Traffic Sign Recognition using Multi-Class Morphological Detection

Thien Huynh-The *, Hai Nguyen Thanh †, Hung Tran Cong ‡
* Department of Computer Engineering
Kyung Hee University, Gyeonggi-do, 446-701, Korea
Email: thienht@oslab.khu.ac.kr
†Faculty of Electrical and Electronic Engineering
University of Technical Education, Hochiminh, Vietnam
Email: nthai@hcmute.edu.vn
‡Training & Science Technology Department
Posts and Telecoms Institute of Technology, Hochiminh, Vietnam
Email:conghung@ptithcm.edu.vn

***Abstract*—In this paper, a novel method is proposed for the Traffic Sign Recognition (TSR) using the Principle Component Analysis (PCA) and the Multi-Layer Perceptrons (MLPs) network. In particular, the candidate signs are individually detected from two chroma components in the YCbCr space and then classified into three shape classes: circle, square, and triangle based on computing the rotated version correlations. The PCA-based features of these objects will be used for the MLPs as the training system corresponding to previously determined class. This approach not only reduces the time but also increases the performance of the recognition process. In simulation, the proposed method is estimated with over 500 statistic images for the accuracy rate up to 96%.**

Keywords - Intelligent Transport System; Traffic Sign Recognition; Principle Component Analysis; Multi-Layer Perceptrons; Morphology Detection.

I. Introduction

1. As one of the important fields in the Intelligent Transport System (ITS), the TSR system immediately gives the warning from road signs and suggests some advices to drivers. Based on the cascade of the Support Vector Machines (SVM), candidate regions containing traffic objects as the Maximally Stable Extremal Regions (MSERs) [1] have been recognized with the set of the Histogram of Oriented Gradient (HOG) features. Zaklouta et al [2] suggested a real-time traffic detection system as a component of the Driver Assistant Systems (DAS) for only circular and triangular signs based on the HOG-based SVM algorithm. In order to improve the accuracy rate, the segmentation using the enhanced red color channel was applied for an identification process with the tree classifiers. In [3], the recognition system using the linear SVM for a shape classification was proposed in a traffic-sign detection stage. Due to the Hue Saturation Intensity (HSI)-based segmentation, the candidate objects were recognized by the Gaussian-kernel SVM approach. Another approach, proposed by Deli Pei et al [4], was focused on an extension of the original unsupervised model with an additional supervised term to restrain the classification errors of the recovered feature representations, called the Supervised Low-Rank Matrix Recovery (SLRMR) model.

The eigen-based TSR [5] was represented by applying the PCA algorithm to extract the most significant components of input images for the categorization process. A set of weights was computed from the most effective eigenvectors of a database and then unknown objects would be classified by the Euclidean distances. In addition, the existing method based on a computation of color eigenvectors was proposed by Tsai et al [6], in which the extracted candidate signs from road scenes would be recognized by the Radial Basis Function (RBF) network. In another research, a proposed framework to extract candidate regions containing information about color segmentation, shape simplification, and shape decomposition was presented by Fleyeh [7]. The recognition task could be achieved through the direct matching with templates for closed candidate shapes. For unclosed shapes, it needs to be computed the minimum geometric differences between objects and templates . As a novel approach to detect and recognize traffic signs based on Ridge Regression [8], a precise segmentation in the RGB color space was obtained with the same performance as other learning machines. Moreover, to resist the illumination variations and distortions, Jiang et al extracted features using the Otsu method for a recognition process. However, this method is unappropriate for the subsequent frames.

The challenges in improving the accuracy rate of classification and recognition for traffic signs can be resolved by a proposed method with main contribution as follows:

- Classify the doubt objects into geometric classes such as, circle, square, and triangle.
- Improve the recognition performance by applying the Multi-Layer Perceptron networks for each geometric class.
- Reduce computation cost with Principle Component Analysis algorithm for feature extraction.

For detail, a novel TSR method in this paper is proposed by the morphological detection and classification to identify candidate objects. Based on the examination of correlations of each object and morphological samples under various degrees in rotation, the appropriate classes for these objects will be

978-1-4799-6956-2/14/$31.00 ©2014 IEEE

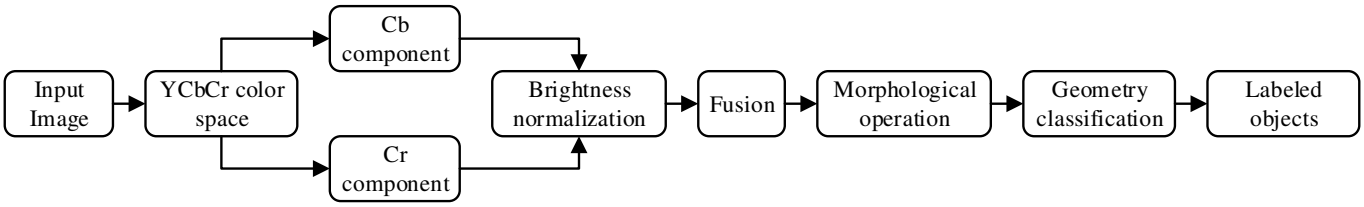

Fig. 1. The flowchart of the morphological detection and classification.

decided. The recognition process as the combination of the PCA algorithm and the MLPs network is used to deal for the high accuracy rate. This paper is organized as follow: the section II described the proposed morphological detection and classification. The recognition based on PCA-MLPs network will be represented in section III. Finally, the experimental results and the conclusion are written in section IV and V in this paper, respectively.

II. Morphological Detection and Classification

In a proposed framework of the detection and classification stage, the authors detect doubt objects in separated components in the YCbCr color space based on the Canny method. For the next step, these objects is then filtered by the scaling condition to reject unfitting items, and fitting objects will be classified into three groups by mapping their shapes to the existed samples. In order to be clear, the detail of this section is shown in the Fig.

A. Doubt Region Extraction Using Contrast Enhancement

In this stage, the authors propose an effective detection method based on the morphological analysis. The innovation of this method is that candidate objects are detected and categorized based on the morphology to reduce the time and improve the accuracy rate of the recognition process. The Canny segmentation is employed as the edge detection technique by looking for the local maxima of the gradient magnitudes. In particular, the level of edging can be controlled through the threshold value. Moreover, to improve the performance of the segmentation process, the input video frames will be analyzed in the YCbCr color model instead of in the RGB model [2]. These input color frames need to be converted to the YCbCr color model and separated into three components: the luminance Y, the blue-difference Cb, and the red-difference Cr. The analyzing based on the YCbCr color space is appropriate for the chrome of traffic signs when the most of them are usually red or blue. Compared with the RGB-based analysis, the areas containing doubt objects are quite difficult to observe clearly due to the low contrast. A sample of the YCbCr separation is shown in the Fig. 2, in which the red or blue regions are not easy to identify. However, the contrast are fully increased for each component through the following equation:

$$g_o = L \times \left(\frac{g_i - g_{\min}}{g_{\max} - g_{\min}} \right) \quad (1)$$

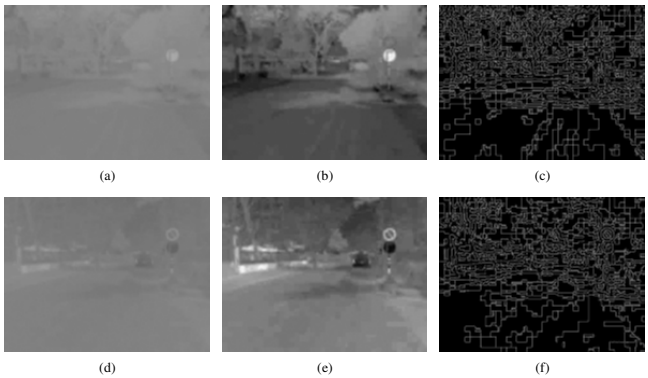

Fig. 2. The Canny segmentation for two separated color component. First row: Cb-component. Second row: Cr-component

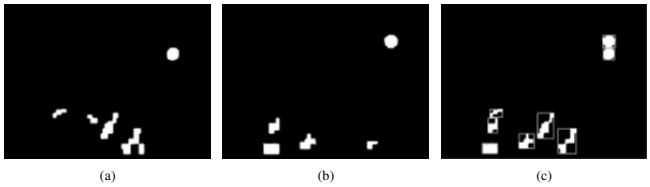

Fig. 3. Detection of doubt regions based on the fusion of two components: (a) Cb-component, (b) Cr-component, (c) The fusion.

where g_o and g_i are the gray values of the input image and the output image. $g_{\min}$ and $g_{\max}$ are the maximum and minimum gray values of the input image, L is the range of the output histogram. Basically, the equation (1) is also a linear transformation of brightness to increase the contrast.

The areas containing objects can be plainly observed when applying the contrast enhancement as shown in the Fig. 2(b) and (e). In addition, the binary output images of the sample after segmentation with the Canny method are also shown in the Fig. 2(c) and (f), corresponding to Cb and Cr. To determine doubt regions containing signs, the holes, a set of background pixels, which can not be reached by filling on the background from edges, are blocked up in the binary images as shown in the Fig. 3(a) and (b). The fusion of them is represented as the Fig. 3(c) using the following equation:

$$F = Cb \cup Cr \quad (2)$$

where Cb and Cr denoted the binary images after filling holes corresponding to two separated components.

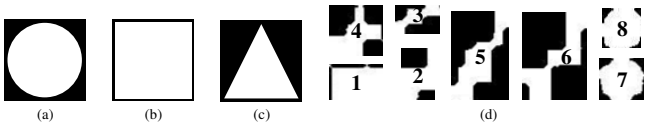

Fig. 4. Three templates corresponding to three groups for classification and the objects have been extracted as the individual images: (a) Triangle, (b) Circle, (c) Square, (d) 8 Doubt Objects.

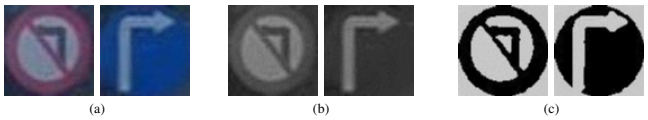

Fig. 5. Normalization of two extracted objects based on their positions: (a) Cropped version, (b) Gray-scale version, (c) Normalized version.

B. Morphological Classification Based On Correlation

In the next step, these objects are individually extracted into the separated images and categorized based on the morphology. The main idea is that these objects need to be compared with the templates which is corresponding to three classes, such as triangle, circle, and square as shown in Fig. 4(a) - (c). Classification of these objects, that can reduce the time of recognition on a neural network, can be listed in detail through the following steps:

1) Remove objects with unacceptable size: If the doubt objects are candidate traffic signs, the bounding boxes of them as shown in Fig. 3(c) need to be the square (for circle or square signs) or the rectangle (for triangle signs). From this geometric specification, the authors utilize the ratio of widths, denoted C, which can be determined by the following equation:

$$C = \frac{x\_width}{y\_width} \tag{3}$$

where x_width and y_width denote the size of extracted object, and C can be seen as the size condition in this research. As the recommendation, the value of C has been set in the range $(0.8 \leq C \leq 1.2)$. The suggested range is proposed to avoid ignoring candidate objects due to the effects of perspective projection.

2) Compute the normalized correlation factors: The main ideal in this step is that the correlations of the doubt objects and each template are computed and compared for categorization. In particular, both template and object will be rotated with the angle $\alpha = 30$ degrees for each step in accumulation to determine the correlation factors. The decision of class for the object depends on the mean value:

$$m\_nc2_{o,s} = \frac{\sum_{i=1}^{N} nc2_{o,s,i}}{N} \tag{4}$$

where $m\_nc2_{o,s}$ is the mean of normalized correlation factors between the o object and the s template, $nc2_{o,s,i}$ is the normalized correlation when rotating with the angle $\beta = i \times \alpha$ degrees for both, and N $\left(= \frac{360}{\alpha}\right)$ is the total number of rotated versions. The equation for computing correlation is expressed as follows:

$$nc2 = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} O(i,j) \times S(i,j)}{m \times n} \tag{5}$$

where $O_{i,j}$ and $S_{i,j}$ are the gray values located at coordinate (i, j) of the object and the template which have been normalized in the same size $m \times n$ (set at 80×80). It is noted that the values of $O(i,j)$ or $S(i,j)$ are the set 1 for bit 1 and -1 for bit 0. Therefore, the value of $(O(i,j) \times S(i,j))$ is either 1 or -1.

For each object, three values corresponding to the means of correlation factors between the object and three templates are determined. The object is chosen as the candidate sign for recognition by comparing these coefficients to the threshold referring the acceptable level of the object. In particular, the algorithm for classification can be described based on the consideration of maximum of means (4):

$$\max\_nc2_o = \max\left(m\_nc2_{o,s=cir,squ,tri}\right) \tag{6}$$

where $\max\_nc2_o$ is the maximum of the mean correlation values for all classes.

Categorization is produced based on comparing $\max\_nc2_o$ to the threshold y. In particular, if $\max\_nc2_o$ is larger than threshold, the doubt object will be considered as the area containing traffic sign. Furthermore, the geometry-based classification is also achieved by examination of $m\_nc2_{o,s=cir,squ,tri}$ with some rules which are summarized in the Table I as the clear explanation. The result of classification for the sample image is represented in the Table. II, in which only two objects 7 and 8 detected as the candidate signs, will be recognized by the circle sign database. Based on the positions, two objects

TABLE I
CHECKING CONDITIONS FOR CLASSIFICATION (F: FIT, UF: UNFIT).

Case	C (3)	Threshold y	$\max\_nc2_o$	Decision
1	UF	NaN	NaN	Reject
2	F	Smaller	NaN	Reject
3	F	Larger	$m\_nc2_{o,cir}$	Circle
4	F	Larger	$m\_nc2_{o,squ}$	Square
5	F	Larger	$m\_nc2_{o,tri}$	Triangle

TABLE II
RESULT OF THE CLASSIFICATION FOR A SAMPLE WITH $y = 0.8$

Object	C (3)	$m\_nc2_{o,cir}$	$m\_nc2_{o,squ}$	$m\_nc2_{o,tri}$	Decision
1	0.659	NaN	NaN	NaN	Reject
2	1.552	NaN	NaN	NaN	Reject
3	0.641	NaN	NaN	NaN	Reject
4	0.957	0.275	0.274	0.429	Reject
5	1.529	NaN	NaN	NaN	Reject
6	1.351	NaN	NaN	NaN	Reject
7	0.949	**0.851**	0.567	0.217	Circle
8	1.029	**0.817**	0.675	0.249	Circle

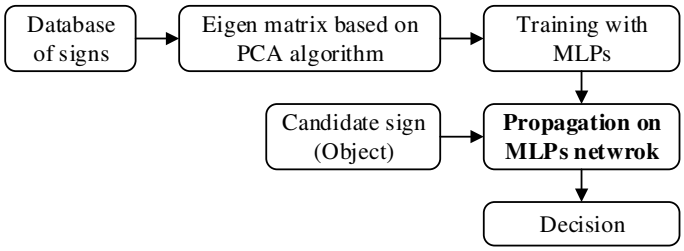

Fig. 6. The flowchart of recognition based on the PCA-MLPs algorithm.

are extracted as the individual images to normalize as shown in Fig. 5.

III. Recognition based on the PCA-MLPs Algorithm

For the recognition, the authors used the PCA algorithm to extract the object's features which will be then propagated on the MLPs network with the trained set. In order to be more detail, the algorithm is described in Fig. 6.

A. The PCA Algorithm for Images

In recent years, the Pincipal Component Analysis (PCA) that alters the data into the new structure based on its variance was used as the solution for face recognition [9]. In particular, the 2-D image is represented as the 1-D vector by concatenating each column (or row) into a vector as follows:

$$X_j = \begin{bmatrix} x_{1,j} \\ x_{2,j} \\ \vdots \\ x_{N,j} \end{bmatrix} \tag{7}$$

where N is the total pixels in an image. $x_{1,j}$, $x_{2,j}$ ... $x_{N,j}$ are the the gray value of the j^{th} image. For M images with the same size, the set of the 1-D vectors is described as:

$$X = \begin{bmatrix} X_1 & X_2 & \cdots & X_M \end{bmatrix} \tag{8}$$

where X_1, X_2 ... X_M are the 1-D vectors of the images. Let m represent the mean image.

$$m = \frac{1}{M} \begin{bmatrix} x_{1,1} + x_{1,2} + & \cdots & +x_{1,M} \\ x_{2,1} + x_{2,2} + & \cdots & +x_{2,M} \\ \vdots & \ddots & \vdots \\ x_{N,1} + x_{N,2} + & \cdots & +x_{N,M} \end{bmatrix} = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_N \end{pmatrix} \tag{9}$$

The images centered by subtracting the mean image from each image vector are represented as the following equation:

$$w_j = X_j - m \tag{10}$$

From (9), the matrix W, the set of 1-D vectors w_j, is constructed by placing side by side the column vectors w_j. Instead of computing the eigenvectors and eigenvalues of the covariance matrix WW^T with the size $N \times N$, the the eigenvectors d_j and eigenvalues μ_j of the covariance matrix W^TW with matrix of size $M \times M$ are computed based on the common theorem in linear algebra as follows:

$$W^TWd_j = \mu_j d_j \tag{11}$$

By multiplying left to both sides by W:

$$W^TW\left(Wd_j\right) = \mu_j \left(Wd_j\right) \tag{12}$$

which means that the first M eigenvectors and eigenvalues of WW^T are given by Wd_j and μ_j. The eigenvectors are sorted from high to low according to their corresponding eigenvalues. In particular, the smallest eigenvalue is associated with the eigenvector that finds the least variance. A facial image is projected on L $(L \ll M)$ dimensions using the equation:

$$\Omega = \left[v_1 v_2 \ldots v_L\right]^T \tag{13}$$

where v_i is the i^{th} coordinate of the facial image in the new space, which come to be the principle component.

B. The MLPs Neural Network

The Multi-Layer Perceptrons (MLPs) have been assigned to implement the recognition core after feature extraction. As the improvement of original perceptrons, the MLPs have been upgraded by cascading one or more extra layers [10], called hidden layers, which are not directly connected to the external environment. The main algorithm of the MLPs learning is the propagation of the error backwards. This method can be shown briefly as:

• Correct the output layer of weights using the following formula:

$$w_{ho} = w_{ho} + \left(\eta \delta_o o_h\right) \tag{14}$$

where w_{ho} is the weight connecting the hidden node h with the output node o. η is the learning rate and o_h is the output at hidden node h. The δ_o is given by the following equation:

$$\delta_o = o_o\left(1 - o_o\right)\left(t_o - o_o\right) \tag{15}$$

where o_o is the output at the node o of the output layer, and $(t - o)$ is the target output for that node.

• Correct the input weights using the following equation:

$$w_{ih} = w_{ih} + \left(\eta \delta_h o_i\right) \tag{16}$$

where w_{ih} is the weight connecting node i of the input layer with the node h of the hidden layer. η is the learning rate and o_i is the input at the node i of the input layer. In (16), δ_h is described as follows:

$$\delta_h = o_h\left(1 - o_h\right)\sum_o \left(\delta_o w_{ho}\right) \tag{17}$$

.

• In order to calculate the error E, the following function can be utilized by taking the average difference between the target and the output vector:

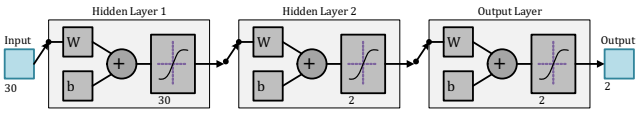

Fig. 7. Structure of MLPs network for each class.

TABLE III
SOME PARAMETERS OF THE MLPS NETWORK FOR EACH CLASS

Parameters	Value
Activation function	Sigmoid
Performance Function	MSE
Number of hidden layers	2
Number of inputs	30
Number of outputs	2
Number of neurons in the first hidden layer	30
Number of neurons in the second hidden layer	2
Maximum Number of training Epochs	500
Minimum Performance Value	10^{-6}

$$E = \frac{\sqrt{\sum_{n=1}^{p} (t_o - o_o)^2}}{p} \tag{18}$$

where p is the number of nodes in the output layer, t_o and o_o are the target and output vectors at the output node o.

In simulation, the authors used 50 samples for 2 signs as the database for each class. However, only 30 principle components corresponding to 30 eigenvectors having largest eigenvalues are chosen for training process. It is noted that there are three networks are generated to respond to three classes. In particular, the candidate sign will be identified by the appropriate network. The MLPs network in this paper has been designed for the number of nodes in the output layer to equal to the number of signs and the number of neurons in the input layer corresponding to the number of eigenvectors. The training system is implemented based on the Mean Squared Normalized Error (MSE) performance function with some settings and parameters as listed in Table III. For each object, the node of an output layer will express two coefficients which are the sums of weights. By comparing the maximum of them to the threshold, the current object is decided as the sign in database or not:

Maximum of sum of weights is less than threshold $\Rightarrow$ non-signs.

Maximum of sum of weights is greater than threshold $\Rightarrow$ signs.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Some results of simulating the proposed method for statistic images are represented in this section. There are 500 test images from urban and rural region used for evaluation of 6 traffic signs with three classes. The Fig. 8 shows the traffic signs which need to be recognized and the Fig. 9 represents some samples for recognition. More results are also shown in the Table. IV, IV, and VI. The authors divided an

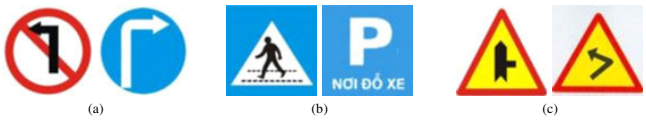

Fig. 8. Set of traffic signs used for evaluation.

Fig. 9. Result of simulation with some sample images.

assessment process into two stage: the object extraction and the recognition stage.

In the first simulation, the test images are estimated for both detection and classification step. Objects detected from each image have to include at least one candidate signs and others. Therefore, the detection will be unsuccessful if there is no any candidate sign in a collection of extracted objects. Moreover, an assessment for multi-sign samples is also considered. It is important to note that the classification is only implemented for successful samples in the detection. In practice, these images have been captured on Vietnam roads which have many objects such as electric posts with wiring maze, street lights, or advertisement boxes can negatively affect the detection performance. Furthermore, the intermittent weather can be seen as the challenge in dangerous warning for drivers.

For the second simulation as shown in the Table VI, the Fail[1] represents event of incorrect classification in the first stage, meanwhile, the Fail[2] denotes event of incorrect recog-

TABLE IV
EXPERIMENTAL RESULTS OF MORPHOLOGICAL DETECTION

Class	No.Inputs	Detection rate	False Alarm Rate
Circle	150	98.0%	2.0%
Square	200	96.0%	4.0%
Triangle	150	97.3%	2.7%

TABLE V
THE CONFUSION MATRIC OF MORPHOLOGICAL CLASSIFICATION

	Circle	Square	Triangle	Recall
Circle	1.000	0.000	0.000	1.000
Square	0.020	0.958	0.022	0.958
Triangle	0.025	0.016	0.959	0.959
Precision	0.957	0.984	0.977	

TABLE VI
EXPERIMENTAL RESULTS OF RECOGNITION.

Class	Input	Undetected	Fail[1]	Fail[2]	Success
Circle	150	2.0%	0.0%	1.3%	96.7%
Square	200	4.0%	4.0%	0.0%	92.0%
Triangle	150	2.7%	4.0%	0.7%	92.6%

nition. Through the results in the Fail^2 column, the method achieves the high performance in recognition based on the PCA-MLPs network due to positive effect of classification. The candidate sign is recognized in the 2-signs MLPs instead of the 6-sign MLPs network, that is, not only the accuracy rate is improved but also the time for identification is reduced significantly.

Based on the Scale-Invariant Feature Transform (SIFT), the Kus's method [11] only achieved the best performance when adding the color and orientation information. As a comparison, the proposed method is simpler than Kus's research when using less features for the training. In order to achieve almost 90% in accuracy with the SIFT feature extraction, Guo [12] represented the visual attention mechanism as a novel method. However, the drawback of Guo's method is that the performance is reduced in disturbance conditions.

V. CONCLUSION

In this paper, a novel traffic sign recognition method was represented using the PCA-MLPs algorithm with two main stages. In the morphology detection and classification, the objects were determined separately from two chroma components of the YCbCr space and then categorized into three class: triangle, circle, and square by computing the correlations of these objects and samples in different rotated versions. The PCA-MLPs recognition is implemented simpler and faster due to the classification. In addition, the method was evaluated on statistic images captured on Vietnam roads with the accuracy rate up to 96%. However, the method need to be improved in segmentation step to achieve the better results.

REFERENCES

[1] J. Greenhalgh and M. Mirmehdi, "Real-time detection and recognition of road traffic signs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, pp. 1498–1506, 2012.

[2] F. Zaklouta and B. Stanciulescu, "Segmentation masks for real-time traffic sign recognition weighted hog-based trees," *IEEE conference on Intelligent Transportation Systems*, pp. 1954–1959, 2011.

[3] S. Maldonda-Bascon, S. Lafuente, P. Gil-Jimenez, H. Gomez-Moreno, and F. Lopez-Ferreras, "Road-sign detection and recognition based on support vector machines," *IEEE Transactions on Intelligent Transportation System*, vol. 8, pp. 264–278, 2007.

[4] D. Pei, F. Sun, and H. Liu, "Supervised low-rank matrix recovery for traffic sign recognition in image sequences," *IEEE Signal Processing Letters*, vol. 20, pp. 241–244, 2013.

[5] H. Fleyeh and E. Davami, "Eigen-based traffic sign recognition," *IET Intelligent Transport System*, vol. 2, pp. 190–196, 2011.

[6] L.-W. Tsai, J.-W. Hsieh, C.-H. Chuang, Y.-J. Tseng, and C.-C. L. K.-C. Fan, "Road sign detection using eigen colour," *IET Computer Vision*, vol. 2, pp. 164–177, 2008.

[7] S. Xu, "Robust traffic sign shape recognition using geometric matching," *IET Intelligent Transport System*, vol. 3, pp. 1301–1309, Nov 2009.

[8] Y. Jiang, S. Zhou, Y. Jiang, J. Gong, G. Xiong, and H. Chen, "Traffic sign recognition using ridge regression and otsu method," *IEEE Intelligent Vehicles Symposium,*, 2011.

[9] K. I. Kim, K. Jung, and H. J. Kim, "Face recognition using kernel principal component analysis," *IET Signal Processing Letters*, vol. 9, pp. 40–42, 2002.

[10] S. Haykin, Ed., *Neural Networks and Learning Machines*. New Jersey, USA: Prentice Hall, 2011.

[11] M. C. Kus, M. Gokmen, and S. Etaner-Uyar, "Traffic sign recognition using scaling invariant feature transform and color classification," *International Symposium on Computer and Information Sciences*, pp. 1–6, 2008.

[12] H.-R. Guo, X.-J. Wang, Y.-X. Zhong, and P. Lu, "Traffic signs recognition based on visual attention mechanism," *The Journal of China University of Posts and Telecommunications*, vol. 189.