Proceedings of the 9th International Conference on Neural Information Processing (ICONIP'02), Vol. 2
Lipo Wang, Jagath C. Rajapakse, Kunihiko Fukushima, Soo-Young Lee, and Xin Yao (Editors)

DETECTION AND RECOGNITION OF ROAD SIGNS USING SIMPLE LAYERED NEURAL NETWORKS

Hirofumi Ohara, Ikuko Nishikawa

Ritsumeikan University
Department of Computer Science
Shiga, 525-8577, Japan

Shigeto Miki, Noboru Yabuki

Tsuyama National College of Technology
Dept. Electronics and Computer Engeering
Okayama, 708-8509, Japan

ABSTRACT

A road sign detection method is proposed, using 2 simple 3-layered neural networks. Multiple preprocessing steps are taken for masking the irrelevant areas and selecting the candidate areas from an original input color images, and both neural networks are modules for this selection process. One network is for matching a color of an input pixel with a road sign, and the other is for matching a shape of an input object. The final recognition is given by a template matching, and an auxiliary re-detection step is added to improve the efficiency. The experiments using a large number of pictured images under several different conditions show the high detection rates over 95% in mose cases, while the computational cost is low owing to the smallness and the simplisity of the neural networks.

1. INTRODUCTION

The technology in the field of ITS (Intelligent Transportation Systems) makes remarkable developement in the recent years, and the image processing is one of the key technologies to supply real time information for a driver. Neural networks are also used as intelligent tools for the image processing in ITS [1, 2]. In this paper, small and simple structured neural networks are used as modules for the proposed system of the road sign detection from a natural color image. The experimental results are given to see the efficiency and the computational cost of the proposed method.

2. DETECTION PROCESS USING 2 TYPES OF LAYERED NEURAL NETWORKS

The outline of the detection process is shown in Fig.1. The original color image is first reduced the relevant area by LOG filter. Next Color NN classifies each pixel whether the pixel has a color of a road sign or not. Then a Shape NN outputs whether each sub-region in the reduced area contains an object wirh the shape of a road sign. If a shape similarity is higher than a threshold value, then a template matching is done for the final recognition. There is also an additional process of the re-detection for the region which is rejected by the previous steps. The details of each steps are described in the following sections.

step 0 Original image of 24-bit color

step 1 Masking the irrelevant region for the road sign detelction by LOG filter

step 2 Selection of the relevant region using Color NN for matching the color

step 3 Selection of the relevant region using Shape NN for matching the shape

step 4 Matching with the temple for the final recognition

step 5 Auxiliary Selection from the previouslt rejected region

Figure 1: Outline of the proposed detection method

2.1. Masking the region by LOG filter

Several procedures in the first region reduction step are shown in Fig.2. An original image is each flame of video data pictured by a digital camera, and 24-bit color image with 320×240 pixels. This is transformed into 8-bit gray scale image of brightness. Then Laplacian of Gaussian filter (LOG filter) shown in Eq.(1) is applied as the first region selection.

$$\nabla^2 G(x,y) = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) e^{\frac{x^2+y^2}{2\pi\sigma^2}} \qquad (1)$$

,where σ is a positive constant and x, y are 2 dimensional coordinates of an image. $\sigma = 0.85$ is used in the step 1-1 — Transformation of a 24-bit color image into a 8-bit brightness image

step 1-2 — Application of LOG filter

step 1-3 — Transformation into binary image using Discriminant Analysis Method

step 1-4 — Removal of noisy areas

step 1-5 — Masking the original color image by the obtained binary image

Figure 2: Masking process (step 1 in Fig.1)

present experiments, by which the Gaussian factor restricts the operator range mainly to the adjacent pixels. An output form the filter is shown to be similar to a reverse of an output obtained by the application of the Median filter for 3 or 4 times repetitively. Therefore, LOG filter is used as a simplified version to reduce the computation.

An output from the filter is quantized into a binary using a cut-off value obtained by Discriminant Analysis Method (DA method). In the following experiments, a cut-off value, which is obtained by some preliminary results, corresponds to 50% pixels for each. Neighboring pixels with the same value are clustered into one region, and small regions with less than 100 are deleted. Finally, the binary image is used as a mask region for the original color image, that is, the regions with small derivative values are taken as irrelevant for the sign detection. The sky areas are usually deleted by this process successfully, and almost one third of the original image is removed in average.

2.2. Selection of relevant pixels by Color NN

As the second step, Color NN selects each pixel in the regions reduced by the previous step. The inputs to the Color NN are 3 color attributes of one pixel; R and B values of the pixel by RBG color system and a luminosity. The colors used for road signs are 2 in most cases, one is from several primary colors, such as red, blue and yellow, and the other is black or white. An output unit of the Color NN corresponds to each color used in the road signs, and one additional unit to detect that the input pixel color does not match any color of signs. The network has an ordinary 3 layered structure with a sigmoidal function for each unit. Thus the Color NN classifies whether the input pixel has a color of any road sign or not, by winner-takes-all. In the following experiments, only road signs painted with red and white are considered, therefore the Color NN has 2 output units, one for red and the other for non-red . The network learns by Back-Propagation using a large amounts of pixels taken from sample images with and without a road sign as learning data.

After the Color NN classifies each pixel, neighboring pixels that belong to the same class are clustered into one, and small regions with less than 100 pixels are taken as a noise and deleted from the relevant region. Large regions with more than 3,000 pixels are also deleted to remove some commercial signboards in the present experiments.

2.3. Selection of relevant regions by Shape NN

The third step is matching the shape using Shape NN. The Shape NN is also 3-layered network with a sigmoidal function. Each output unit corresponds to a shape of a road sign, and detect a similar shaped object. An input to the network is a sub-area with $N \times N$ pixels in the regions selected by the previous steps. The network has $N \times N$ input units, and a gray-scale pixel value is given to a unit. The network learns by BP using binary images of road signs as training data.

If one output exceeds a certain threshold, then the input sub-area becomes a candidate for the final step of a template matching. For more than one output units over a threshold, the matching is done in order of the output value.

2.4. Template matching for the selected regions

The fourth and the final step is the template matching on the selected sub-area of size $N \times N$. The sub-area image is normalized by the following several procedures before a template matching. Small defects and noises are filled in or deleted. Inflation and deflation is also used to delete isolated points. The pixel values are normalized by the maximum and the minimum values of the input sub-area. The size is also normalized depending on the template to be matched. The closeness with a template image is given by the following neasure D,

$$D = \sum_x \sum_y \frac{|f(x,y) - g(x,y)|}{N^2} \quad (2)$$

,where $f(x,y)$ and $g(x,y)$ are pixel values of the input and the template, respectively.

2.5. Re-selection from the rejected region

This section describes an additional process for the region missed by the series of above steps. Here we focus on the miss detection at the final step by Shape NN.

The most misses by Shape NN are owing to the split of one sign region into more than one pieces. This is caused by the characters or symbols written in white, which split the background red area. The output from the Color NN is often divided into upper and lower pieces for a stop sign, and it is divided into more than two for a one-way sign, for example.

To solve this particular miss detection problem, we again use the mask information obtain in Sec.2.1, and apply the mask to the image with the reduced scales of colors. That is, an original full color image is divided into coarse-grained pieces by the color quantization. If one piece contains at least one pixel which is labeled as red by the Color NN, then the whole piece is filled in by red, and becomes an input to the Shape NN again. If the new output from the Shape NN exceeds a certain threshold, then the input region is matched with a template just as described in Sec.2.4. This is an auxiliary procedure after the whole steps described in Secs.2.1 to 2.4 are done, and is not applied to the region, where a sign is ever successfully detected.

3. STRUCTURES AND THE LEARNING OF NEURAL NETWORKS

3.1. Color NN

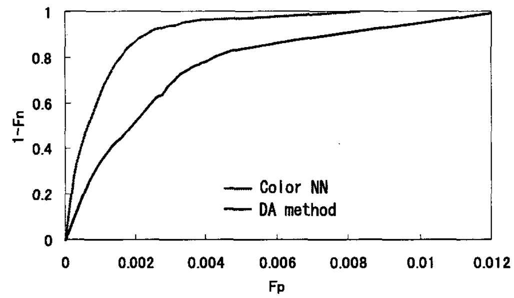
3.1.1. Structures and the learning

The network is 3-layered with 3 input units, 4 hidden units and 2 output units. 3 inputs are for R and G values, and a luminosity of a input pixel. 2 outputs are for red and non-red. The training data are 6,691 pixels taken from multiple pictures with road signs. The pixels in the red region of a sign are labeled red, while the other pixels are non-red. Among 6,691 pixels of the training data, 1,905 pixels are red on a sign. The other pixels are 1,242 in a yellow domain, 800 in a blue domain and 2,744 of others, where the domains are categorized on RG color plane. Some pixels in a yellow domain are orange close to red, and some pixels in a blue domain are purple close to red, both of which make the learning difficult. A convergence condition for the learning is the relative error less than 10%.

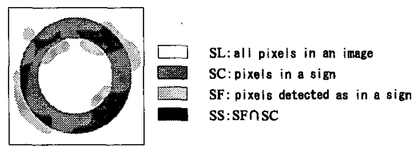
3.1.2. ROC curve of the network output

As the evaluation of the Color NN, the obtained results are compared with another network with one output unit. This unit is trained to output 1.0 for a red pixel and 0.0 for the others, and after the training an output value in [0.0, 1.0] is made binary using a cut-off value obtained by DA method.

Fig.3 (a) shows ROC (receiver operating characteristic) curves given by the Color NN with 2 output units, and given by the network with one output unit plus DA method. The horizontal axis is the rate of False



(a) ROC curves of the Color NN and DA method



(b) Definition of F_p and F_n

Figure 3: Evaluation of the Color NN by ROC curve

Alarms F_p, which comes from falsely detected pixels as on a sign, and the vertical axis is 1.0 minus the rate of Miss F_n, which comes from pixels on a sign that failed to be detected. Therefore the upper left part in the figure indicates a successful classification of low miss and low false alarm rates. The curve of the Color NN shows a better result. Here the definitions of F_p and F_n are given by the following Eqs. (3) and (4),

$$F_p = \frac{S_F - S_S}{S_L - S_C} \tag{3}$$

$$F_n = \frac{S_C - S_S}{S_C} = 1 - \frac{S_S}{S_C} \tag{4}$$

, where S_L, S_C, S_F and S_S are the numbers of all pixels in an image, of the pixels in a sign area, of the pixels detected as in a sign area and of the pixels correctly detected as in a sign area, respectively (See Fig.3 (b)).

3.2. Shape NN

The network is 3-layered with 32×32 input units, 30 hidden units and 3 output units. The number of the inputs corresponds to the size of the input sub-area. 3 outputs are for 3 different signs shown in the next chapter. BP is used for the learning, and a convergence condition is the relative error less than 10%.

4. EXPERIMENTAL RESULTS

4.1. 3 types of road signs

The road signs used for the experiments are 3 signs shown in Fig.4, a speed limit sign, no parking and a stop signs. All 3 signs are painted by red and write. The shape of a speed limit and no parking signs is a circle, while the shape of a stop sign is an inverted triangle.

Figure 4: Shapes of 3 road signs used in the experiments (left: a speed limit, middie: no parking and right: a stop sign)

4.2. Detection procedure

Fig.5 shows an example of the detection precoess obtained by the proposed method. Fig.(a) is an original 24-bit color image with 2 road signs. Fig.(b) is the output of LOG filter, which is applied to the 8-bit gray image, and Fig.(c) is a binary image given by a cut-off value obtained by DA method. The regions with a big spatial derivative are colored black, and used as non-masked regions for the original image of Fig.(a). Fig.(d) is a color image after the masking. The Color NN after the learning described in the previous section is applied to each pixel in the non-masked regions, and Fig.(e) is the result of the output from the Color NN. The pixels classified as red are shown in the figure. Fig.(f) is the output from Shape NN, where a threshold $t = 0.25$ is used in the following experiments. Both speed limit and non stop signs are successfully detected and recognized in this example.

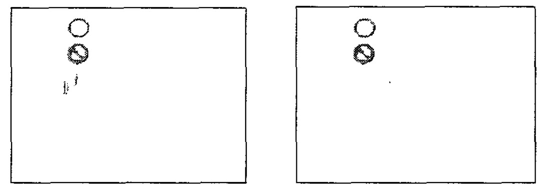
Fig.6 is an example of the failure to detect a stop sign. Fig.6(a) is an original color image and Fig.(b) is the output from the Color NN to detect the sign region, which is split into two. Fig.(c) is the output from Shape NN, which fails to detect anything. Additional re-detection procedure described in Sec.2.5 is applied in this case. Fig.7(a) is the masked image of 3-bit color. If a piece contains any red pixels shown in Fig.6(b), then it is completely filled in by red. The resulting output from the Shape NN is shown in Fig.7(b), which successfully detects the stop sign.

4.3. Data and the results of the experiments

Image data are the flames of video pictures of road signs taken at 4 places from a moving car, at 2 different times

(a) Original color image (b) Output from LOG filter

(c) Binary image by DA (d) Masked color image



(e) Output from Color NN (f) Detection of the signs

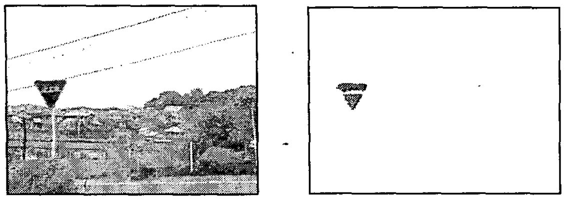
Figure 5: An example of a successful detection process

and dates for each. Table 1 shows the place, dates and signs of the data.

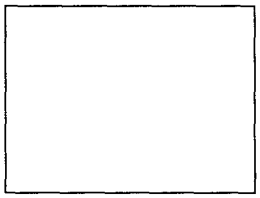
The results of the detection are shown in Table 2. A hit rate 1 H_1 is the result without the re-detection, and a hit rate 2 H_2 is the result of the re-detection, which is only applied to the missed cases of the rate $1 - H_1$. Total hit rate is therefore obtained by $H_1 + (1 - H_1) \times H_2$. In most data, the total hit rate is over 95%, while for Data No.9 it remains 87%.

Fig.8 is a typical example of a miss even after the re-detection. Fig.(a) is a part of the original color image with a stop sign. Fig.(b) is the output from the Color NN, where the sign region is split into two. Fig.(c) is the masked image with reduced numbers of color scales, and Fig.(d) is the colored region by the filling-in procedure. By this re-detection process, the split region shown in Fig.(b) is connected, but there also appear some additional string-like noises around the sign, which causes a miss detection again. Inflation-deflation procedure is not able to solve this problem in our experiments.

The results in this chapter shows the effectiveness of the proposed method using the neural networks and the re-detection. However, the re-detection is not always effective, and it is necessary to consider more careful design of the parameters and the adaptation to each data for the further improvement.

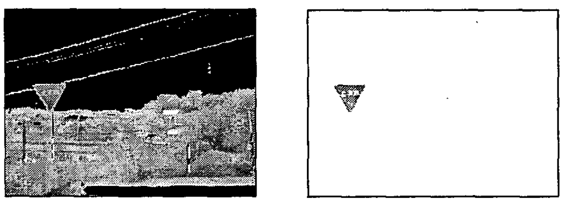


(a) Original color image (b) Output from Color NN



(c) Miss detection of the sign

Figure 6: An example of a miss detection process



(a) Masked and quantized image (b) Re-detection

Figure 7: An example of a successful re-detection

5. CONCLUSION

A road sign detection method is proposed and evaluated by the experiments. 2 neural networks are used for the pre-processing of the selection and masking in the proposed method; one for the color matching and the other for the shape matching. Both are simple 3-layered networks with BP algorithm, therefore the computational cost is rather low, while the experiments show the high recognition rate over 95% in most data.

6. REFERENCES

[1] N. Yabuki and S. Miki, Recognition and Detection of Speed Limit Signs from Road Images, Trans. of IEICE, J77-D-II, pp. 1393-1394, 1997.

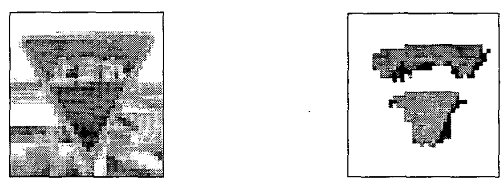
[2] H. Kodama et.al., Location Identification of a Speed Limit Sign using a Layered Neural Network, RTA-97-44, pp. 55-58, 1997.

Table 1: Pictured image data used in the experiments

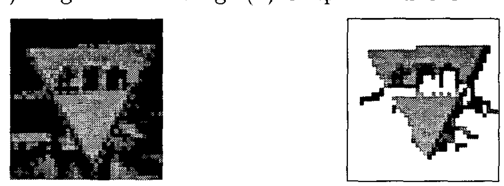
Data	Place	Date and time	Signs in the image
No.1	A	5/12/'01 9:00	a speed limit and no parking signs
No.2	B		
No.3	A	5/12/'01 14:00	
No.4	B		
No.5	A	5/15/'01 9:00	
No.6	B		
No.7	C	5/15/'01 9:00	a stop sign
No.8	D		
No.9	C	5/15/'01 12:00	
No.10	D		
No.11	C	5/16/'01 13:00	
No.12	D		

Table 2: Recognition Rates

Data	Hit rate H_1	Hit rate H_2	Total hit rate
No.1	100%	–	100%
No.2	100%	–	100%
No.3	96%	0%	96%
No.4	100%	–	100%
No.5	100%	–	100%
No.6	100%	–	100%
No.7	100%	–	100%
No.8	93%	100%	100%
No.9	65%	64%	87%
No.10	30%	93%	95%
No.11	100%	–	100%
No.12	100%	–	100%



(a) Original color image (b) Output from Color NN



(c) Masked and quantized image (d) Filled in area

Figure 8: An example of a re-detection failure