Detection, categorization and recognition of road signs for autonomous navigation

Alaa E Abdel-Hakim

Proceedings of ACIVS

DETECTION, CATEGORIZATION AND RECOGNITION OF ROAD SIGNS FOR AUTONOMOUS NAVIGATION

Aly A. Farag and Alaa E. Abdel-Hakim
Computer Vision and Image Processing Laboratory
University of Louisville, Louisville, KY 40292
E-mail: {farag, alaa }@cvip.uofl.edu
http://www.cvip.uofl.edu

Abstract

In this paper we present a novel and robust approach for detection, categorization and recognition of road signs. It is known that the standard road signs contain few and easily distinguishable colors, such as red for prohibition, yellow for warnings, green, blue and white. We use a Bayesian approach for detecting road signs in the captured images based on their color information. At the same time, the results of the Bayes classifier categorize the detected road sign according to its color content. The SIFT transform is employed in order to extract a set of invariant features for the detected road sign label(s). Recognition is done by matching the extracted features with previously stored features of standard signs. We illustrate the accuracy and robustness of this approach.

1. INTRODUCTION

The main objective of the project, we are working in, is to develop a smart autonomous vehicle that can navigate through an environment and, through a sensor suite, collect data about the environment which feeds into an on board intelligent system for understanding the environment and performing certain tasks of interest. One of the goals of our system is to provide a Driver Support System (DSS) that will be employed in a pedestrians aiding system. In such applications, road sign detection and recognition (RSR) is very important, since the road signs carry much information necessary for successful, safe and easy driving and navigation.

The RSR approach, proposed in this paper uses a Bayes classifier to detect the road signs in the captured image (e.g., [1]), based on its color content. The color category of the road sign is very important in the recognition of it. For example, two identical signs with different colors may have a completely different interpretation because of the difference between their colors. The Bayes classifier does not just label the captured image only, but it categorizes the labels to the appropriate category of the road sings as well.

This work is supported by US Army under grant DABT60-02-P-0063.

Based on the results obtained by the Bayes classifier, an invariant feature transform, namely the Scale Invariant Feature Transform (SIFT) is used to match the detected labels with the correspondent road sign [2]. The contribution in this paper is in using an invariant feature approach (SIFT) for the RSR problem. Using the SIFT transform for the matching process achieves several advantages over the previous work in RSR. For example, it overcomes some difficulties with previous algorithms such as the slowness of template matching based techniques [3], the need for a large number of various real images of signs for training like the neural-based approaches [4], or the need for need a priori knowledge of the physical characteristics of the lighting illumination of the signs like in [5]. As another advantage of using the Bayes classifier is the acceleration of features extraction and matching operations of the SIFT transform by shrinking the matching area to the labels only. Also, it limits the search subspace of the SIFT transform by determining the color category of the detected sign.

2. RELATED WORK

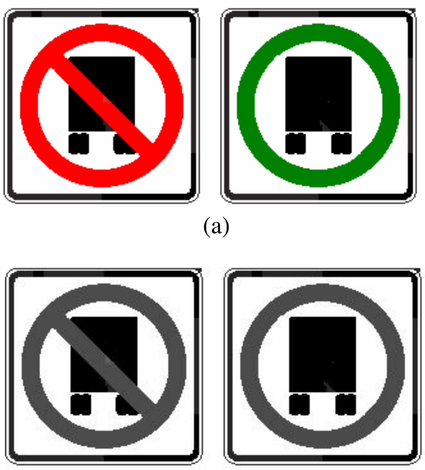
There are many researches in the literature deal with RSR problem. In this section, we will explore some of those approaches and show their advantages and their weak points, which are overcome by using our proposed approach.

In [3], the authors used template matching for recognition of the road signs in the regions of interest (ROI) in the captured image. The ROI of the road image is determined by expecting the possible location(s) of the sign or by using the color information of the road image. The approach of [3] inherits the difficulties of the template matching schemes, namely, the relatively slowness and the need for various shapes for each template to consider different deformations resulted from changes in scale, orientation, rotation ...etc.

In [6], the authors used Laplace kernel classifier for road sign classifications. They used the Laplace kernel classifiers in the decision tree. The smoothing parameters of the Laplace kernel are optimized by the pseudo-likelihood cross-validation method. They used the Expectation-Maximization algorithm to maximize the pseudo likelihood function. That approach used the local orientations of the edges for the matching process. Therefore, it has some difficulties in recognition of square signs in urban areas, where lots of horizontal and vertical lines exist in the image. The template matching approach used in [7] doesn't suffer from the square signs case. However, it still doesn't respond to strong shape distortion at all in addition to the relatively long time consumed in template matching.

In [5], the authors used a combination between physics-based approach for color detection and a template matching-based approach for sign recognition. The use of a physics-based approach for detection requires well knowledge of the appropriate physical model and needs to keep in mind the changes in the model parameters to accommodate the natural variations like illumination and lighting conditions. On the other hand, the difficulties of using template matching, as mentioned before, still exist.

Other approaches, such as [8]-[12], neglect the color components in the acquired image which yield to a significance loss in the signs information. The significance of the color information is illustrated in the example shown in Fig. 1. Figure 1-a shows the images of two colored signs. Figure 1-b shows the gray level images of the two signs. It is clear that neglecting the color information can easily lead to a complete different interpretation of the images.



(a)

(b)

Figure (1): The effect of discarding the color information in the road signs images (a) Colored signs (b) Gray signs

3. BAYES CLASSIFIER

In this section, we introduce a brief description of the Bayes classifier and how it is employed in our road sign detection and categorization problem.

Assume that the captured image is to be labeled with respect to a number of classes *C*. For each pixel *x* in the image, the conditional probability,

$$P(w_i/x),\ \ i \in [1,c] \tag{1}$$

is calculated where w_i is the i^{th} class. This probability value gives the likelihood that this pixel belongs to that class. The Bayes classifier chooses the class based on the maximum conditional probability. In other words

$$x \in w_i, \text{ if } P(w_i/x) > P(w_j/x)\ \ \forall j \neq i \tag{2}$$

The priori probability of each class $P(w_i)$ can be used in calculating the conditional probability of Eq. (1), as shown in Eq. (3).

$$P(w_i / x) = \frac{P(x / w_i) P(w_i)}{P(x)} \tag{3}$$

The priori probability for each class can be assumed/estimated easily. Specifically, in our problem for RSR, we consider 5 color classes (red, yellow, green, blue and white). We assume equal priori probability for each color. A background class is detected, if the maximum posteriori probability is less than a certain threshold.

Parametric and non-parametric estimation can be used to estimate the classes conditional densities $P(x/w_i)$. In this paper, we consider the parametric estimation approach. In this approach, the design data are used to estimate the parameters of the class densities. In this paper, we assume that $P(x/w_i)$ Gaussian distribution function. Hence, the only parameters needed to fully describe the conditional densities $P(x/w_i)$ are the *covariance matrix* Σ and the *mean vector* $\boldsymbol{\mu}$. There are many approaches in the literature to estimate these two parameters. One of the most common approaches is the maximum likelihood estimator (MLE) [1], which is used in our approach. In MLE, the estimated mean vector $(\hat{\mu})$ is obtained as in Eq (4-a) and the estimated covariance matrix $(\hat{\Sigma})$ is obtained as in Eq. (4-b).

$$\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} x_k \tag{4-a}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \hat{\mu})(x_k - \hat{\mu})^t \qquad \text{(4-b)}$$

Where n is number of pixels in the design set, used for estimation. We'll consider the problem as a one-dimensional Gaussian distribution. The one dimension in the input space represents the hue color components in the HSI model. This is because the hue component is invariant to changes in brightness and shadows, which is the suitable case for the RSR problem. Thus, x_k is the hue component for the k^{th} pixel. RGB components can be used for the inputs as well, but in that case, the classification error will be larger.

4. LOCAL INVARIANT FEATURES

Local invariant features descriptors are description vectors which contains some keys that describe a local image region in a manner invariant to spatial transformation and other distortion factors. To be efficient in features matching, the descriptors should be distinctive and at the same time robust to changes in viewing conditions as well as to errors of the point detector. The variations, exist in road images, in lighting conditions, illumination, scaling, rotation and may be affinity, motivate the use of an invariant features approach for RSR. In this section we will explore some of the known invariant features approaches and explain the reason for using the SIFT approach specifically.

Local photometric descriptors computed at interest points are distinctive, robust to occlusion and do not require segmentation. Many of the recent work has been concentrated on how to make these descriptors invariant to image transformation like scaling, translation and rotation. Also, on making them robust with respect to the changes in the gray level, illumination, brightness …etc. The main common idea of the approaches based on these descriptors is to construct invariant "image regions" which are used as support regions to compute invariant descriptors.

There are several researches have been made in this area. For example, Mikolajczyk and Schmid [13] have developed affine invariant interest points with associated affine invariant regions. Tuytelaars and Van Gool [14] construct two types of affine invariant regions, one based on the combination of interest points and edges and the other based on image intensities. In [15], by computing Gaussian derivatives, steerable filters and differential invariants are used for obtaining the local descriptors. Complex filters are proposed in [16] by computing a kernel for each pixel in the image weighted by a Gaussian function.

In [2], Lowe has proposed scale-invariant regions based on local extrema in scale-space built with difference-of-Gaussian (DoG) filters, the SIFT approach. The next section will illustrate it in more details.

Mikolajczyk and Schmid [17] have made a performance evaluation for SIFT descriptors [2] versus other invariant feature descriptors. They concluded that SIFT is the best with respect to the scale, rotation, and illumination changes. Therefore, the SIFT approach adequate for the considered problem of RSR which has many of those variations.

5. SIFT FEATURES

The Scale Invariant Feature Transform, SIFT, was developed by David Lowe [2]. SIFT is used mainly for generation of invariant features in an image for object detection, recognition and tracking applications. The potential of SIFT lies in the invariance of the generated features to image translation, scaling, rotation, and partially to illumination changes and affine projection. Due to these characteristics, it can be considered as the most feature descriptor for the RSR problem. The main contribution in this paper is employing SIFT for the RSR problem to overcome the difficulties appear with other approaches, as discussed in section 2.

The first step in SIFT is the determination of the key locations, or the points of interest. This is done by constructing a Gaussian pyramid, using gaussian smoothing and subsampling. Each level in the pyramid is constructed using a difference-of-Gaussian image, obtained by subtracting the image from its gaussian smoothed image. Then subsampling is performed to construct the next level in the pyramid. Feature locations are identified by detecting the maxima and minima relative to surrounding pixels and adjacent scales. By this way, SIFT guarantees that the key points are located at regions and scales of high variations, which make these locations stable for characterizing the image. Then, the keys are used as input to a nearest-neighbor indexing method that identifies candidate object matches. Final verification of each match is achieved by finding a low-residual least-squares solution for the unknown model parameters. Lowe has called this method for image feature generation "Scale Invariant Feature Transform" (SIFT).

6. RSR APPROACH

The road signs are designed using particular shapes and distinctive colors such that it can be easily distinguished from the surrounding environment.

Depending on this fact, the proposed approach in this paper makes use of both color and shape information. As shown in Section 3, a Bayes classifier is used for labeling the captured image according to the color information in the image, hence the obtained labels are classified according to the standard road signs color classes.

The SIFT descriptors for the standard road signs are built and saved offline. According to the color class decided by the Bayes classifier for the detected sign label, a matching process is performed between the SIFT descriptor of the detected sign label and those of the standard signs in the same color class. The detailed SIFT matching process is explained by Lowe in [2].

The number of invariant features or points of interest varies from one sign to another. Hence, matching based on the absolute number of correspondent features may be misleading. Therefore, the matching process is performed according to the objective function shown in Eq (5)

$$f_i = \frac{n}{\sqrt{n_b n_i}} \qquad (5)$$

Where f_i is the objective value for i^{th} standard road sign, n is the number of matches, n_i is the total number of the invariant features in the i^{th} standard sign and n_b is the total number of the invariant features in the label to be matched. The maximum value of f is 1.0 when full matching occurs and the minimum is zero when no matching occurs. The matching decision is made according to the standard sign with the highest objective value. If the highest objective value is less than a certain threshold, the detection is discarded. The proposed approach can be summarized as shown in Algorithm (1).

As stated in the introduction, the Bayes classifier is not useful in categorization of the detected labels in the captured image only, but it has another important advantage of limiting the search space SIFT descriptors matching process. Hence, the matching speed is improved.

7. EXPERIMENTAL RESULTS

In this section, we will present some results for the red (prohibition) signs. The first row of Figure 2 and Figure 3 shows samples of road images. The second row shows the results of the Bayes classifier on the road images set. The third row shows the gray level labels for the red color in the captured images. The last row shows the matching results. Figures 4-7 illustrate how the matching process is implemented for a specific sign. Figure 4 shows an example for captured road images. As shown in Figure 5, the maximum objective value is at standard image number 27, which is shown in Figure 6. The second maximum is at standard image number 36 shown in Figure 7. It is clear that there are many common features between those two standard images. However, the algorithm succeeded in matching the sign with the correct one.

1. *Build the SIFT descriptors for the known standard road signs (offline).*
2. *Acquire the road image.*
3. *Label the acquired image using Bayes Classifier with respect to the color content.*
4. *Build SIFT descriptors for each label.*
5. *For each label, match its descriptors with those of the standard signs in the same color category.*
6. *For each label, find the standard sign with the maximum objective value ($f_{i\,max}$).*
7. *For each label, if ($f_{i\,max}$) is less than a certain threshold, then no sign detected for that label.*
8. *Else, the standard sign with ($f_{i\,max}$) is detected in the acquired image, for that label.*

Algorithm (1): The Proposed RSR Approach

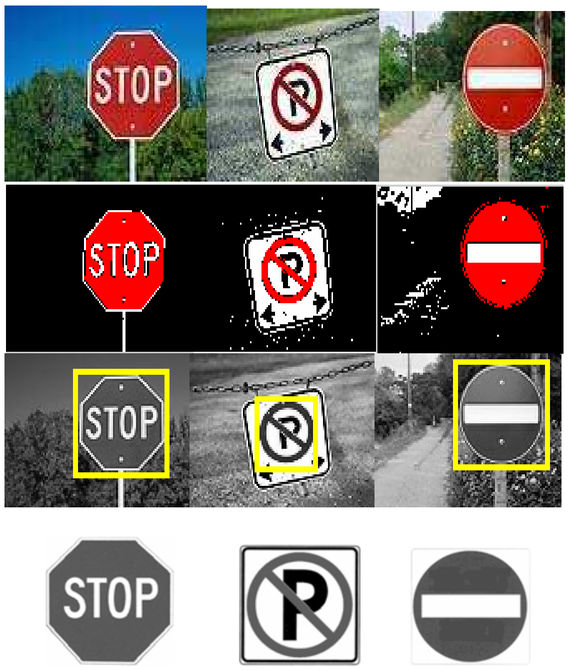


Figure (2) Sample results (1)

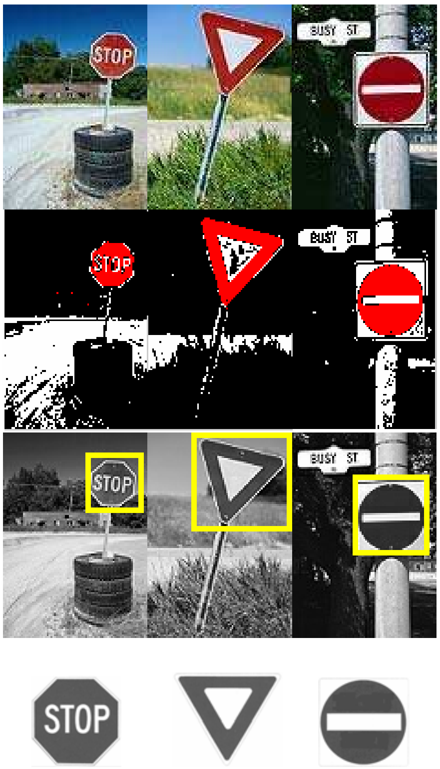


Figure (3) Sample results (2)

Figure (4) Sample captured image

8. CONCLUSION

In this paper, we presented a novel approach for road signs detection and recognition. The detection process is based on the Bayes classification for the color information in the captured image. The recognition process is based on the SIFT approach for matching the invariant features between the captured image and number of standard road images. The proposed approach is to be employed in an integrated project for smart vehicle navigation and DSS system.

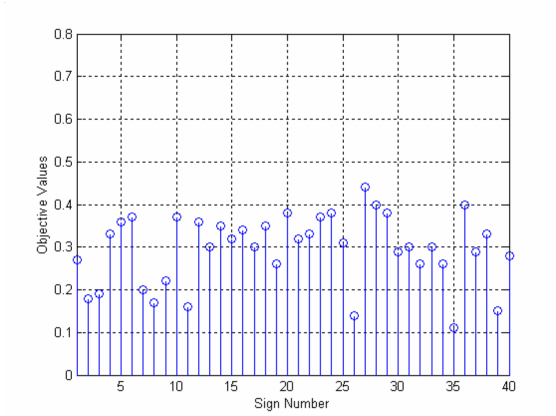


Figure (5) Objective values of the correspondences between the captured image and the standard red signs

Figure (6) The standard image with the highest objective value

Figure (7) The standard image with the second highest objective value

9. FUTURE WORK

Although the detection and recognition of the road sign (RSR) in the captured image is very important in the following stages of the AI navigation, it is not enough to make further driving decisions. For example, the human driver may recognize several road signs in the image seen while driving; the decision is usually made according to just one or two of those signs. Neglecting of many other detected road signs is due to the relatively large distance between the driver's vehicle and the detected signs. Thus, estimating the distance between the vehicle and the detected sign is an important issue that should be considered when developing an autonomous vehicle navigation system. Our future work in DSS is an integration of the proposed RSR approach with the approaches used in

3D reconstruction for the estimation of the distance between the vehicle and the detected sign.

10. REFERENCES

[1] R. O. Duda, P. E. Hart and D. Stork., *Pattern* Classification, 2nd edition , Wiley, 2001.

[2] David G. Lowe, "Object Recognition from Local Scale Invariant Features", ICCV 1999, pp. 1150-1157, 1999.

[3] C. Huang, S. Hsu, "Road Sign Interpretation Using Matching Pursuit Method", 4th IEEE Southwest Symposium on Image Analysis and Interpretation, pp. 202-208, April 2000.

[4] C. Fang, C. Fuh, S. Chen and P. Yen, "Road Sign Reccognition System Based on Daynamic Visual Model", pp 750-755, CVPR 2003.

[5] Y. Lauziere, D. Gingras, F.Ferrie, "A model-based road sign identification system", IEEE Computer Conference on Computer Vision and Pattern Recognition. CVPR 2001, pp. 1163-1170, 2001.

[6] P. Paclik, J. Novovicova, P. Somol, P. Pudil, "Road Sign Classification using Laplace Kernel Classifier", 11th Scandinavian Conference on Image Analysis, SCIA'99, Kangerlussuaq, Greenland, pp. 275-282, 1999

[7] P. Paclik, J. Novovicova, P. Somol, P. Pudil, " Road Sign Classification using Laplace Kernel Classifier ", Pattern Recognition Letters, vol.21, num.13-14, pp.1165-1173, 2000

[8] G. Piccioli et al., "Robust road sign detection and recognition from image sequences", Intelligent Vehicle Symposium, pp. 278-283, 1994.

[9] M. Betke and N. Makris, "Fast object in noisy images using simulated annealing", MIT A.I. Memo No. 1510, 1994.

[10] S. Azami et al., "Rout guidance sign identification using 2-D structural description", IEEE Intelligent Vehicle Symposium, pp. 153-158, 1996.

[11] S. Prince and R. Bergevin, "Road sign detection and recognition using perceptual grouping", International Symposium On Automotive Technology and Automation, pp.163-170, 1997.

[12] D. Gavrila and V. Philomin, "Real-time object detection for smart vehicle", Proceedings of ICCV '99, pp. 87-93, 1999

[13] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector", ECCV, pp. 128–142, 2002.

[14] T. Tuytelaars and L. Van Gool, "Wide baseline stereo matching based on local, affinely invariant regions", BMVC, pp. 412–425, 2000.

[15] W. Freeman and E. Adelson, "The design and use of steerable filters", PAMI, 13(9):891–906, 1991.

[16] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets", ECCV, pp. 414–431, 2002.

[17] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors", CVPR, pp. 257-263, 2003.