Real-time Traffic Sign Recognition System with Deep Convolutional Neural Network

Seokwoo Jung[1], Unghui Lee[2], Jiwon Jung[3] and David Hyunchul Shim[4]

[1] Department of Aerospace Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea
(Tel : +82-10-2034-5941; E-mail: swjung92@kaist.ac.kr)

[2-4] Department of Aerospace Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea
(Tel : +82-42-350-3764; E-mail: hcshim@kaist.ac.kr)

Abstract – TSR (Traffic Sign Recognition), a part of ADAS (Advanced Drive Assistance System), helps driver (or car) to recognize traffic signs ahead with using front camera. According to EURO NCAP rating policy, car should be able to warn the driver when the car's speed is above the set speed threshold. It is thought that various types of traffic sign should be recognized to get more detailed information of road. In this paper, 6 types of traffic sign images are trained by LeNet-5 convolutional neural network architecture. In the detection phase, light-weight color-based segmentation algorithm and Hough transform algorithm are applied to extract candidate regions of traffic signs. The recognition system nearly achieves real-time performance. On-line recognition test is performed on the KAIST campus road, and the result shows all 16 traffic signs are recognized successfully through the driving. The recognition system is implanted into autonomous vehicle 'Eurecar'. Different types of traffic signs are trained consistently and development of clustering algorithm is considered as a future work for robust recognition system.

Keywords – TSR (Traffic Sign Recognition), CNN (Convolutional Neural Network), Real-time Image Processing, Autonomous Vehicle

1. Introduction

Recently, well-known automobile companies like BMW, Mercedes-Benz, and etc., are enthusiastically developing ADAS (Advanced Drive Assistance System). Mobileeye, technology company that develops vision-based ADAS, commercialize ADAS system equipped not only LKAS (Lane Keeping Assist System), but also TSR (Traffic Sign Recognition) system for speed limit signs. And according to EURO NCAP rating policy, car should be able to warn the driver when the car's speed is above the set speed threshold. Speed limit value could be also known by digital map data, but this method is restricted to roads that are parts of the map data. Therefore, vision-based traffic sign recognition system is needed for other cases.

Also, it is thought that various types of traffic sign should be recognized to get more detailed information of road. However, it is hard to make a specific country-version traffic sign recognition system (except speed limit signs) because only US and German traffic sign datasets are available freely.

In this paper, state of the art deep learning technology is considered as a main algorithm to recognize traffic signs. Especially in the deep learning architecture, convolutional neural network (hereinafter referred to as CNN) has been widely used for image classification problem. CNN is a kind of multilayer perceptron that uses small sub-region (called a receptive field). Receptive fields are tiled to cover the entire input image, then produce feature maps by share the same weight and bias. So, it could reduce the number of parameters needed for neural network learning, and increases learning efficiency.

In this paper, a firsthand data is used for training Korean-version traffic signs. The real-time application is final goal, so appropriate light-weight detection algorithm is chosen for extracting ROI (region of interest) of traffic signs.

In chapter 2, methodology of the recognition algorithm is described briefly. In chapter 3, on-line experiment environment is depicted and the test result is analyzed. In the last chapter, improvement method is discussed.

2. CNN-based traffic sign recognition algorithm

2.1 Extraction of traffic sign candidates

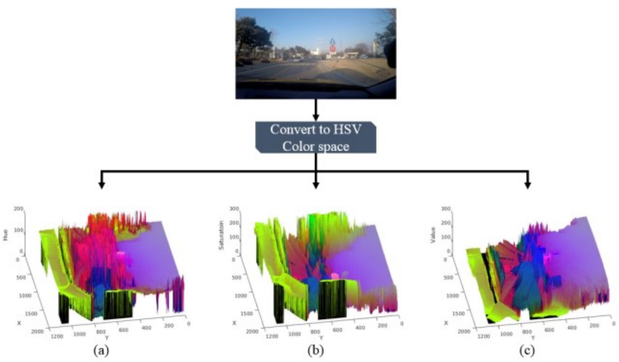

Fig. 1. Distribution of (a) Hue (b) Saturation (c) Value.

Including LeNet-5 model used in this paper, almost of CNN models that are constructed for image classification are set to use crop images as input data. The pixels of crop images are composed mostly of object regions, because It is more advantageous to extract features of specific objects. That's why the region proposal algorithm is needed before the classification phase. Selective search [1] is recently used as a preprocess of object detection. It searches regions likely to contain objects by computing similarity of various color spaces and texture. It is designed to trade off between the robustness and the speed. To reach satisfactory detection performance, almost of the cases are too slow for real time application. So, we suggest simple color segmentation method for region proposal, the HSV (Hue, Saturation and Value) thresholding operation. Hue value has a 0 to 180 range, and the others have 0 to 255 range.

Traffic signs have distinct color features. For example, pedestrian crossing signs are blue, and stop signs are red. Because of this reason, the simple HSV thresholding could extract regions for some kinds of traffic signs. Figure 1. shows the hue, saturation and value distribution of an image which contains "Stop" and "Pedestrian crossing" signs. The hue of stop sign is high (larger than 160) and the hue of pedestrian crossing sign is around 100. Saturation and value also depend on objects. We selected minimum and maximum H, S and V numbers of 5 thresholding operations like Table 1. These are arranged with extraction algorithm of stop, sharp curve, slow, pedestrian crossing and speed bump signs. Minimum and maximum numbers of HSV are obtained by following equations.

$$
\begin{aligned}
H_{\min} &= \mu_H - \alpha\sigma_H,\ H_{\max} = \mu_H + \alpha\sigma_H \\
S_{\min} &= \mu_S - \beta\sigma_S,\ S_{\max} = \mu_S + \beta\sigma_S \qquad (1) \\
V_{\min} &= \mu_V - \gamma\sigma_V,\ V_{\max} = \mu_V + \gamma\sigma_V
\end{aligned}
$$

In eq. (1), μ and σ are mean and standard deviation of non-zero pixels. α, β and γ are decisive factors of the margin of H, S and V range. Appropriate sizes of α is 1.1. On the other hand, β and γ are set to 2.5. Sufficient margin of saturation and value are reflected in max. and min. selection., because both of them are greatly changed according to illumination condition. Especially in stop signs, min. and max. numbers of hue are set to heuristically because we use HSV conic color space model. The hue of red color ranges in 170~180 and 0~10.

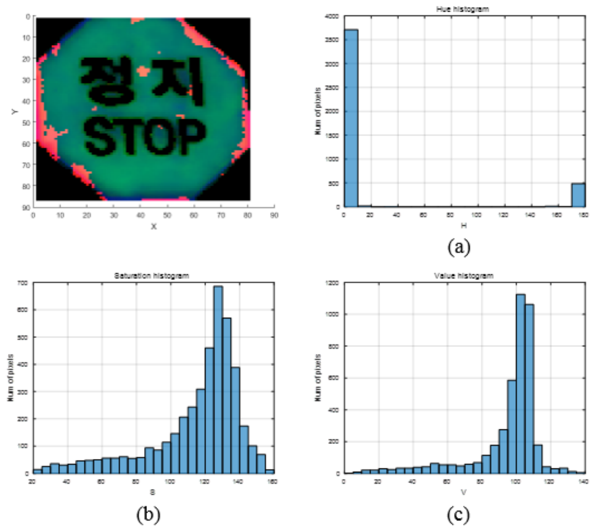

Fig. 2. Histogram of (a) Hue (b) Saturation (c) Value of the stop sign.

Table 1 Minimum and maximum number of HSV.

Operation Number	Min. threshold number			Max. threshold number		
	H	S	V	H	S	V
1	0	47	43	20	181	145
2	11	73	0	31	252	206
3	0	16	17	40	149	143
4	106	39	16	111	169	155
5	13	93	0	24	226	204

The shape of speed limit sign is circular, and this feature could be distinct from the shape of other objects. Therefore, Hough transform [2] is used to detect circular shape in the input image. Before Hough transformation algorithm, Gaussian filter (5x5 kernel with standard deviation 1.2) and Canny edge detection algorithm [3] are implemented. Figure 3. shows the result of region proposals of a front camera image. Regions whose size are too large or small or aspect ratio are large (long rectangle shape) have been excluded in detection phase.

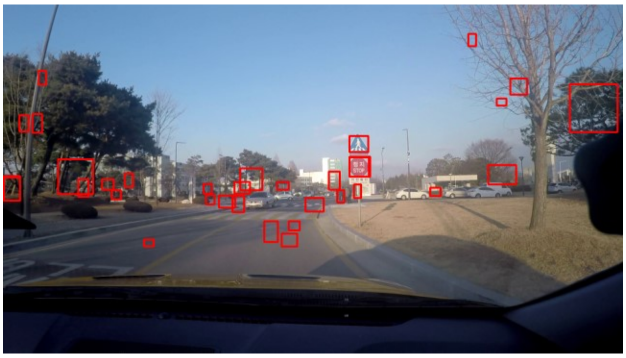

Fig. 3. A reuslt of region proposals algorithm.

2.2 Classification with LeNet-5 CNN architecture

Each proposed region is entered to CNN model. The images are resized by 100x100 pixel, and also have RGB channels. In other words, the input data of CNN has 100x100x3 dimension. Output layer is designed that has 8 nodes, 6 is the number of the classes of chosen traffic signs, and another one is the class of the other traffic signs. And the last 1 node shows the false positive class output. Figure 4. shows the image of chosen traffic signs. It contains some Korean words.

Fig. 4. List of selected traffic signs.

The deep convolutional neural networks is constructed based on LeNet-5 model [4]. Some numbers of nodes have been modified. In 2 convolutional layers, 5x5 size receptive field is used and it tiles the input image with stride 1. After each convolutional layer, max pooling layer resamples the data and reduces size of the data by using 2x2 kernel with stride 2 arrangement.

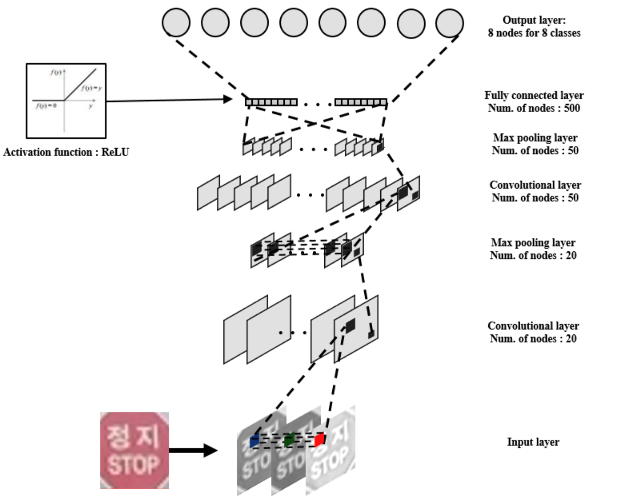

Fig. 5. LeNet-5 architecture for traffic sign classification.

We collected the driving video by using GOPRO HEREO 4 camera attached on the wind shield of the autonomous car whose name is 'Eurecar' [5]. At first, we trained only with positive samples created by simple affine transformation from hand cropped images. we have collected false positive classification results after several test. And, we set them to be negative samples. Finally, we mixed the training data with 25000 positive samples and 78000 negative samples (Fig.6 shows an example of the samples).

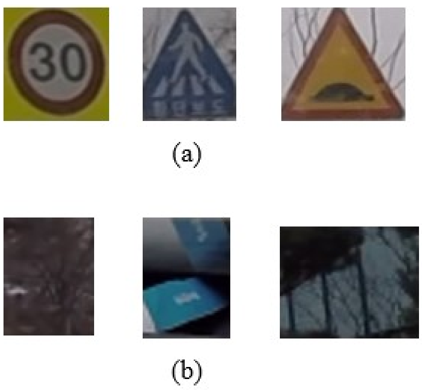

Fig. 6. (a) Positive samples (b) Negative samples.

In this paper, TSR system is developed by C++, CUDA 7.5 environment. We used the Caffe [6] deep learning framework that include modularity and efficient computer operations. Specifications of the computer used in training is i7-6820HK CPU, Titan X graphic card with 12gb GPU memory.

3. Test result

3.1 On-line field test

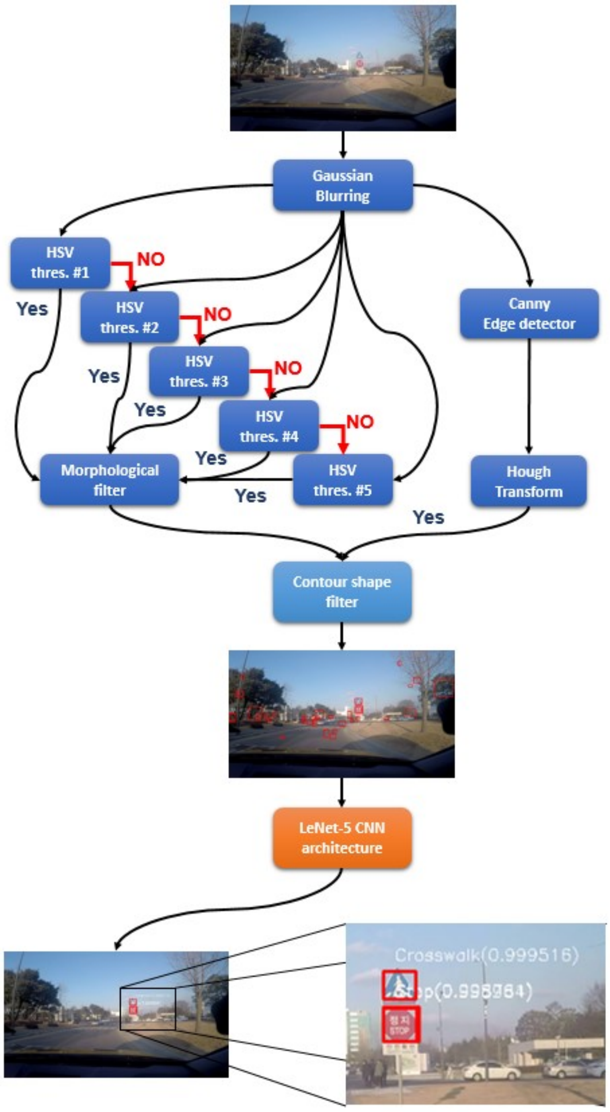

Fig. 7. TSR system flow chart and recognition result.

The test is performed on the KAIST campus road. It is easy to see many kinds of traffic signs in the campus for security reasons, so it is suitable to test TSR system. During prior field test, the output node threshold value has been tuned as 0.95. In other words, if the class A node from a input cropped image has output value 0.97, the TSR system decides the image is in class A.

There are 16 traffic signs on the road within the observable range. All of the traffic signs were recognized successfully during driving the course. Figure 7. shows the overall process from an input frame to an output classification result. The input image is obtained from a field test video. As shown in bottom image of fig. 7, pedestrian crossing and stop traffic signs are classified with well-fitted bounding box. Average fps (frames per second) is about 16.9Hz on high-end notebook which is connected with exterior Titan X.

Fig. 8. An overlapped recognition result.

However, there remains the further problems to solve. In this paper, each region proposal treated individually. It means that it shows overlapping results when several region proposals point to a same traffic sign. Figure 8. shows a speed limit traffic signs are recognized twice. During detection phase, Hough transform finds the circular shape of the sign and HSV thresholding operation finds the red color. Therefore, the speed limit signs detected twice by independent method. However, the location of proposed region is slightly different, so the TSR system recognizes 2 traffic signs, not one. It is thought that this problem could be solved by developing clustering algorithm of detection (or recognition) results.

4. Conclusion

Most studies in the field of vision based ADAS have been developed by using static and light-weight algorithms. It is fast, but the accuracy depends on the environments like illumination condition. Recently, parallel programming has been widely used because of advances in GPU (Graphical Processing Units). So, we tried to apply a complex learning model to real-time application. As a result, it works satisfactorily in terms of accuracy and speed. Above all things, our model could learn the data consistently, and the accuracy will increase whenever it trains with new data. Some remaining problems are accounted for, and the solutions might be developed soon.

Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. 2010-0028680).

References

[1] J. R. R. Uijlings et al., “Selective Search for Object Recognition”, IJVC, 2013.

[2] D.H. Ballard, “Generalizing the Hough Transform to Detect Arbitrary Shapes”, Pattern Recognition, Vol. 13, No. 2, p.111-112, 1981.

[3] Canny, J., “A Computational Approach To Edge Detection”, IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679-698, 1986.

[4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition.”, Proceedings of the IEEE, 1998

[5] Unghui Lee, Sangyol Yoon, David Hyunchul Shim, Pascal Vasseur, Cedric Demonceaux, "Local path planning in a complex environment for self-driving car", IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems, 2014.

[6] Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor, “Caffe: Convolutional Architecture for Fast Feature Embedding”, arXiv preprint arXiv:1408, 5093, 2014