Real-time Illumination-invariant Speed-limit Sign Recognition Based on a Modified Census Transform and Support Vector Machines

Kwangyong Lim
Computer Science,
Yonsei University, 50 Yonsei-ro,
Seodaemun-Gu, Seoul, Korea
+82-2-2123-3876
kylim@yonsei.ac.kr

Taewoo Lee
Computer Science,
Yonsei University, 50 Yonsei-ro,
Seodaemun-Gu, Seoul, Korea
+82-2-2123-3876
gunmong@yonsei.ac.kr

Changmok Shin
Hyundai Mobis Co., Ltd.,
Mabuk-Ro, Giheung-Gu Yongin-Si,
Gyeonggi-Do, Korea
+82-31-8021-4735
lensless@mobis.co.kr

Soonwook Chung
Hyundai Mobis Co., Ltd.,
Mabuk-Ro, Giheung-Gu Yongin-Si,
Gyeonggi-Do, Korea
+82-31-8021-4735
sw.chung@mobis.co.kr

Yeongwoo Choi
Computer Science, Sookmyung
Women's University
Chungpa, Yongsan-Gu, Seoul, Korea
+82-2-710-9763
ywchoi@sookmyung.ac.kr

Hyeran Byun
Computer Science,
Yonsei University, 50 Yonsei-ro,
Seodaemun-Gu, Seoul, Korea
+82-2123-2719
hrbyun@yonsei.ac.kr

ABSTRACT

In this paper, we propose a robust illumination system for speed-limit sign recognition in real-time. Real-time traffic sign detection with various illuminations is one of the challenges in a vision-based intelligent vehicle system, as illumination varies greatly in real-world road images based on factors such as driving time, weather, lighting conditions, and driving directions. Our method uses a MCT (Modified Census Transform) as an illumination-invariant method for the real-time detection of traffic signs and uses a SVM (Support Vector Machine) as a classifier for detection and validation. With the proposed method, we have obtained a very high detection rate of 99.8% and recognition rates of 98.4% on various real-world driving images.

Categories and Subject Descriptors

I.4.9 [**Image Processing and Computer Vision**]: Application; I.5.1 [**Pattern Recognition**] Models – Neural nets; I.5.4 **[Pattern Recognition**]: Application – Computer Vision;

General Terms

Algorithms, Measurement, Verification, Experimentation

Keywords

Traffic Sign, Road Sign, Traffic Sign Detection, Traffic Sign Verification, Traffic Sign Recognition, Speed-limit Sign Recognition

1. INTRODUCTION

Advanced Driver Assistance Systems (ADAS) are systems that provide assistance in the driving process, which are becoming more common in modern vehicles. A Traffic Sign Recognition (TSR) system such as ADAS requires high accuracy and real-time processing. For real applications in ADAS, accurate sign detection is first required.

Various methods have been proposed for TSR [1,2]. In one German traffic sign recognition benchmark competition [3,4], the CNN/MLP method of Ciresan [5] achieved the highest recognition rate of 99.15%. However, the competition only addressed the classification of the sign content and did not include the detection of traffic signs in continuous images. In another German traffic sign detection benchmark competition, teams using the shape appearance, SVM (Support Vector Machine) and Haar-like features archived the highest detection rate [6,7,8]. However, this benchmark does not address the processing time.

Research into Traffic Sign Recognition remains active. However, it is difficult to satisfy the requirements both in real-time processing and in accuracy. Traffic signs have specific colors and shapes to attract the driver's attention. Color-based traffic sign detection methods were studied in [9,10,11], but these methods were sensitive to the illumination changes in real-world images. Shape-based traffic sign detection methods were studied in [12,13,14], and some of them archived high accuracies. However, their performances depend heavily on the preprocessing quality of the images. The main problem of traffic sign detection is that the image captured by the camera cannot maintain color constancy as the illumination changes.

In this paper, we will focus on the problems of circular (speed-limit) sign detection and recognition with various illumination changes. To achieve traffic sign detection, we propose a modified 8-bit MCT (Modified Census Transform). To recognize the contents of the signs, we propose a feature descriptor and a multi-level SVM structure. For testing, we used the images captured by a HDR (High Dynamic Range) camera installed in moving cars. These data consist of various road scenes including highways and city streets. We define each scene of a dataset as the moment of a sign appearance to

IMCOM (ICUIMC)'14, January 9–11, 2014, Siem Reap, Cambodia.

the moment of disappearance. This paper is organized as follows. Section 2 presents a 9-bit MCT and a modified 8-bit MCT including feature generation. Section 3 describes the method used for sign verification, and section 4 describes the multi-level SVM approach and descriptor. The experimental results are presented in Section 5.

2. TRAFFIC SIGN DETECTION USING AN 8-BIT MODIFIED CENSUS TRANSFORM

Illumination change is a serious problem in object detection and recognition. The image, as captured by the camera installed in the moving cars, reflects various environments, such as weather conditions and illumination changes. The color and shape of the traffic signs satisfy the standard law. Thus, many researchers have been using these characteristics. However, color and shape characteristics have weaknesses during the night or in backlit scenes. Thus, a robust detection method that is stable even with illumination changes is required. For this reason, the best feature to consider is only the reflectance value, without an illumination value in the intensity. In digital imaging, the brightness value $I(X)$ is defined as the product of the illumination value $L(X)$ and reflectance value $R(X)$, as follows:

$$I(X) = gL(X)R(X) + b \tag{1}$$

where X is the location of each pixel (x, y). According to equation (1), it is impossible to evaluate the reflection value $R(X)$ without any knowledge of the illumination value $L(X)$. However, if we ignore the illumination value $L(X)$, the remaining factor is only the reflectance value $R(X)$. As the Modified Census Transform (MCT) only uses the reflectance value $R(X)$ in equation (1), it is robust to illumination changes.

2.1 MCT (Modified Census Transform)

The modified census transform is a non-parametric local transform that is a modification of the census transform by B. Fröba and A. Ernst [15]. It is an ordered set of comparisons of pixel intensities in a local neighborhood to determine whether the intensity of pixels is higher or lower than the mean pixel intensity.

The feature is defined as a set of 3x3 kernels that emphasize the local spatial structure of an image. MCT is used to compute an index of the kernels and consists of an ordered set of binary comparisons of pixel intensities between all the pixels in the 3x3 kernels and the mean intensity of all the pixels of the kernels (Fig. 1).

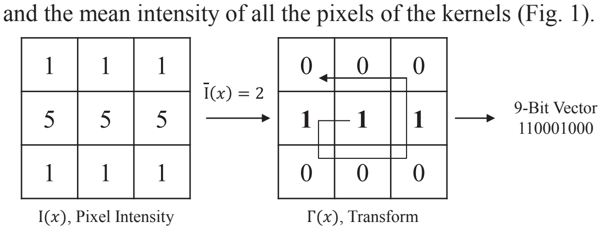

Figure 1. Example of the Modified Census Transform

Let $N'(x)$ be a 3x3 local spatial neighborhood of the pixel x such that $x \in N$, and $\bar{I}(x)$ is the intensity mean of the pixel intensities of the neighborhood. The MCT at x is defined as below:

$$\Gamma(x) = \otimes_{y \in N'} \xi(\bar{I}(x), I(y)) \tag{2}$$

where $\xi(\bar{I}(x), I(y))$ is a comparison function that yields 1 if $\bar{I}(x) < I(y)$, otherwise 0. The symbol $\otimes$ denotes the concatenation operation to form a bit vector(see Fig. 1). If we consider a 3x3 neighborhood, we are able to determine 511 structure kernels. Fig. 1 shows an example of the MCT. In this example, $\bar{I}(x)$ is 2.333, but we make it an integer for the fast computation. Each pixel on $N'(x)$ is transformed to 0 or 1 by applying a comparison function $\xi(\bar{I}(x), I(y))$. Bit vectors are concatenated in order, as shown in Fig. 1. Fig. 2 shows an illustrative example of the modified 8-bit MCT, visualized as an MCT kernel index image where the kernel index determines the pixel intensity.

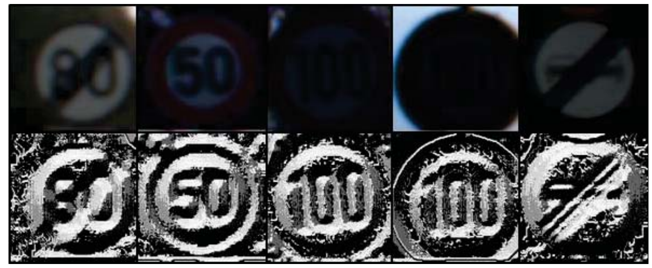

Figure 2. Examples of the illumination invariance of the MCT

2.2 Modified 8-bit MCT

The MCT feature consists of a 9-bit kernel index, which can express the total 512 types of local structures. The MCT feature can be used for both object recognition and object detection. In object detection, the 512 types of local structures can identify details at the object surface. It is a very useful element that can detect tiny defects on the object surface. However, a local structure made of the 8 surrounding pixels in the 3x3 window can also detect an object quite well. Therefore, the 9-bit MCT can consume unnecessary computation time and can create some problems such as the curse of dimensionality. The modified 8-bit MCT is also a non-parametric local transform that is a modification of the census transform. It is an ordered set of comparisons of pixel intensities in a local 8-neighborhood to find out whether the intensity of pixels is higher or lower than the mean pixel intensity. The feature is defined as a set of 3x3 kernels that emphasize the local spatial structure of an image. The MCT is used to compute the index of the kernels and consists of an ordered set of binary comparisons of pixel intensities between all the pixels in the 3x3 kernels and the mean intensity of all the pixels of the kernels, as shown in Fig. 3.

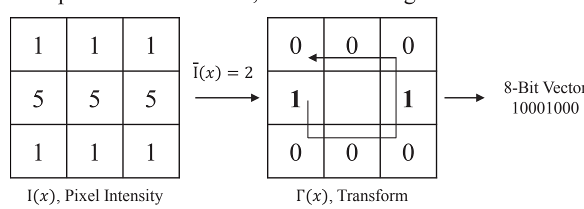

Figure 3. Example of the modified 8-bit MCT

The features of the modified 8-bit MCT can be created by using equation (1) but with the difference that the center pixel is excluded

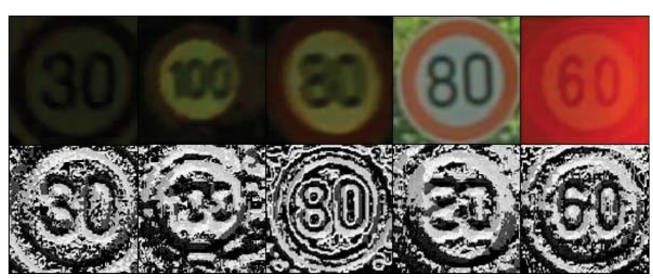

Figure 4. Examples of the illumination invariance of the modified 8-bit MCT

from the set of 3x3 window pixels. Fig. 4 shows an illustrative example of the modified 8-bit MCT, visualized as an MCT kernel index image where the kernel index determines the pixel intensity.

3. TRAFFIC SIGN VERIFICATION

A car can be driven in various areas and environments. Therefore, there are many road images including various real environments unlike the images used in computer vision research. The detected sign candidates can include false detections. In the downtown area, similar regions such as circular objects and emblems could be included as sign candidates, as shown in Fig. 5. Due to false detection, several regions of interest could be recognized incorrectly. In previous works, additional methods for sign verification are introduced involving color distribution and shape analysis to prevent false recognition. Verification methods using the traffic sign's ordinary features have achieved quite high results in a general environment image. However, in special environments such as night or backlit environments, verification is difficult. Therefore, we propose a feature for sign validation based on a histogram of local structures. The proposed feature can be written as follows:

$$f = \sum_{i=1}^{255} \frac{1}{v_{max}} H_i \tag{3}$$

where H represents the histogram of the kernel index for MCT in the ROI image and v_{max} is set to be the highest kernel index excluding the 0-th bin. In general, the kernel index 0 is the largest and has no edge response. Thus, before the normalization, histograms must be created excluding the 0-th bin. After the normalization, the sign candidates are verified using a binary SVM.

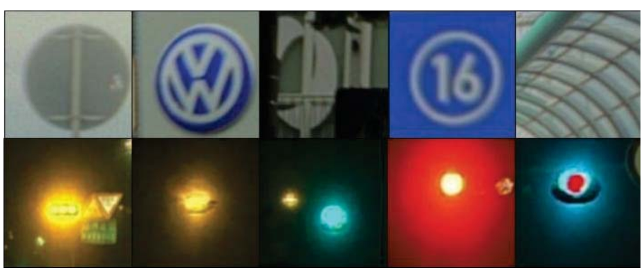

Figure 5. Examples of incorrectly detected ROIs

4. TRAFFIC SIGN RECOGNITION

After the regions are detected and verified, there are many approaches for content recognition. The most useful approach is extracting the sign pictogram followed by inner character recognition. Other approaches extract the feature vectors based on the pictogram intensity and use various classifiers to recognize them. Most recently, a method utilizing a Convolutional Neural Network (CNN) has been proposed to achieve a higher accuracy. Such approaches show a reasonable accuracy, but the image must be preprocessed to address extreme illuminations such as night or backlit scenes due to the high reliance on the intensity values. In this chapter, we propose a MCT feature descriptor, which is robust to illumination variance, and a recognition method using a multi-level SVM.

4.1 Feature Descriptor

In section 3, a method using modified 8-bit MCT feature histograms was introduced for sign verification. In the verification stage, the MCT feature itself is not only capable of classifying the observations into sign/non-sign but also capable of the recognition task. However, it is not sufficient to deal with detailed content to determine whether the target is a speed-limit sign. We achieved higher accuracy by adding another descriptor regarding spatial information in addition to the MCT feature. The proposed descriptor can be written as follows:

$$d = \sum_{i=1}^{4} \sum_{j=1}^{255} \frac{1}{v_{\max\_i}} H_{ij} \tag{4}$$

where H represents the histogram of the kernel index for MCT in the target region; i represents a pixel index in the region; and $v_{\max\_i}$ is set to be the highest kernel index excluding the 0-th bin. An example of the descriptor regarding spatial information is shown in Fig. 6.

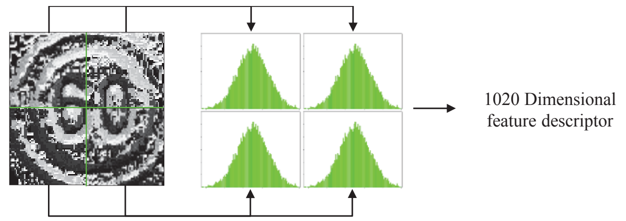

Figure 6. Example of feature descriptor extraction

4.2 Sign Recognition by Multi-level SVM

Traffic sign recognition methods using various methods are still considered. However, a multi-class SVM is a good candidate for sign recognition. In our work, the multi-class SVM should handle:

1) Non-target signs should be rejected.
2) Target signs should be correctly recognized.

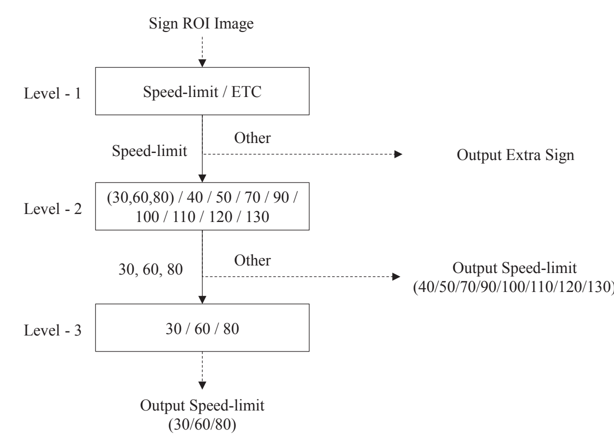

Figure 7. Example of a 3-level SVM structure

Even though the multi-class SVM performs quite well in recognition, it can still classify non-target objects as the correct classes. The classification performance of the multi-class SVM is lower than the performance of a binary SVM. As shown in Fig. 7, we solve this problem by using a multi-stage SVM structure.

We design a 3-stage SVM structure. The first stage is to distinguish between speed-limit signs and other signs. The second stage is to recognize the characters of the signs, and in this stage the classifier recognizes the numerals. However, the numerals 30, 60 and 80 are recognized as a single class and further classified in detail at the third stage due to their similar features and the possibility of confused recognition results.

5. EXPERIMENTAL RESULTS

In our experiment, the speed-limit sign detector is trained using the modified 8-bit MCT features of the positive and negative sign samples. The four cascaded strong classifiers are created through an Adaboost algorithm, and each strong classifier consists of a linear combination of several weak classifiers. We used 300 positive samples and 3,000 negative samples in the training. The entire procedure is performed on 4-level image pyramids to support various sized traffic signs. The lower half of the image is discarded to save computational costs because of the road. For the sign verification, a SVM is trained using 23,623 images consisting of 17,874 signs and 5,749 non-sign objects. For the sign recognition, a 3-level SVM is trained using 13,140 speed-limit signs and 4,734 extra signs.

In the experiments, we used the test data captured by a HDR (High Dynamic Range) camera installed in the moving cars. The data consist of various road scenes including highways, express ways, and local streets.

The driving scene dataset starts from the moment of the sign appearance to the moment of the sign disappearance. The scenes were taken under various lighting conditions such as sunny, rainy, cloudy, day and night times. This dataset contains 12,442 continuous images with a resolution of 960 × 504 pixels and consists of 372 driving scenes. Each scene has an average of 22 frames. Table 1 shows the specification of the test sets. The test sets consist of three individual sets, and the data were captured in Europe and South Korea.

To test the performance of the traffic sign recognition system, a frame-by-frame comparison is used in general. However, the accuracy of our final recognition result is evaluated more elegantly. The event scenes are detected from the raw sequence of the frames by two cascaded queues. This scheme is shown to be useful in real applications, regardless of several instantaneous noisy recognitions due to the temporal voting through the whole sequence of scenes containing the signs.

The detection and recognition results are summarized in Table 2 where the recall means the rate of correct detection and the precision means the rate of correct recognition. As shown in Fig. 8, the detection and recognition performance are measured in three different test sets. Our method achieved a 98.4% detection rate in the day-time test sets, as shown in Fig. 8 (a) and (b), and a 96.4% detection rate with 100% precision in the night test set, as shown in Fig 8 (c). However, the precision depends only on the recognition accuracy on detected sign candidates. Thus, we evaluate our overall accuracy with an F1 score regarding both detection and recognition, as below.

$$score_{F1} = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \quad (5)$$

Table 1. Specification of Test Sets

Dataset Name	Type	Frame length	Number of scenes	Resolution
Europe 1	Daylight, Rain, Cloudy	4,829	161	960 × 504
Korea 1	Daylight, Rain, Cloudy	4,528	99	960 × 504
Korea 2	Night	3,085	110	960 × 504

Table 2. Detection and Recognition Performance of the Proposed Method

Dataset	Signs	True Positive	False Negative	False Positive	Precision	Recall	F1 Score	Avg. Time
Europe 1	161	158	1	2	0.981	0.994	0.9874	54ms
Korea 1	99	99	0	2	0.980	1.000	0.9898	52ms
Korea 2	110	109	0	1	0.991	1.000	0.9954	52ms

(a)
(b)
(c)

(a) Europe daylight results (Europe 1) (b) Korea daylight results (Korea 1) (c) Korea night results (Korea 2)

Figure 8. Detection and recognition results with various driving scenes

The test sets that we used were captured while driving thousands of kilometers under various illuminations, which are difficult to handle for existing methods using the ordinary features of traffic signs. In addition, our test sets that were captured at night include high motion-blur, low intensity, and flare due to traffic lights, which all impose great difficulty. Nevertheless, the proposed method shows a high accuracy due to the MCT's robustness to illumination variance.

6. Conclusion

In this paper, we propose a robust illumination system for speed-limit sign recognition in real time. Real-time traffic sign detection with various illuminations is one of the challenges in a vision-based intelligent vehicle system, as the illumination changes in real-world road images vary wildly with factors such as the driving time, weather changes, lighting conditions, and driving directions.

Our detection approach utilizes a modified 8-bit MCT as an illumination-invariant method for the real-time detection of traffic signs. It is useful not only for detection but also for verification and recognition. This paper introduces a traffic sign recognition system using an MCT feature descriptor and a multi-level SVM. Consequently we adjust the recognition result by a voting process via event spotting while the signs appear in the view to further improve the recognition accuracy.

The proposed method guarantees robustness to various environmental, day/night and other illumination changes while achieving superior performance in computational cost. Future work will include the optimization of the algorithm and hardware for smaller devices, enabling real-life installation of the system.

7. ACKNOWLEDGMENTS

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government(MEST) (No. NRF-2012R1A1A2041343).

8. REFERENCES

[1] C. Bahlmann, Y. Zhu, and V. Remesh, 2005. A system for traffic sign detection, tracking, and recognition using color, shape and motion formation, *in Proceedings of the IEEE Symposium on Intelligent Vehicles*, 255-260.

[2] C. keller, C. Sprunk, C. Bahlmann, J. Giebel, and G. baratoff, 2008. Real-time recognition of u.s. speed sign, *in Proceedings of the IEEE Symposium on Intelligent Vehicles*, 518-523.

[3] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, 2011. The German Traffic Sign Recognition Benchmark: A multi-class classification competition, *in International Joint Conference on Neural Networks*, 1453-1460.

[4] German Traffic Sign Recognition Benchmark (GTSRB), http://benchmark.ini.rub.de/?section=gtsrb

[5] D. Ciresan, U. Meier, J. Masci and J. Schmidhuber, 2008. A Committee of Neural Networks for Traffic Sign Classification, *in Proceedings of International Joint Conference on Neural Networks*, 1918-1921.

[6] C.G. Kiran, L.V. Prabhu, R. V. Abdu and K. Rajeev, 2009. Traffic Sign Detection and Pattern Recognition Using Support Vector Machine, *in Proceedings of International Conference on Advances in Pattern Recognition*, 87-90.

[7] Jialin Jiao, Zhong Zheng, Jungme Park, Yi L. Murphey, 2009. A Robust Multi-class Traffic Sign Detection and Classification System using Asymmetric and Symmetric features, *IEEE International Conference on Systems, Man, and Cybernetics*, 3421-3427.

[8] Getman Traffic Sign Detection Benchmark (GTSDB), http://benchmark.ini.rub.de/?section=gtsdb

[9] J. N. Chourasia and G. H. Raisoni, 2010. Centroid Based Detec-tion Algorithm for Hybrid Traffic Sign Recognition System, *in proceedings of International Conference on Emerging Trends in Engineering and Technology*, 96-100.

[10] J. Abukhait, I. Abdel-Qader, Jun-seok Oh, O. Abudayyeh, 2012. Road sign detection and shape recognition invariant to sign defects, *IEEE International Conference on Electro/Information Technology (EIT)*, 1-6.

[11] Woong-Jae Won, Minho Lee, Joon-Woo Son, 2008. Implementation of Road Traffic Signs Detection Based on Saliency Map Model, *IEEE International Symposium on International Vehicle Symposium (IVS)*, 542-547.

[12] N. Barnes, A. Zelinsky, and L. Fletcher, 2008. Real-time speed sign detection using the radial symmetry detector, *IEEE Transactions on Intelligent Transportation Systems*, 9, 2, 322-332.

[13] R. Belaroussi and J.-p. Tarel, 2009. A real-time road sign detection using bilateral Chinese transform, *in proceedings of IEEE Symposium on Visual Computing*, 1161-1170.

[14] S. Houben, 2011. A single target voting scheme for traffic sign detection, *in Proceedings of IEEE Symposium on Intelligent Vehicles*, 124-129.

[15] B. Fröba, and A. Ernst, 2004. Face Detection with the Modified Census Transform, *in Proceddings of the sizth IEEE International Conference on Automatic Face and Gesture Recognition(FGR'04)*, 91-96.

[16] Y. Zhong, K. Karu, AK Jain, 1995. Locating Text in Complex Images, *Pattern Recognition*, 28, 1523- 1535.

[17] Y. Freund, and R. E. Schapire, 1997. A Decision Theoretic Generalization of On-Line Learning and an Application to Boosting, *Journal of Computer and Systems Science*, 55, 119-139.

[18] Y. Freund, and R. E. Schapire, 1999. A Short Introduction to Boosting, Journal of Japanese Society for Artificial Intelligence, Vol.14, No.5, pp.771-780, 1999.

[19] P. Viola and M. Jones, 2002. Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade, *Advances in Neural Information Processing Systems*, 2, 1311-1318.

[20] N. Barnes, A. Zelinsky, and L. Fletcher, 2008. Real-time speed sign detection using the radial symmetry detector, *IEEE Transactions on Intelligent Transportation Systems*, 9, 2, 322-332.