Vehicle Color Recognition via Representative Color Region Extraction and Convolutional Neural Network

Kwang-Ju Kim [*1], Pyong-Kun Kim [*2], Kil-Taek Lim [*3], Yun-Su Chung [*4], Yoon-Jeong Song [*5], Soo In Lee[*6],

Doo-Hyun Choi [#1]

ETRI
Daegu, Korea
[1] kwangju@etri.re.kr, [2] iros@etri.re.kr, [3] ktl@etri.re.kr, [4] yoonsu@etri.re.kr,
[5] yjsong@etri.re.kr, [6] silee@etri.re.kr
#*School of Electronics Engineering, Kyungpook National University*
Daegu, Korea
[1] dhc@ee.knu.ac.kr

Abstract—**Vehicle color recognition is one of the important part in ITS (Intelligent Transportation System). This paper presents a new vehicle color classification technique for CCTV systems via representative color region extraction and Convolutional Neural Net (CNN). The Harris corner point detection method is used to generate a probability map of a representative color region. From the probability map, point are randomly selected to generate an input image for CNN. Finally, we trained CNN model with it. In order to evaluate the performance of the proposed method, we acquired a total of 5,941 images from camera on highway. We conducted 5-fold cross validation for performance evaluation. Our vehicle color recognition method performance of about 96.1% was shown.**

Keywords—Vehicle; Color Recognition; Probability Map; CNN;

I. INTRODUCTION

The Intelligent Transportation System (ITS) is a next-generation transportation system that combines intelligent technologies with existing transportation systems. Recently as the increase of safety requirement, the installation of CCTV cameras to monitor the roads is increasing rapidly. To operate these CCTVs, many monitoring agents are needed. However, there are difficulties in monitoring all of these CCTVs. Therefore, recent CCTV system is evolving into an intelligent system that automatically detects objects and recognizes properties (e.g., car color, car type, car brand, etc.). One of the representative cases of intelligent CCTV system is the automatic license plate recognition for the detection of speed violation [1, 2]. However, the case of crimes is different from that of speed violation. Most of the witness of the crimes explain their witness using the vehicle properties such as color and type at first [3, 4]. It could be very helpful to grasp the route of the criminal using the vehicle if CCTVs can automatically recognize the vehicle's color and type. Conventional vehicle color recognition methods can be classified into two types. One is a method of extracting color

from the entire vehicle, and the other is a method of using a large part of the region of the frontal vehicle [5-7]. In this paper, we propose a new vehicle color classification technique for CCTV via representative color region extraction and Convolutional Neural Net (CNN). Fig. 1 shows the flowchart of the proposed vehicle color recognition method. (1) The vehicle front image is extracted based on license plate detection [8]. (2) Harris corner point detection method [9] is applied to generate a probability map of a representative color region. (3) From the probability map, points were randomly selected to generate an input image for CNN. (4) Finally, we trained CNN model with it. In order to evaluate the performance of our proposed method, we acquired a total of 5,941 images from camera on highway.
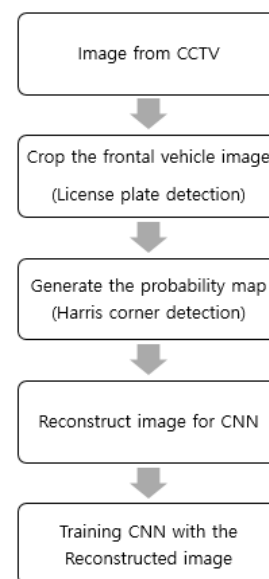

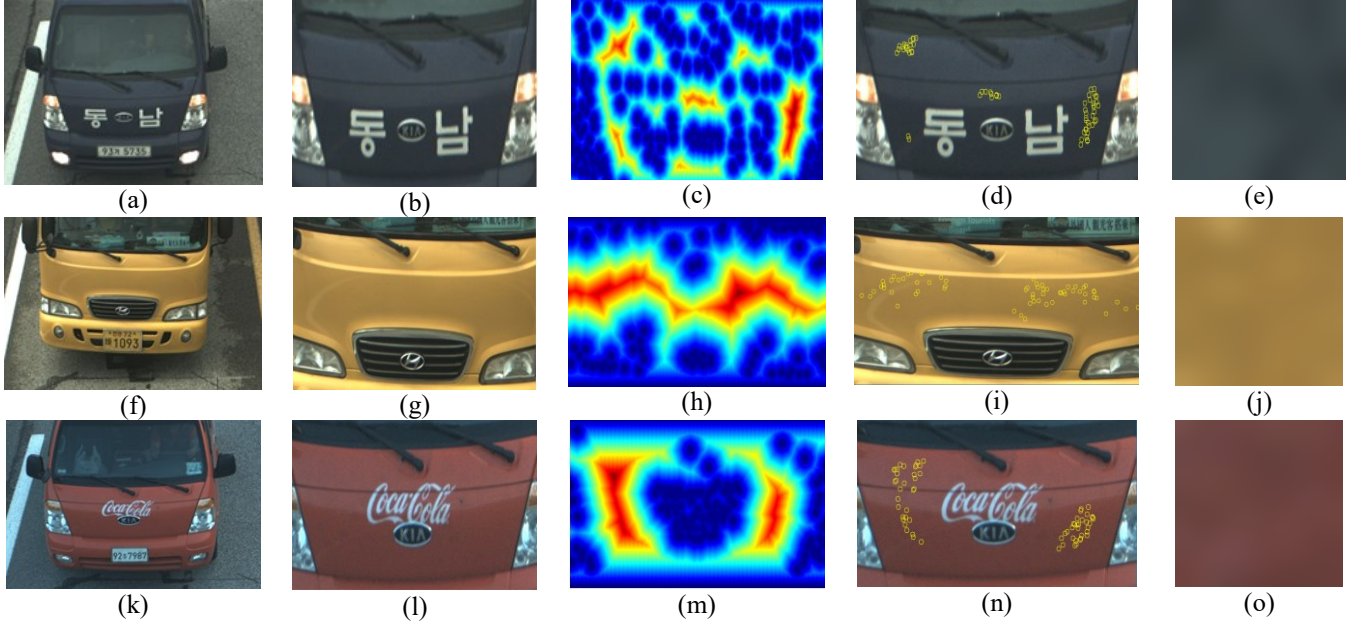Fig. 1. Flow chart of vehicle color recognition

Fig. 3. Input images (a)(f)(k), Extracted vehicle front images (b)(g)(l), Probability map of vehicle representative color region (c)(h)(m), Randomly selected points in reliable color regions (d)(i)(n), Reconstructed input images for CNN (e)(j)(o)

II. PROPOSED VEHICLE COLOR RECOGNITION METHOD

A. Vehicle front image extraction

The region containing vehicle representative color is generally located at the upper of the license plate area. After detecting the license plate area of the vehicle, the vehicle front image is extracted as shown in Fig. 2 using the license plate position information. Using the vehicle image without extracting the front image has negative effects on the vehicle color recognition since noises are included such as road and vehicle glass part.

Fig. 2. Extraction of vehicle front image

Fig. 4. Examples of noisy bonnet images

For a conventional vehicle color recognition method shown in [5-7], the bonnet area above the license plate position is set as the ROI and is recognized by histogram analysis method. The bonnet area image includes the representative color, but is also includes noise such as advertisement or emblem as shown in Fig. 4 below.
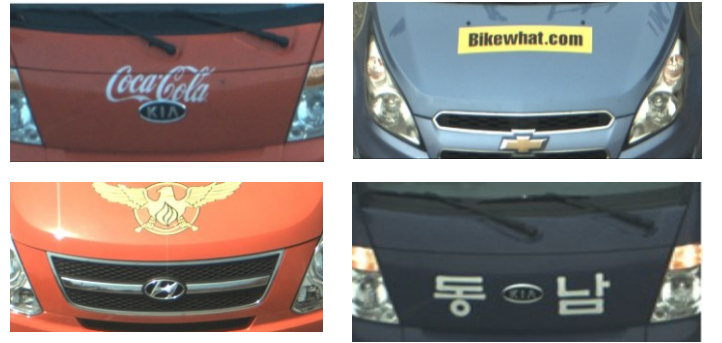
Therefore, in some cases, it is hard to recognize the color by simply using the bonnet image as in the conventional vehicle color recognition method. Hence, using parts that do not include an advertisement or an emblem in the bonnet image could enhance the color recognition accuracy. The process and result of generating the vehicle representative color from the input image are shown in Fig. 3.
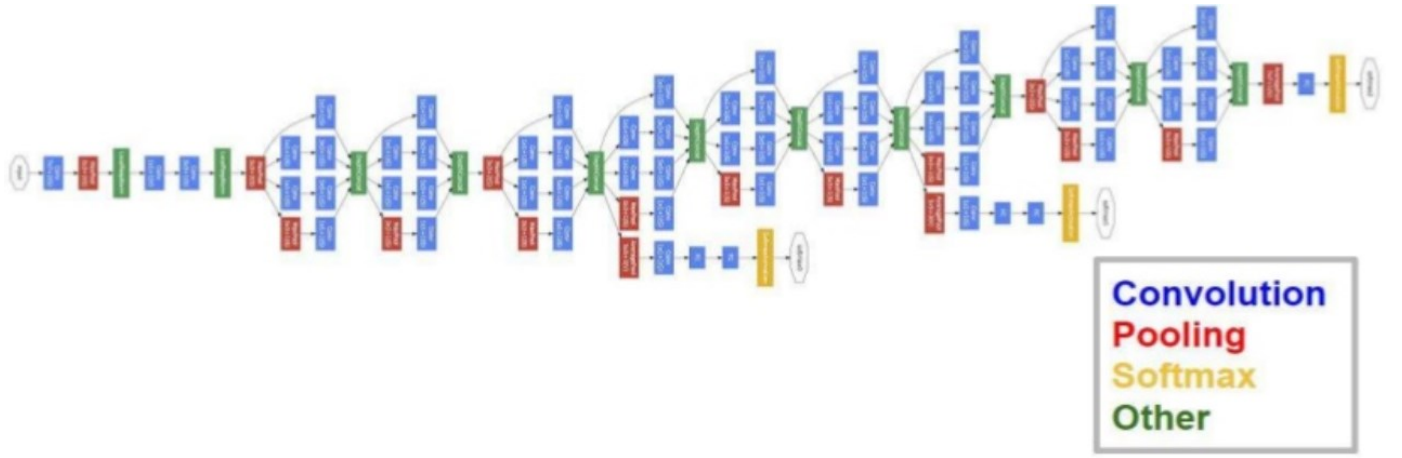
Fig. 5. Google net architecture

B. Harris corner detection

To extract the region that does not contain noise, the method of dividing the closed region formed was applied by the corner points. One of popular methods of finding corner points is the Harris corner detector. The Harris corner detector computes the locally averaged moment matrix computed from the image gradients, and then combines the eigenvalues of the moment matrix to compute a corner strength, of which maximum values indicate the corner positions.

$$c(x,y) - \sum_{w_g}[I(x_i,y_i) - I(x_i + \Delta x, y_i + \Delta y)]^2 \tag{1}$$

Where I is corner positions, (x_i, y_i) is the point inside the Gaussian window W_g. The Taylor expansion of the region shifted by $(\Delta x, \Delta y)$ is expressed as the following expression.

$$I(x_i + \Delta x, y_i + \Delta y)$$
$$\approx I(x_i,y_i) + [I_x(x_i,y_i)i_y(x_i,y_i)]\begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix} \tag{2}$$

Where I_x, I_y represents a gradient for x and y, respectively. It can be summarized the above (1), (2) as follows.

$$c(x,y) = [\Delta x \Delta y]C(x,y)\begin{bmatrix}\Delta x \\ \Delta y\end{bmatrix}$$
$$C(x,y) = \begin{bmatrix} \sum_{w_g}(i_x(x_i,y_i))^2 & \sum_{w_g}i_x(x_i,y_i)I_y(x_i,y_i) \\ \sum_{w_g}i_x(x_i,y_i)I_y(x_i,y_i) & \sum_{w_g}(i_y(x_i,y_i))^2 \end{bmatrix} \tag{3}$$

Corner Response Function (CRF) is defined to determine the corner degree using equation (3).

$$C_r(x,y) = \det(C(x,y)) - K \cdot [Trace(C(x,y))]^2$$
$$\det(M) = \lambda_1\lambda_2 \tag{4}$$
$$Trace(M) = \lambda_1 + \lambda_2$$

Where C(x, y) is the weight function. λ_1, λ_2 are the eigenvalue of C(x, y).

C. Generation of probability map

A probability map using the Harris corner detector is generated to candidate the representative color region.

$$E_{(x,y)} = \sum_{u,v}|C_{(u,v)} - I_{(x,y)}|$$
$$= \sum_{u,v}\sqrt{(C_u - I_x)^2 + (C_v - I_y)^2} \tag{5}$$

$$P_{(x,y)} = \frac{E_{(x,y)}}{E_{max}}, E_{max} = \max_{x,y} E(x,y) \tag{6}$$

Where $C_{(u,v)}$ is corner point. $I_{(x,y)}$ is pixel in the image. $E_{(x,y)}$ is a Euclidean distance from $C_{(u,v)}$ to $I_{(x,y)}$. A representative color region probability map is generated by calculating the Euclidean distance of all pixels and corner points as shown (5), (6). For example, if the distance between the corner points is close, it indicates that the probability of existing an advertisement or an emblem is high as shown Fig. 3 [10-12]. This probability map can choose the representative color region.

D. Reconstructed color image

After detecting the vehicle representative color region probability map, a new image to be used as a CNN input is

generated by randomly selecting pixels as shown in Fig. 6. Gaussian blur was applied to minimize the defects caused by noise due to pixel rearrangement.
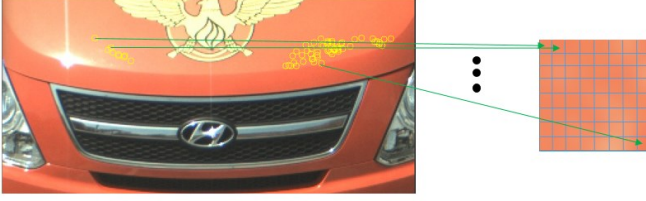

Fig. 6. Image reconstruction for CNN

E. Train CNN

Our models trained using stochastic gradient descent with 64 batch size, momentum of 0.9 and weight decay of $1e^{-05}$. Total 5,941 vehicle images from CCTVs on the highways are acquired for the experiments. They include 7 classes of vehicle color, which are black, gray, silver, white, blue, red, and yellow. Before the data processed for training, it resized to 224x224@3 and subtracted by mean image of the training data. Standard data augmentation technique was also used. Learning rate was 0.001 and reduced continuously by a factor of 10 at epoch 10 as shown in Fig. 7. DIGIT system and Caffe framework for google net was used [13]. DIGIT system provides Caffe framework with google net. The training process done in GPU hardware. We used 4 GPUs to reduce the training time.
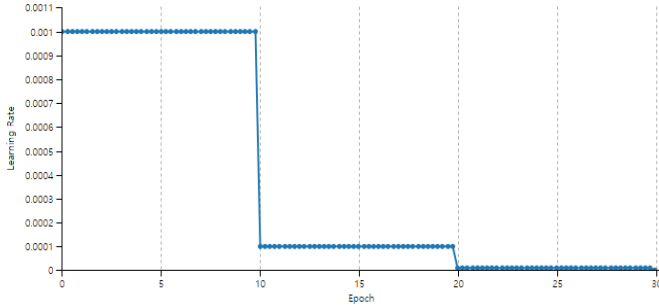

Fig. 7. Changes in learning rate

III. EXPERIMENTS

Images (size: 1624x1224) containing the vehicle were acquired from CCTV on the highway. Some example images of the dataset can be viewed in Fig. 8.

Fig. 8. Example images of dataset

F. 5-Folds Cross Validation

Our data was spilt into 5 different subsets. 4 subsets were used to train out data and leave the last subset as test data. Then we averaged result against each of the subsets as shown in Fig. 9.
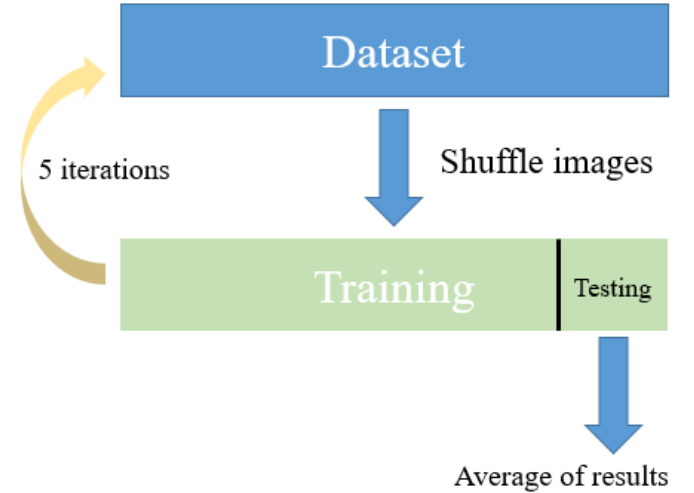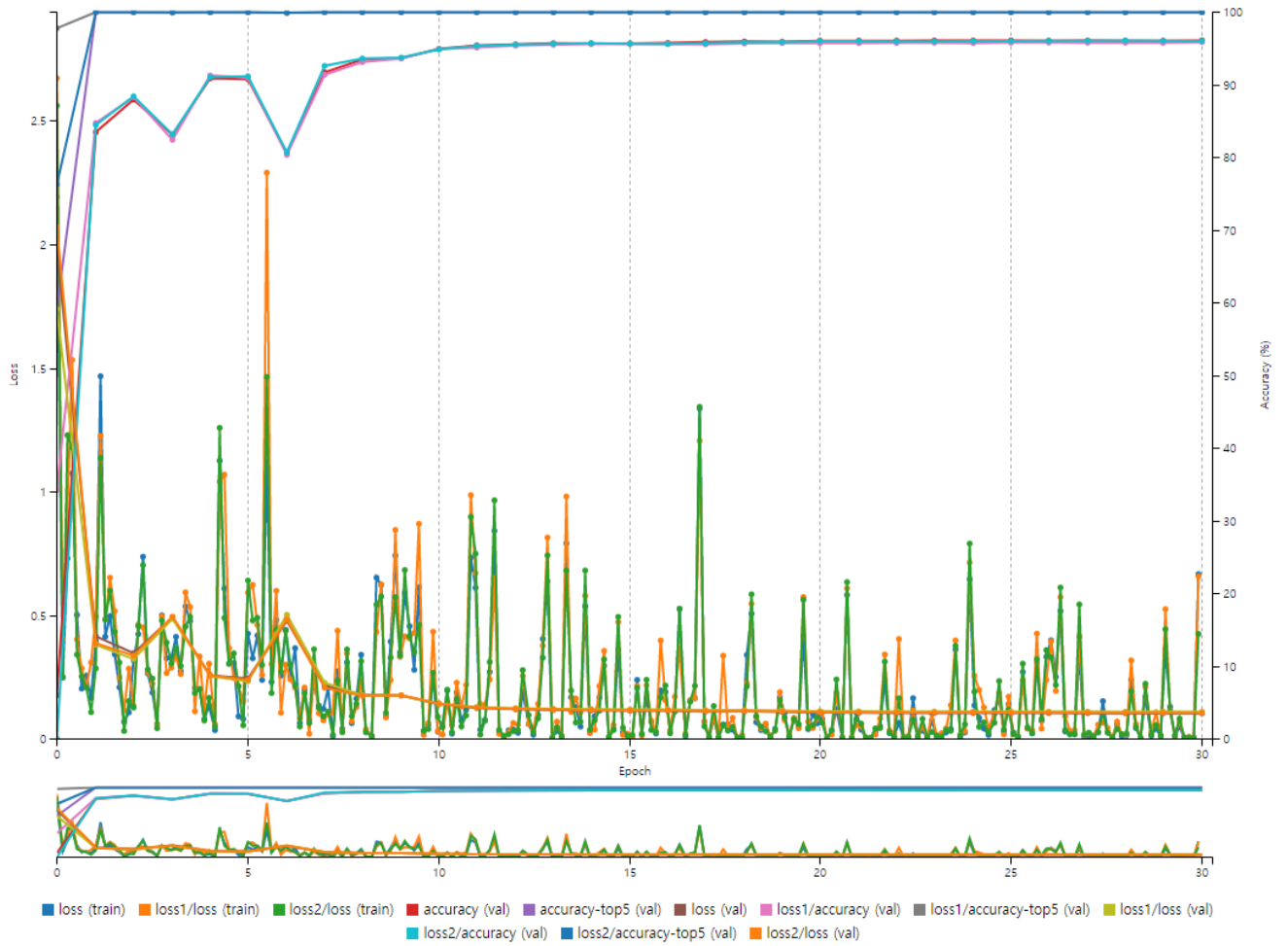

Fig. 9. 5-folds cross validation

Fig. 11. Graph of vehicle color recognition result

Legend: loss (train), loss1/loss (train), loss2/loss (train), accuracy (val), accuracy-top5 (val), loss (val), loss1/accuracy (val), loss1/accuracy-top5 (val), loss1/loss (val), loss2/accuracy (val), loss2/accuracy-top5 (val), loss2/loss (val)

G. Results

Table 1 shows the results of experimenting 5,941 images with 5-fold cross validation. The proposed vehicle color recognition method performance of about 96.1% was shown. Hyper parameter sweep of learning rate was performed for improved model accuracy.

TABLE I
RESULTS OF ACCURACY AND LOSS IN VEHICLE COLOR RECOGNITION

Epoch	Iteration	Learning rate	Accuracy	loss
5	3250	0.001	0.907766	0.244969
10	6500	0.001	0.949480	0.141192
15	9750	0.0001	0.956986	0.114438
20	13000	0.0001	0.960739	0.107537
25	16250	0.00001	0.961028	0.104845
30	19500	0.00001	0.961172	0.103609

Fig 11 showed the graphic representation of Table 1. As the epoch increases, the accuracy increase and the loss decreases. Figure 11 shows an example of CNN correctly classifying our

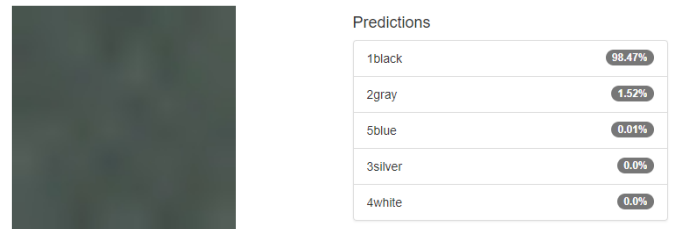image of a black vehicle with 98.47 confidence as shown in Fig. 10.

Predictions

1black	98.47%
2gray	1.52%
5blue	0.01%
3silver	0.0%
4white	0.0%

Fig. 10. An example of CNN classification result

IV. CONCLUSION

This paper presents a novel vehicle color classification technique for CCTV via representative color region extraction and CNN. In order to evaluate the performance of the proposed method, we acquired a total of 5,941 images from camera on highway. Using vehicle front image is extracted based on license plate detection, Harris corner point detection method is applied to generate a probability map. From this probability map, points were randomly selected to generate an input image for CNN. Then we trained CNN model with it

and analysed the accuracy, loss by 5-fold cross validation. Our vehicle color recognition method performance of about 96.1% was shown.

We plan to acquire more data in the future, especially in the rainy weather. We are also going to research algorithms that increase vehicle color recognition accuracy rate in various weather environments.

REFERENCES

[1] Y. Kim, "A vehicle speed measurement system implementation using a stereo camera and a license plate recognition algorithm," *Journal of the Institute of Electronics and Information Engineers*, vol. 7, pp. 78-84, 2006.

[2] B. Oh, "Number region extraction of license plates using colors and arrangement of numbers," *Journal of Korea Multimedia Society*, vol. 14, pp. 1117-1124, 2011.

[3] A. Psyllos, "Vehicle model recognition from frontal view image measurements," *Computer Standards & Interfaces*, vol. 33, pp. 142-151, 2011.

[4] A. Psyllos, "Vehicle logo recognition using a sift-based enhanced matching scheme," *Intelligent Transportation Systems IEEE Transactions*, vol. 11, pp. 322-328, 2010.

[5] T. Kim, "The study of vehicle color recognition using the fuzzy c-means algorithm based quantization," *Korea Multimedia Society*, vol. 13, pp. 235-238, 2010.

[6] T. Kim, "The study of representative color extraction of vehicle and vehicle color recognition using ART2 algorithm," *The Institute of Electronics Engineers of Korea*, vol. 33, pp. 2134-2137, 2010.

[7] M. Yang, "Vehicle color recognition using monocular camera," *Wireless Communications and Signal Processing, International Conference on. IEEE*, pp. 1-5, 2011.

[8] B. Han, "Real-time license plate detection in high-resolution videos using fastest available cascade classifier and core patterns," *ETRI Journal*, vol. 37, pp. 251-261, 2015.

[9] C. Harris, "A combined corner and edge detector," *Proceedings of the 4thAlveyVisionConference,* pp. 147-152, 1988.

[10] P.E. Danielsson, "Euclidean distance mapping," *Comput. Graphics Image,* vol. 14, pp. 227-248, 1980.

[11] G. Borgefors, "Distance transformations in arbitrary dimensions," *Comput. Vision Graphics Image Process*, vol. 27, pp. 321-345, 1984.

[12] K. Kim, "Extraction of vehicle representative color regions using feature points and probability map," *Journal of Korean Institute of Communications and Information Sciences*, vol. 43, pp. 597-602, 2018.

[13] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 1–9, 2015.