Color-Based Road Sign Detection and Tracking

Luis David Lopez and Olac Fuentes

Computer Science Department University of Texas, El Paso 79902, USA
{ldlopez, ofuentes}@utep.edu

Abstract. This paper describes a general framework for the detection and tracking of traffic and road signs from image sequences using only color information. The approach consists of two independent parts. In the first we use a set of Gaussian distributions that model each color for detecting road and traffic signs. In the second part we track the targets detected in the first step over time. Our approach is tested using image sequences with high clutter that contain targets with the presence of rotation and partial occlusion. Experimental results show that the proposed system detects on average 97% of the targets in the scene in near real-time with an average of 2 false detections per sequence.

1 Introduction

Automatic detection of road signs has recently received attention from the computer vision research community. The main objective of these algorithms is to detect signs with different maintenance conditions and variable light sources from image sequences acquired from a moving vehicle. Many algorithms have been proposed to solve this problem. These methods can be divided into shape-based [5, 8, 11, 16] and color-based segmentation [3, 4, 13].

Shape-based methods detect the signs using a set of predefined templates. In general, these methods are sensitive to total or partial occlusion and target rotation. Color-based methods detect signs in a scene using the pixel intensity in RGB [3], CIELab [4] or other color spaces [13]. A few typical problems in the detection of road signs using color information are that vandalism, long exposure to the sun, or camera sensitivity produce a change in the apparent color of the sign. Other approaches combine shape and color segmentation to improve the percentage of correct detection, while minimizing the number of false detections [1, 6, 7]. But due to time requirements, these methods are not suitable for real-time applications.

In this paper we are proposing to detect road signs in CIELab color space, modeling the pixel intensity values using a set of Gaussian distributions. We combine the segmentation and tracking processes in order to obtain accurate detection results, while satisfying real-time constraints.

Our computational results show that this method can detect signs in sequences with different outdoor lighting conditions, in the presence of obstacles such as cars and trees that partially occlude the sign, and with motion blur that is originated by the vehicle motion and vibration.

Some of the advantages that motivate the use of video instead of single images in this framework are:

- Redundant information: The use of several images from the same scene can increase the robustness of the detection system, especially in sequences with high level of noise. Thus, a failure in the detection in a given image does not necessarily result in the failure of the detection in the whole sequence.
- Targets with different sizes: We only process targets whose area is bigger than a predefined threshold and reject the smaller targets. With this method we reduce the false target detection rate that may be generated by the noise in the image.
- Several points of view of the same target: with every step the angle of the camera changes; this information can be used to make the sign classification process easier [12].

This paper is organized as follows. Section 2 shows an overview of the proposed framework. Sections 3 and 4 describe the algorithms used to detect and track the signs, respectively. Several results that validate our system are reported in Section 5, and finally Section 6 contains concluding remarks.

2 Framework

The proposed road sign detection and tracking framework is shown in Figure 1. In our experiments, we first convert the rgb-scale image into CIELab color space. The segmentation phase uses a mixture of three Gaussian models [14] to characterize each road sign color (red, green, blue, yellow, and orange). These models are updated after the execution of the segmentation algorithm in order to improve the discrimination of the background in each step. All the candidate signs are tracked over time using the Conditional Density propagation algorithm *Condensation* [10], which is a probabilistic method that iteratively propagates the targets' position using Bayes' rule. After tracking the targets for k frames, the segmentation algorithm is executed again in order to obtain more information about the road sign; this helps to deal with partial occlusion and changes in size. The details of these algorithms are given in the next two sections.

3 Color-Based Sign Detection

In this approach we discriminate the road sign from the background using a set of Gaussian models. We use a perceptually uniform color space called CIELab [15] that measures how similar the reproduction of a color is to the original when viewed by a human observer. In this space, every pixel has a luminance component L and two chrominance components a and b. We create a system that is sensitive to color but invariant to intensity changes using only the luminance components, as also done in [4].

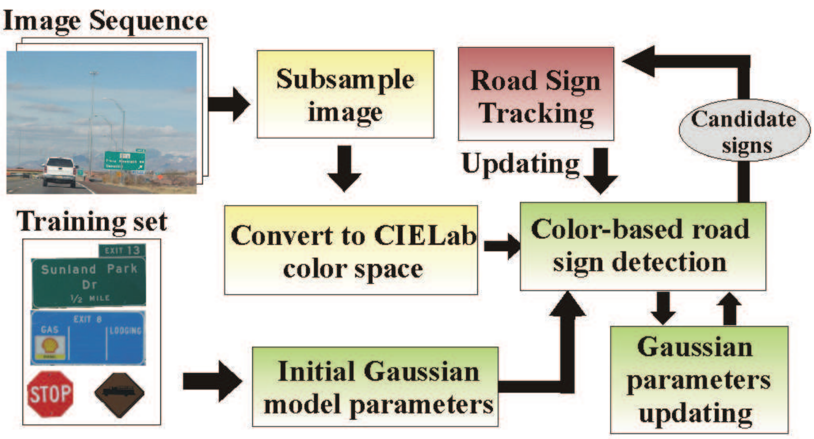


Fig. 1. Road Sign detection framework

Gaussian models have been successfully applied to automatic detection of mobile targets using color information [14, 9]. Traditionally, this method models the background color of a given image, so that everything that is similar enough to the model is viewed as a part of the background. The results are improved at every step using a learning function that updates the mean value of the distributions.

Unlike traditional methods that model the background, in our framework we use a Mixture of Gaussian distributions to model each road sign color (red, green, blue, yellow, orange and brown), where every Mixture of Gaussian distributions consists of three Gaussian models. Our detection algorithm consists of three basic steps: initialization, model generation and target detection. The following paragraphs explain each of these steps in detail.

Gaussian initialization A Gaussian Model consists of a mean value and a covariance matrix. In this step we compute the initial values of these parameters using a set of image patterns. Every model of the same mixture is trained using images with excellent, regular, and poor illumination conditions, respectively. In the training process the mean and the covariance matrix of every Gaussian model are computed using Eq. 1 and 2, respectively.

$$\mu_i = \frac{1}{n}\sum_{j=1}^{n} color_{i,j} \tag{1}$$

$$\Sigma_i = \frac{1}{n-1}\sum_{j=1}^{n}(color_j - \mu_i)(color_j - \mu_i)^T \tag{2}$$

Model Generation In a Gaussian model each pixel p is modeled with a single probability distribution $\eta(p_t, \mu_t, \Sigma_t,)$, as shown in Eq. 3, where μ_t and Σ_t are the mean value and covariance matrix of the distribution at frame t, respectively. Pixels where observed colors p_t are close enough to the background distribution are classified as background points, while those too far away are classified as foreground points.

$$\eta(p_t, \mu, \Sigma_t) = \frac{1}{(2\pi)^{n/2}|\Sigma_t|^{\frac{1}{2}}} e^{\frac{1}{2}(p_t-\mu)^t \Sigma_t^{-1}(p_t-\mu)} \tag{3}$$

In order to improve the accuracy of the detector, we use a multi-modal model; in this model the probability of observing the current pixel value is computed using a mixture of K multiple independent distributions (Eq. 4), where $\mu_{i,t}$ and $\Sigma_{i,t}$ are the i^{th} mean and covariance at time t, respectively, and $\omega_{i,t}$ is an estimate of the weight of the ith Gaussian in the mixture at time t.

$$P(p_t) = \sum_{i=1}^{K} \omega_{i,t} * \eta(p_i, \mu_{i,t}, \Sigma_{i,t}) \tag{4}$$

Rather than explicitly specifying the value of the signs' colors, we model the value of a particular color using Eq. 4, with $K = 3$ and $\omega_{i,t} = 1$. In our model, if the probability $P(p_t)$ is higher than a predefined threshold, we classify the pixel as part of a road sign.

Target Detection In this step, for every pixel in a given image we compute the difference between the modeled and the real pixel value. A small difference means that this pixel is accurately described by the model and this pixel is considered an element of a sign, while a high difference represents a pixel that does not belong to a road sign. In this framework, it is enough if one model of the distribution generates a similar value to consider the pixel as part of the foreground. The pixels that were labeled as foreground are grouped using connected components [2], creating regions of interest that are used for later tracking.

4 Target Tracking

In the previous step we generate a set of targets that are represented by a state vector x_i, which contains the position and size for every target. We use a modified version of the condensation algorithm to track this set of targets over time.

The proposed tracking algorithm is shown graphically in Figure 4 and expressed as pseudo-code in Table 1. This method has three steps: initialization, observation, and selection. The following paragraphs describe these steps in detail.

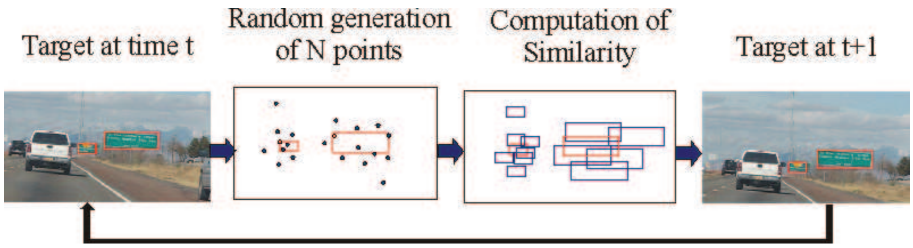


Fig. 2. Target tracking

Table 1. The algorithm to track road signs

Tracking(Target[ ] T)
- *where* $T = (\boldsymbol{x}, width, height)$ *and* $\boldsymbol{x} =$ *(positionX, positionY, velocity)*
For i = 1 to maxTargets
- *Select the ith target from the Targets' list*
Select T_i from T
- *Generate N particles for every target*
For k = 1 to N
$x'_{i,k} = T_i.\mathbf{x} + r(T_i.\mathbf{velocity}, \sigma)$
$X'_i = (x'_{i,1}, x'_{i,2}, ..., x'_{i,k})$
- *Compute similarity*
$P'_i = p(obs \mid X'_i)$
- *Update target i at time t+1*
prediction = $argmax(P'_i)$

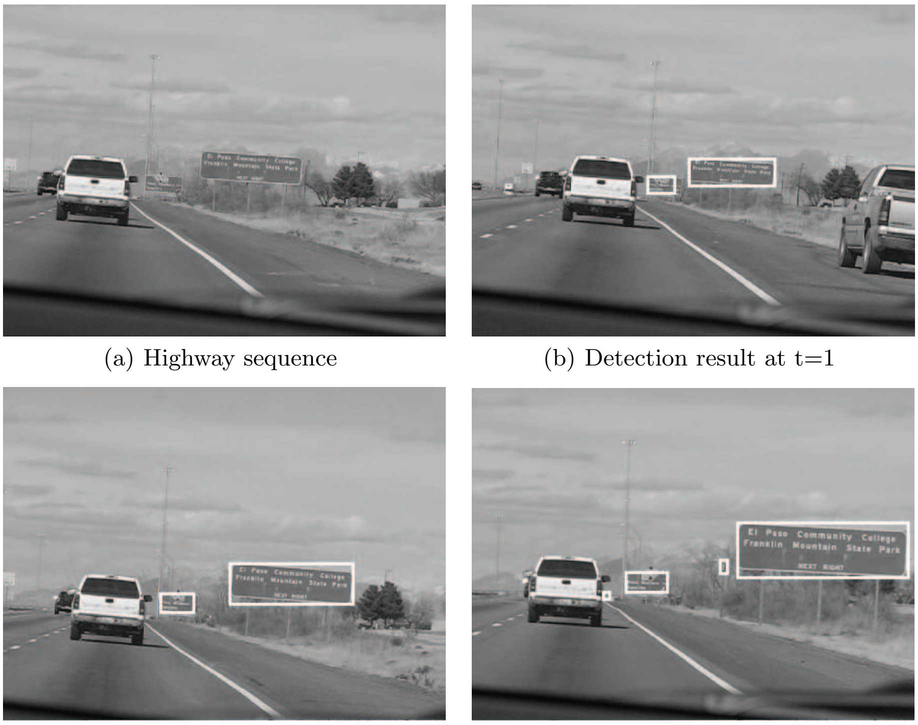
Initialization After the target detection step, the state vector T_i stores the position and size of all the targets at time t. In this approach, we randomly generate a set of N coordinates (x, y) that estimate the target position at time $t + 1$. The estimation of the target position at time $t + 1$, $\widehat{T}_{i,n}$, is defined in Equation (5), where r is a Gausssian random value with mean equal to the current velocity.

The current targets' velocity can help improve the prediction, but the detection algorithm can not compute this parameter, thus, in the first iteration we estimate the next position using a velocity equal to zero. After the first tracking step the velocity in pixels of the targets is computed as the difference between the positions at time t and $t + 1$.

$$\widehat{T}_{i,n} = T_i.\mathbf{x} + r(T_i.velocity, \Sigma) \tag{5}$$

Observation In the observation process we compute the similarity between every target and its N predicted positions using a simple correlation method that works with gray scale images. We use gray scale images instead of images in CIELab color space because the tracking algorithm has approximately the same results in both cases, but the time required to compute the correlation is significantly lower using a gray scale image. The result of this step is an array that represents the similitude of the predicted position i at time $t+1$ with the target at t.

Selection We determine which of the predicted positions contains the target in the image at time $t+1$. In this framework we select the highest correlation value, which represents the most similar prediction as the position of the target.



(a) Highway sequence

(b) Detection result at t=1

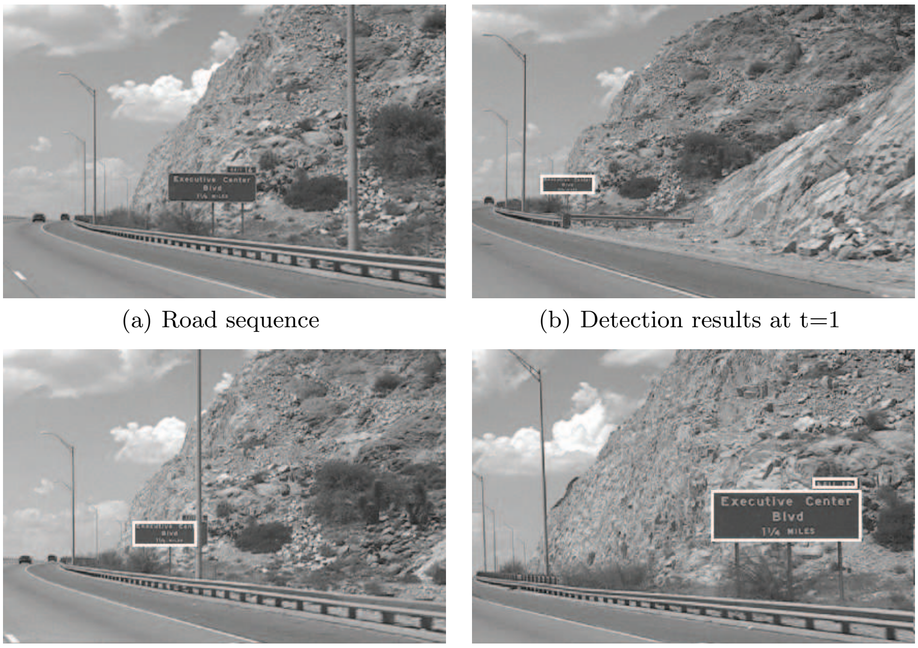
(c) Detection result at t=100

(d) Detection result at t=200

Fig. 3. Road sign detection results in Highway sequence

5 Experimental Results

In this section we show the experimental results of our approach. The algorithm was tested with a database of ten image sequences with a frame rate of 20 images per second, captured by a digital camera mounted on a vehicle and



(a) Road sequence

(b) Detection results at t=1

(c) Detection results at t=80

(d) Detection results at t=160

Fig. 4. Road sign detection results in Road sequence

different driving situations. Figures 3(a), 4(a), and 5(a) show three sample image sequences: *Highway*, *Road*, and *Parking lot*, respectively. The *Highway* sequence (Figure 3) contains four green and one yellow signs, some of them with partial occlusion and rotation on a straight highway. The *Road* sequence (Figure 4) contains two green signs of different sizes in images with a large amount of noise. The *Parking lot* sequence (Figure 5) contains several objects with colors that can be detected as road signs. The database also contains a large amount of shrubs and trees with color that is similar to the green signs.

Table 2 summarizes the principal features of the sample sequences and the computational results. In the table, Time processing shows the time required to process an image of the sequence, Pd is the probability of detecting a target with size larger than 150 pixels, and NFt is the number of false detections per sequence.

By applying the algorithms described in sections 3 and 4, the signs are detected and tracked in each frame. Figures 3(b), 3(c), and 3(d) show the road signs detected in the Highway sequence; in these images our method detected all the road signs in the scene with only one false detection. The two signs that are overlapped are detected as one single sign. This problem is generated by the camera perspective, but it is solved when the vehicle is closer and the angle from the camera to the signs places them on different positions in the image plane. In

(a) Parking sequence

(b) t = 1

(c) target at t = 50

(d) t = 100

Fig. 5. Road sign detection results in Parking lot sequence

these images the milepost and other partially occluded signs are detected after 180 images when their size is larger than the threshold defined in the detection algorithm.

Figures 4(b), 4(c), and 4(d) show the results using the road sequence; in these images we can observe that the proposed method detects all the road signs in the scene, and discriminates objects with similar colors to road signs, obtaining zero false detections.

The results for the parking lot sequence are shown in Figures 5(b), 5(c), and 5(d). In this sequence our method detects correctly 95 % of the targets in the

Table 2. Computational results

Sequence	Image size	Frames	Targets	Time processing	Pd (%)	NFt
Highway	640x480	360	4	0.25 secs	96	1
Road	720x480	240	2	0.32 secs	98	0
Parking lot	2272x1240	240	2	1.60 secs	95	5

scene. The false detections are objects that can not be differentiated from road signs using only color information, but this number is not significant compared to the number of images in the sequence.

In these figures the false detections are shown with a white circle, while the detections obtained from the segmentation algorithm are marked with a white square. These results show that the proposed system can detect rotated signs and signs with different sizes and colors in sequences with varying characteristics.

6 Conclusions

This paper describes a general framework for the detection and tracking of traffic and road signs from image sequences using only color information.

In the detection step we are proposing the use of Gaussian models to detect road signs. This method combines detection and tracking in order to reduce the computational time required to process the whole image sequence, making this framework suited for real-time applications.

The proposed approach was tested using a set of image sequences that contain signs with different sizes and colors in the presence of occlusion. Experimental results show that our method is invariant to in-plane rotation, being able to detect signs with different sizes and colors viewed at any orientation. On average the system detects 97% of the signs in highway, road and street environments.

As future work we plan to use a machine learning method to improve the learning function in the detection algorithm. We will also perform experiments with nighttime image sequences and use road shape information in order to reduce the number of false detections.

References

1. Claus Bahlmann, Ying Zhu, Visvanathan Ramesh, Martin Pellkofer, and Thorsten Koehler. A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. In *Proceedings of the IEEE Intelligent Vehicles Symposium (IV 2005)*, 2004.
2. D. Ballard and C. Brown. *Computer Vision*. Prentice-Hall, Englewood Cliffs, New Jersey, 1982.
3. M. Benallal and J. Meunier. Real-time color segmentation of road signs. *Proceedings of the Canadian Conference on Electrical and Computer Engineering, 2003 (IEEE CCECE 2003)*, pages 1823–1826, 2003.
4. Murphy Chutorian and Trivedi M.M. N-tree disjoint-set forests for maximally stable extremal regions. In *Proceedings of the British Machine Vision Conference*, Edinburgh, PA, September 2006.
5. A. de la Escalera and L. Moreno. Road traffic sign detection and classification. *IEEE Transactions on Industrial Electronics*, 44:848–859, 1997.
6. Chiung-Yao Fang, Sei-Wang Chen, and Chiou-Shann Fuh. Road-sign detection and tracking. *IEEE Transactions on Vehicular Technology*, 52(5):1329–1341, 2003.
7. H. Fleyeh and M. Dougherty. Road and traffic sign detection and recognition. In *Proceedings of the 10th EWGT Meeting and 16th Mini-EURO Conference*, 2005.

8. D. Gavrila and V. Philomin. Real-time object detection for smart vehicles. In *Proceedings of International Conference on Computer Vision*, pages 87– 93, 1999.
9. Luis David López Gutiérrez and Leopoldo Altamirano Robles. Decision fusion for target detection using multi-spectral image sequences from moving cameras. In *Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis*, pages 720–727, Estoril, Portugal, June 2005.
10. M. Isard and A. Blake. Condensation – conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.
11. G. Loy and N.M. Barnes. Fast shape-based road sign detection for a driver assistance system. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 70–75, 2004.
12. C. Martínez and O. Fuentes. Face recognition using unlabeled data. *Iberoamerican Journal of Computer Science Research*, 7:123–129, 2003.
13. W. G. Shadeed, D. I. Abu-Al-Nadi, and M. J. Mismar. In *Proceedings of the 10th IEEE International Conference on Electronics, Circuits and Systems*, pages 890–893.
14. C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages II: 246–252, 1999.
15. V. Vezhnevets, V. Sazonov, and A. Andreeva. A survey on pixel-based skin color detection techniques. In *Proceedings of Graphicon*, 2003.
16. M. Zadeh, T. Kasvand, and C. Suen. Localization and recognition of traffic signs for automated vehicle control systems. In *Proceedings of the Conference on Intelligent Transportation Systems*, Pittsburgh, PA, October 1997.