A METHOD FOR FEATURE EXTRACTION OF TRAFFIC SIGN DETECTION AND THE SYSTEM FOR REAL WORLD SCENE

Jung-Guk Park, Kyung-Joong Kim
machinelearning@sju.ac.kr, kimkj@sejong.ac.kr

Department of Computer Science and Engineering
Sejong University, Seoul, Korea

ABSTRACT

Traffic sign detection is the significant step before recognizing the class of traffic signs. In the detection, most studies rely on region of interest (ROI) from color information. In practice, however, there is no way to cover the various conditions such as illumination effects or weather conditions. To overcome the problem, this work uses the ROI-free detection by the supervised learning in which the predictor trains the positive examples of traffic sign image and negative examples of non traffic sign image. The proposed method is robust to illumination effects although it searches the traffic sign over the input scene. Because the real world scene often contains occluded or overlapped traffic signs, it is required that the detection algorithm should handle the cases. In this work, we introduce a novel feature extraction method inspired by vision perception theory developed in biological system and by power spectrum in frequency domain. The method was combined with support vector classifier. The proposed method showed accurate classification results (99.32%, 5-fold cross validation) over combined image sets of positive and negative traffic signs samples. Finally, we compared the detection ability of the proposed method and a previous work using ROI on real-world traffic scenes.

Index Terms— traffic sign detection, machine learning, Fourier Transform, Support Vector Machine, failure tolerance

1. INTRODUCTION

Traffic sign detection has been tackled by various approaches such as rule-based method, and hybrid algorithms. Most works extracted region of interest (ROI) with color information and were sensitive to weather or illumination effect to verify the traffic sign [1], [2]. It is challenging to exactly extract ROI regardless of such environmental conditions. For example, the segmentation of traffic signs using color pixel is not working for complex or overlapped scenes and requires additional pre-processing

procedure. Also, the previous works tested their approaches on some simple scenes or highways [3], [4]. Recently, there are several detection works using machine learning algorithms instead of the color-based ROI approaches. For example, convolutional neural network [5] and support vector machines have been used for detection over real-world scenes [6]. The fatal fact of scale-space based search methods require high computational cost c for the verification, $c \approx \sum_{i=0}^{n-1} w_i \times h_i$ where n denotes the number of the layer of scale-space, w_i and h_i represent the height and width of the i th layer image, they conclude robust detection. However, CUDA (Compute Unified Device Architecture) or parallel processing on GPU results in the successive cost.

In this work, we propose a novel method to detect traffic sign using machine learning algorithms, and scale-space is adopted to handle the different scale of traffic signs. Inspired by visual recognition system [7], we can obtain local features, and power spectrum analysis is used to extract global features of a cropped image of the sliding window. Figure 1 shows the overview of the proposed system consisting of n layers.

The rest of this paper is organized as follows: section 2 details the proposed feature extraction, and section 3 shows the experimental results. Section 4 concludes this work and discusses the future works.

2. FEATURE EXTRACTION FOR TRAFFIC SIGN

2.1. Local feature extraction

The object outline provides useful information for visual perception, and it has been widely used in computer vision. Thus, we generate Laplace of Gaussian (LOG) image from the input so that edge image is obtained and unnecessary information is removed. The following step contains that the edge image is converted to the feature images by Gabor filters which are similar to those of the human visual system. Once the edge image is produced from a given image, then Gabor filter convolves it as follows
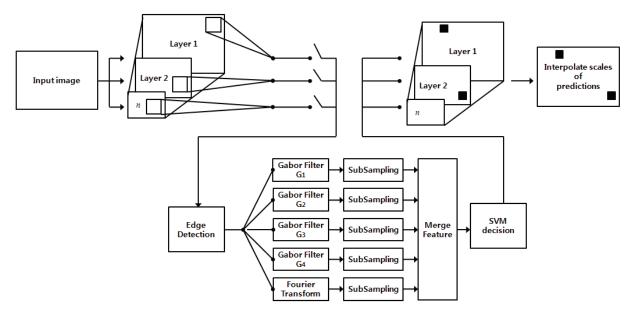
Figure 1. The overview of the proposed system (SVM verifies each layer of pyramid image for the traffic sign)

$$I_{G_k}(x,y) = G_k(x,y) \otimes I_{LOG}(x,y) \quad (1)$$

where k is the index of the filter, $I_{LOG}(x,y)$ and $I_{G_k}(x,y)$ represent the LOG image and the k th Gabor filtered image, respectively. $\otimes$ denotes convolution operator, and the filter $G_k(x,y)$ is the same to

$$G_k(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left(\frac{x_{\theta_k}^2}{\sigma_x^2} + \frac{y_{\theta_k}^2}{\sigma_y^2}\right) + j2\pi W x_{\theta_k}\right\} \quad (2)$$

where $x_{\theta_k} = x\cos\theta_k + y\sin\theta_k$ and $y_{\theta_k} = -x\sin\theta_k + y\cos\theta_k$, σ_x^2 denotes variance for x-axis and σ_y^2 is for y-axis. The term W means spatial frequency. The family function of Equation (2) used in our work is as follows

$$G_{\lambda,\theta_k,\varphi,\sigma,\gamma}(x,y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \varphi\right) \quad (3)$$

in which x' and y' are the same to x_{θ_k} and y_{θ_k} in equation 2, respectively, λ denotes the strength of wavelet, φ is the value which specifies the phase offset of the cosine factor of the function. The term γ affects the ellipticity of the Gaussian function, and the term σ is calculated using the additional variable b which denotes the bandwidth of the filter. The proposed feature is extracted using the four Gabor filters which lie on the different orientations $0°, 45°, 90°, 135°$ to yield the local directional information

and $0 \le k < 4$, because traffic signs can be represented by the four directions.

2.2. The feature enhancement of the power spectrum

Fourier transform converts the signal in spatial domain to frequency domain, and Discrete Fourier Transform (DFT) is widely used for signal processing that is

$$F(k,l) = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)e^{-j2\pi(\frac{kx}{N} + \frac{ly}{N})} \quad (4)$$

where N denotes both the height and width of image, $f(x,y)$ is the edge image in the spatial domain, and the exponential term is the basis function corresponding to each point $F(k,l)$ in the Fourier space. The equation can be interpreted that the value of each point $F(k,l)$ is obtained by multiplying the spatial image with the corresponding base function and summing the result. In the last step, the Gabor filtered images in the section 2.1 and Fourier transformed image are subsampled as follows

$$S(x,y) = \frac{1}{\mu^2} \sum_{i=-\frac{\mu}{2}}^{\frac{\mu}{2}} \sum_{j=-\frac{\mu}{2}}^{\frac{\mu}{2}} I(x+i, y+j) \quad (5)$$

where $S(x,y)$ represents the subsample image from the filtered images $I(.)$ which can be the Gabor filtered images or Fourier transformed image. In Equation (5), μ denotes the length of the subsampling interval.
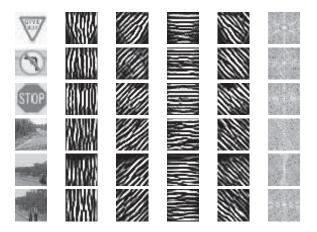
14

Figure 2. The examples to train the classifier and the corresponding features by the proposed method

3. EXPERIMENTAL RESULTS

The experimental results consist of two parts, the one is the proposed feature extraction evaluation, the other shows our simulation of traffic sign detection on the real world scene. Our training examples have two classes as traffic sign image is the positive class and the opposition is non traffic sign image denoting the negative class. The traffic sign image covers the triangle, circular, hexagon and inverted triangle traffic sign. The non traffic sign images are collected from natural road scenes where traffic signs do not appear and they are cropped by the randomly posited quadrangle in the road scenes. The number of the positive examples is 961, and the number of the negative examples is 8550 for learning the various unexpected conditions. Figure 2 describes the images of the positive and the negative examples. The first column shows input images, Gabor filtered features are represented from the second to the forth, and the last column is for Fourier transformed images. Top 3 rows show the positive examples, and bottom 3 rows are the negative examples. In feature extraction, the image to be verified is normalized to 32 by 32 pixels. The parameters of Gabor filters are set to 4 for λ, γ is 0.5, b is equal to 1, and φ is 0 that yields the best performance with 5-fold cross validation. The subsampling factor μ is 8, which generates the 64-dimensional feature vector by the 4 Gabor filters and 16-dimensional feature vector by the fast Fourier transform. The merged feature dimensionality is equal to 80.

The first experiment includes Naïve Bayes classifier, Bayes network, RBF neural network, and polynomial SVM for the performance evaluation of the features. The first two are based on generative method and the others are discriminative method, all predictors use default parameters. Table 1 and Table 2 summarize the positive class evaluation of the Gabor filtered features using receiver operating characteristic (ROC) analysis with 5-fold cross validation and the improved results of each classifier performance with the power spectrum feature, respectively.

Traffic sign detection with the proposed method is simulated on real-world scenes from Google web site and [2]. The input image has no preprocessing method. The layer of pyramid image is 7, the ratio between the image sizes of the above layer and of the below one is 1.2, and the prediction interval ϱ is 2. SVM employs the polynomial kernel $\kappa(\mathbf{x}, \mathbf{x}_i) = (\Gamma \mathbf{x}^t \mathbf{x}_i + 1)^p$ in which p is set to 1, Γ is 1/80 and the term C for the non-separable case is 1. Table 3 shows the performance comparison over the 100 randomly selected scenes that contain red 109 traffic signs. In our simulation for [2], since its original ROI extraction method is sensitive to illumination effects in the given input image, we employ the ROI extraction method using HSI space to obtain the red ROI [8]. The verification for the segmented image of its work is done by SVM which yields 99.73% accuracy with 5-fold cross validation. That verification contains the rejection of the aspect ratio in the original work. In Figure 3 for our detection, the red rectangles which represent all prediction for each layer of pyramid image by

Table 1. The ROC analysis of classifiers for the Gabor filtered features using 5-fold cross validation (bold represents the best performance among the classifiers)

Classifier	Recall	FPR	Precision	F-measure	Accuracy
Naïve Bayes	0.806	0.187	0.327	0.465	81.27 %
Bayesian network	0.843	0.05	0.653	0.736	93.88 %
RBF-NN	0.864	0.017	0.852	0.858	97.10 %
SVM-poly	**0.913**	**0.001**	**0.993**	**0.951**	**99.05 %**

Table 2. The improved performance of classifiers using the enhancement of the power spectrum

Classifier	Recall	FPR	Precision	F-measure	Accuracy
Naïve Bayes	0.832	0.056	0.624	0.713	93.22 %
Bayesian network	0.854	0.021	0.818	0.836	96.60 %
RBF-NN	0.874	0.028	0.781	0.825	96.24 %
SVM-poly	**0.935**	**0.0**	**0.998**	**0.966**	**99.32 %**

Figure 3. The experimental results of the proposed method which covers weather, illumination effects, and occlusion
(The red rectangle expresses the candidate, the green rectangle represents the final decision)

SVM. Each candidate is marked by the square of $2\varrho+1$ pixels and clustered by 8-neighborhood recursive segmentation method. The green rectangle is drawn by the average position of the clustered red rectangles and the largest scale among those.

Table 3. The performance comparison

Evaluation	Nguwi [2]'s feature extraction with HSI ROI	the proposed method
Detection rate	0.6146	0.9174
False alarm	35	8

4. CONCLUSION

This paper introduces traffic detection based on machine learning and shows the robustness which provides the tolerance for the failure from the unexpected condition. The proposed feature extraction yields 99.32% accuracy from SVM with 5-fold cross validation and 91.74% detection rate on real world scenes. The detection includes the multiple detections and occluded traffic signs. It provides ROI-free that it is hard to find an optimized threshold or even there is no optimized value. For future works, the implementation of CUDA would be expected for real-time processing. To reduce the false decision of the predictor, the other prediction approach such as the ensemble predictor is promising for the improvement of the detection performance.

ACKNOWLEDGEMENT

REFERENCES

[1] Maldonado-Bascon, S., *et al.*, "Road-Sign Detection and Recognition Based on Support Vector Machines," Intelligent Transportation Systems, IEEE Transactions on, vol. 8, no. 2, pp. 264-278, June 2007

[2] Y.-Y. Nguwi and S.-Y. Cho, "Emergent self-organizing feature map for recognizing road sign images," Neural Computing & Applications, vol. 19, pp. 601-615, 2010.

[3] Bahlmann, C., Zhu, *et al.*, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," Intelligent Vehicles Symposium, 2005. Proceedings. IEEE pp. 255- 260, June 2005

[4] Tsai, L.-W., *et al.*, "Road sign detection using eigen colour," Computer Vision, IET , vol.2, no.3, pp.164-177, September 2008

[5] C. Garcia and M. Delakis, "Convolutional face finder: a neural architecture for fast and robust face detection," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 26, pp. 1408-1423, 2004.

[6] K. Kwang In, *et al*., "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 25, pp. 1631-1639, 2003.

[7] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex", Journal of Neurophysiology, vol. 58, pp. 1233-1258, December 1, 1987.

[8] H. Gomez-Moreno, *et al*., "Goal Evaluation of Segmentation Algorithms for Traffic Sign Recognition," Intelligent Transportation Systems, IEEE Transactions on, vol. 11, pp. 917-930, 2010.