 Traffic sign primitive extraction

At this point, we are able to extract the primitive chains by reading the primitives from the coded samples. Now, the operator may introduce some negative samples into the sample set. Then the learning process of the model begins. The methodology used is that of active grammatical inference learning described in [6] and [7].

After an argument regular expression ( a context sensitive grammar ) is inferred, a validation test is applied to evaluate how good the system is. Here, the operator may restart the model learning level changing the samples and the learning parameters.

Once the two learning levels are completed, the results are transferred to a recognition system.

### 3 The recognition process

The recognition process of a traffic sign from a road scene is divided into three steps. The first step is the location of the traffic sign. We use a segmentation module which consists of a pre-trained neural net to segment the image. See Figure 3.

Then, a morphological process is applied to remove noise and fill up gaps, and the system then looks for all the objects labelled as traffic sign by the neural net that are located in the right half of the scene.

The objects found are analysed by applying morphological processing and shape contour extraction methods. The smallest square window that contains the traffic sign candidate is located and the system proceeds to the second step.

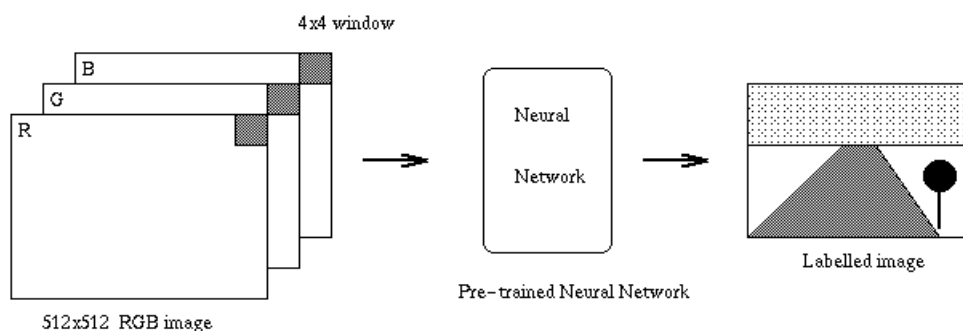


Figure 3 Scene segmentation process.

---

Now the system optimizes the size of the window and the inside of the square window is coded into symbols. After this coding, a size normalization of the traffic sign candidate is applied and noise removal processes are applied to clean up the window.

The third step is the recognition of the traffic sign. This step is divided into two phases.

First, the system recognizes the shape of the traffic sign by finding a distance measure between the extracted symbol chain and each inferred grammar of the traffic sign models. In this phase, an error-correcting parser is used. The traffic signs will be identified by the class with the lowest value of the distance measured if this distance is below a threshold.

At this point we know the shape of all the traffic signs found in the scene. The next phase is to analyse the symbol inside the sign. Once we know both the shape and the symbol, our system will identify the traffic sign.

### 4 Results

In this section we will show some examples of road scene segmentation and traffic sign coding into symbols.

On the left side of Figure 4 we can see two road scenes. On the right side we have the two segmented road scenes. They are labelled from black to white. There are 6 labels

corresponding to unknowns (black), grass, blue sky, road, white lines and traffic signs (white).

---

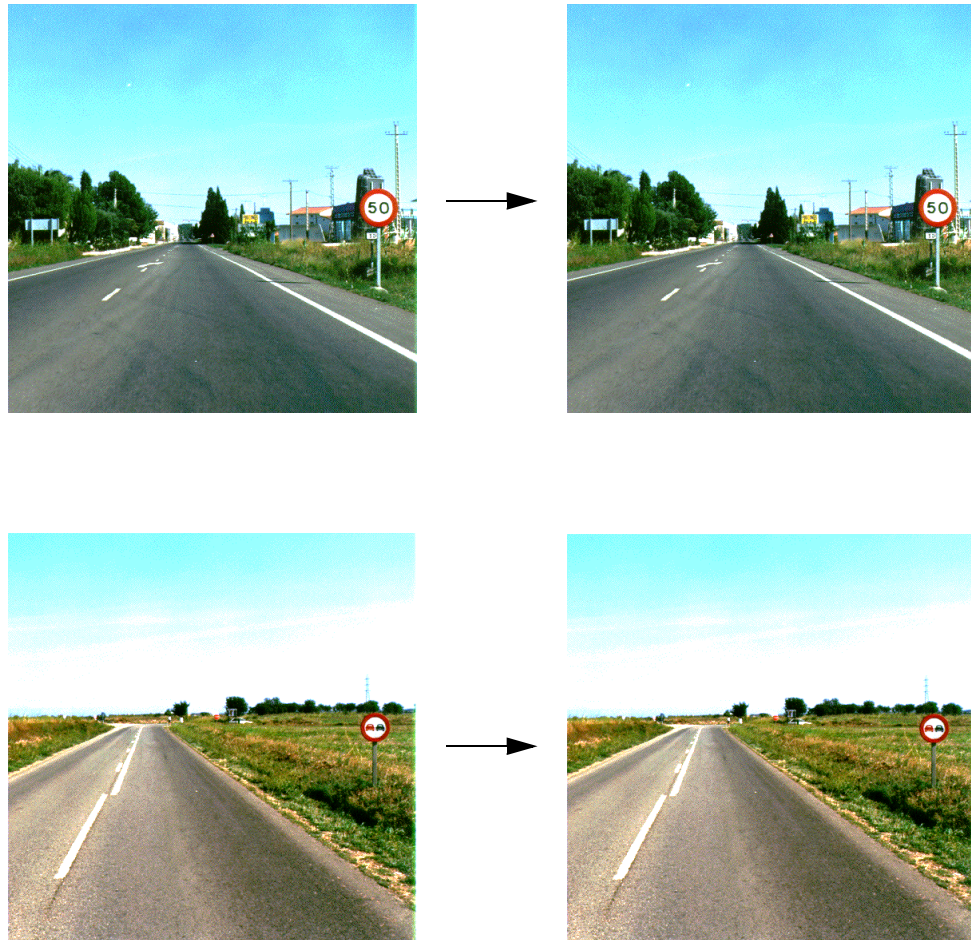


Figure 4: Neural net segmentation results.

---

As we see, the segmentation process gives very good results without using morphological processes. The noise level is very low and can be improved easily by applying noise removal techniques.

Traffic sign detection becomes very easy in those low-noise segmented images.

In Figure 5 we can see the traffic signs from the scenes and the results of coding them into grammar symbols. The traffic signs have different sizes but they are normalized during the coding process. This normalization has to be improved because it introduces noise and shape distortions in the images.



## **5 Conclusions**

At this time, only the segmentation learning process has been completed. Our system is able to segment the road scene and find the traffic sign. The segmentation with neural networks gives very good results on labelling colored images. We have tested our neural nets on different road scenes ( obstacles or shadows on the road, noisy images ...). The system has shown to be very robust. The system is also able to locate the traffic sign and code it into symbols with a very small amount of noise.

We are presently developing the model level learning process.

Traffic sign coding into symbols has been achieved. Presently we are adapting the grammatical inference methodology to the two-dimensional problem. As shown in [7], this methodology gives good results in one dimension. We are evaluating how good the results are with 2D image inputs.

## References

- [1] Charles Thorpe and Tadeo Kanade. *1987 Year End Report for Road Following at Carneige Mellon*, CMU-RI-TR-88-4. The Robotics Institute. Carnegie Mellon University. April 1988.
- [2] Graefe, V., Blöchl, B. *Visual Recognition of Traffic Situations for an Intelligent Automatic Copilot*. PROMETHEUS Workshop, Proceedings of the 5th workshop, Munich, 1991, pp.98-108.
- [3] Austermeier H., Büker U., Mertsching B., Zimmermann S. *Analysis of Traffic Scenes by Using the Hierarchical Structure Code*. Advances in Structural and Syntactic Pattern Recognition, proc. of the International Workshop on Structural and Syntactic Pattern Recognition, Bern, Switzerland, August 1992. Bunke and Wang, Series in machine perception and artificial intelligence, Vol. 5, pp.561-570.
- [4] Català A., Grau A., Morcego B., Fuertes J.M. *A Neural Network Texture Segmentation System for Open Road Vehicle Guidance*. Proc. of the Intelligent Vehicles '92 Symposium. pp. 247-252, 1992.
- [5] Pomerleau, D.A. *ALVINN: An Autonomous Land Vehicle in a Neural Network*, Technical Report CMU-CS-89-107. School of Computer Science. Carnegie Mellon University. 1989.
- [6] R. Alquézar, A Sanfeliu. *A hybrid connectionist-symbolic approach to regular grammatical inference based on neural learning and hierarchical clustering*. Grammatical Inference and Applications, Proc. of the Second Int. Colloquium, ICGI-94, Alicante (Spain), September 1994, R.C. Carrasco, J. Oncina, eds., Springer Verlag, Lecture Notes in Artificial Intelligence 862, pp.203-211.
- [7] A. Sanfeliu, R Alquézar. *Active grammatical inference: a new learning methodology*. Proc. of IAPR Int. Workshop on Structural and Syntactic Pattern Recognition, SSPR'94, Nahariya (Israel), October 4-6, 1994.
- [8] Shun-Ichi Amari. *Mathematical Foundations of Neurocomputing*. Proc. of the IEEE, Vol. 78, No. 9, September 1990. pp1443-1462.
- [9] Fu K.S. *Syntactic Pattern recognition and applications*. Prentice-Hall 1982.