

# Multi-channel Object Detection for Detecting Suspected Trees with Pine Wilt Disease Using Multispectral Drone Imagery

Hae Gwang Park, Jong Pil Yun, Min Young Kim and Seung Hyun Jeong

**Abstract**—In this paper, a multi-channel convolutional neural network (CNN) based object detection was used to detect suspected trees of pine wilt disease after acquiring aerial photographs through a rotorcraft drone equipped with a multi-spectral camera. The acquired multi-spectral aerial photographs consist of RGB, Green, Red, NIR and Red Edge spectral bands per shooting point. The aerial photographs for each band performed image calibration to correct radiation distortion, image alignment to correct the distance error of the lenses of a multi-spectral camera, and image enhancement to edge enhancement to highlight the features of objects in the image. After that, a large amount of data obtained through data augmentation was put into multi-channel CNN-based object detection for training and test. As a result of verifying the detection performance of the trained model, excellent detection results were obtained with mAP 86.63% and average IoU 71.47%.

**Keywords**—Pine wilt disease (PWD), Drone, Remote sensing, Deep learning, Multi-spectral, Convolutional neural network

## I. INTRODUCTION

Pine Wilt Disease (PWD) is caused by close interaction between three factors that are tree, mediated insect and pathogens. It is a fatal disease that cannot be cured and recovered [1]. Pine wilt nematodes, a pathogen that provides a direct cause of PWD, are nematodes that resemble threads of around 1 mm in size [2]. It cannot move on its own, and invades into the body of larvae that normally live in PWD-infected trees. Over time, when the larvae grow into adults and move to a healthy tree and eat the bark, the pine wilt nematodes penetrate into the tree trunks of healthy trees and block the ducts, which are the nutrient passages, to block the nutrient supply [3]. As a result, the trees infected with PWD stop the resin secretion by the rapidly growing pine wilt nematodes, release volatile substances such as alcohol and terpene, thus, brown and die rapidly in a short time as shown in Figure 1. PWD was first discovered in indigenous species in North America, such as the United States, Canada, and Mexico, but now most native

species are resistant [4]. However, the suspected trees infected with PWD, which has spread to other countries, causes browning and death in indigenous species in the region, causing much damage. For example, in 1905, the first damage caused by PWD occurred in Japan in Northeast Asia. Since then, the number of affected areas and countries has increased, including Nanjing City in China in 1982, Taiwan in 1985, and Busan City in Republic of Korea in 1988 [5], [6]. PWD occurring in Republic of Korea are mainly caused by pine species such as *Pinus densiflora*, and *Pinus thunbergii* in April-October [7]. At present, to prevention of PWD is mainly carried out by killing the mediated insects rather than killing the pine wilt nematodes itself, which has no ability to move and breeds in the trees [8]. However, the breeding and movement of mediated insects spread rapidly in unspecified directions and is difficult to predict [9]. Therefore, it is very important to quickly and accurately detect and eliminate the PWD-infected trees in which the larvae of the mediated insect lives.

Currently, researches on image analysis and remote sensing techniques for detecting PWD-infected trees are ongoing. First, Lee *et al.* (2014) studied the spectral reflection characteristics of the hyperspectral images for PWD-infected trees [10]. As a result, it was confirmed that the most significant difference appeared in the red wavelength (688nm) region of the hyperspectral image according to the before and after of PWD. Kim *et al.* (2018) studied the optimal index for detecting PWD-infected trees. [11]. As a result, it was confirmed that among the 10 vegetation indicators, vegetation indicators using red and infrared band wavelengths showed a greater variation in images before and after PWD-infected. Kim *et al.* (2010) have detected the PWD-infected suspicious trees by visual detection using a helicopter based on GIS and GPS information [12]. Park *et al.* (2016) extracted areas where PWD-infected trees lived using an object-based classification model based on IKONOS satellite images [13]. However, satellite images are vulnerable to low spatial resolution and to various noises such as clouds and yellow dust. Also, the visual detection method requires large

Manuscript received August 00, 2020; revised October 00, 2020; accepted November 00, 2020. Date of publication April 00, 2021; date of current version June 00, 2021. This work was supported in part by the National Research Foundation (NRF) of Korea under Grant NBF-2018R1D1A1B07043406. Also, this study has been conducted with the support of the Korea Institute of Industrial Technology (KITECH). (Corresponding author : *Seung Hyun Jeong* and *Min Young Kim*.)

H. G. Park, S. H. Jeong, J. P. Yun are with the Korea Institute of Industrial Technology, Cheonan, 31056, Republic of Korea (e-mail: [dd4680@kitech.re.kr](mailto:dd4680@kitech.re.kr); [shjeong@kitech.re.kr](mailto:shjeong@kitech.re.kr); [rebirth@kitech.re.kr](mailto:rebirth@kitech.re.kr)).

M. Y. Kim is with the School of Electronics Engineering, Kyungpook National University, Deagu, 41566, Republic of Korea (e-mail: [minykim@knu.ac.kr](mailto:minykim@knu.ac.kr)).

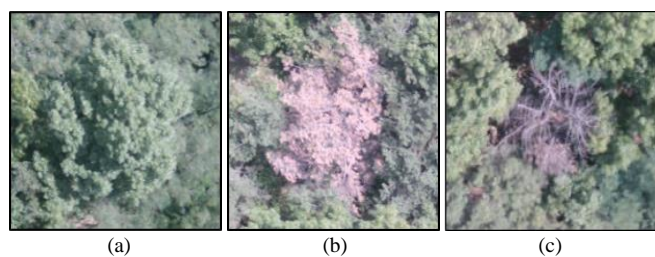


Fig. 1. Aerial photography of trees using drones:(a) is a healthy tree; (b) is a tree that is undergoing browning in the early stages of infection with PWD; and (c) is a tree that has died due to the infection of the PWD and shows an angular form.

amount of time and high costs.

In recent years, the use of unmanned aerial vehicles (UAV) has rapidly expanded, with the ability to support various types of imaging equipment such as RGB and multi-spectral cameras [14]. UAVs are largely divided into rotor or fixed wing type aircraft. The rotorcraft is composed of propeller-type wings and is actively used in forest applications. [15]. Also, image data captured by attaching a multi-spectral camera to a rotorcraft is widely used to distinguish vegetation by providing various spectral band information such as Green, Red, NIR and Red Edge as well as RGB image information [16], [17]. Collected image data is used for image processing and machine learning techniques to quickly and accurately detect tree objects [18, 19]. For example, Valerie Trichon *et al.* (2001), a tree detection technique was performed using specific descriptors such as crown size, crown contour, foliage cover, foliage color and foliage texture from images taken using multi-spectral or RGB cameras [20]. Saeed Al Mansoori *et al.* (2018), based on RGB image and NDVI vegetation index mask, detected 95% of the palm trees using a combination of Circular Hough Transform (CHT) and morphological operators [21]. Takeshi Hoshikawa *et al.* (2020) classified and detected trees infected with pine wilt disease through a model that combines individual tree detection (ITD) and logistic regression-based classification using multispectral and visible color imagery acquired by UAV [22]. As a result, the classification accuracy was 84%. Mutiara Syifa *et al.* (2020) detected trees infected with pine wilt disease through ANN and SVM methods based on the data collected through drone remote sensing [23]. As a result, the SVM model detection performance higher than that of the ANN model was 94.13% and 86.59% for each of the two regions.

Recently, deep learning technologies developed and applied in the field of remote sensing. As an example, Morales *et al.* (2018) conducted a study of automatic segmentation using deep learning technique based on *Mauritia flexuosa* aerial image taken UAV [24]. As a result, the predicted accuracy of the trained model was about 98.1%. Safonova *et al.* (2019) conducted a study to classify Fir Trees damaged by Bark beetle using deep learning techniques based on aerial images captured by UAV [25]. As a result, the predicted accuracy of the trained model was about 98.7%. Xiaoling Deng *et al.* (2020) detected trees infected with Pine wilt disease through an improved Faster R-CNN model based on RGB images rather than multi-spectral images acquired by UAV Remote Sensing [26]. As a result, the detection model performance was 89%. Most deep learning

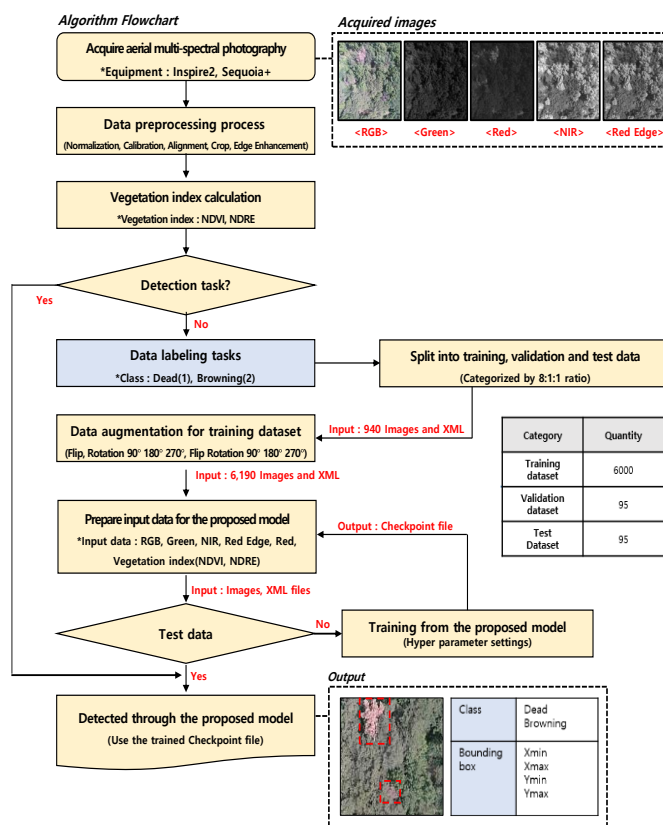


Fig. 2. Proposal detection process. After performing image preprocessing with different combinations of inputs for 7 types of multispectral images (RGB, Green, Red, NIR, Red Edge) and vegetation index (NDVI, NDRE), and then the model is trained. The test result is the classification class and predicted bounding box coordinates in the image.

based algorithms used in computer vision are Convolutional Neural Network(CNN) designed to mimic artificial neural networks, and supervised learning method that predicts after learning feature points (e.g., edge, corner, etc.) of many image data [27, 28]. Among these deep learning-based algorithms, Faster R-CNN is an object detection model and consists of a two-stage network in which RPN (Region Proposal Network) and R-CNN are merged [29].

To our best knowledge, detection of suspicious PWD-infected trees through multi-channel CNN-based object detection has not been conducted yet. Therefore, in this study, multi-spectral aerial images were collected through the rotorcraft drone of *DJI Inspire 2* model, and the suspected trees infected with PWD were detected by using multi-channel CNN-based object detector. Also, the detection accuracies according to the combination of each channel and the use of vegetation index are compared.

In summary, the main contributions of this study are summarized as follows.

- This study proposed a method to detect suspected PWD-infected trees using a multi-channel CNN-based object detection method. Through the experiment, it was confirmed that the detection accuracy can be improved by using the proposed method rather than only 3-channel RGB information.

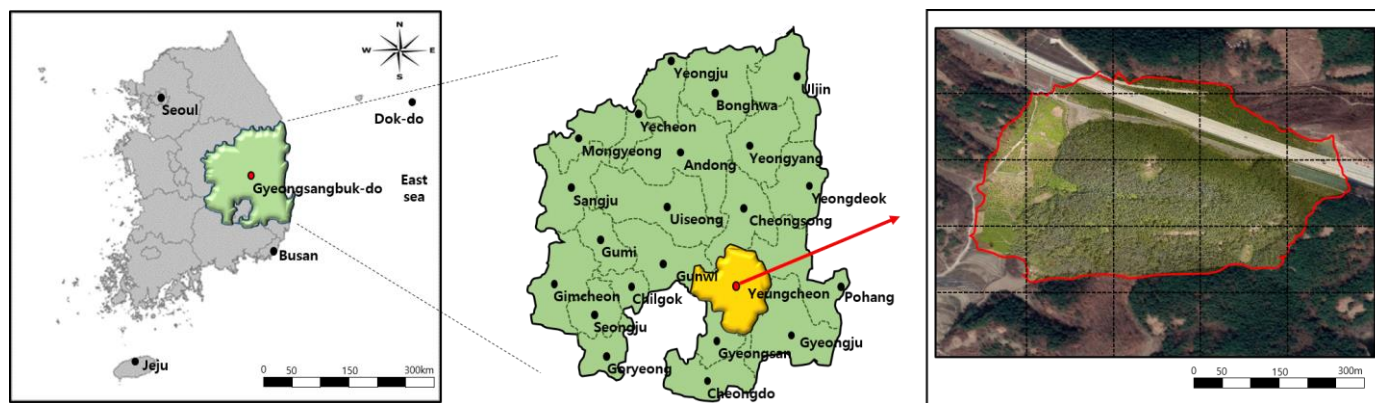


Fig. 3. Study area of the mountain near Daegi-ri, Yeongcheon-si, Gyeongsangbuk-do, Republic of Korea. The area within the red perimeter in the right-most picture is the area where remote sensing was performed in this study.

- We proposed a preprocessing to enhance the accuracy of multi-channel object detection for multi-spectral images. The calibration, alignment, and image enhancement process to be performed for the multi-spectral images are presented.
- A comparative study was conducted according to the combination of channels that increase the detection accuracy. The accuracy was compared according to usage of multi-spectral image and vegetation index. Through the experiments, the accuracy was increased when using RGB, NIR, Red Edge and normalized difference red-edge (NDRE) index.

This paper consists of the four sections. In the second section, the data acquisition and preprocessing of the multi-spectral drone aerial photographs, and the multi-channel object detection method are presented. The application results according to the channel combination are analyzed and discussed in third section. Finally, conclusions and findings will be presented in the fourth section.

## II. MATERIALS AND METHODS

The overall process of proposed method was summarized in figure 2. As shown, the aerial multi-spectral images are acquired by using Inspire2 equipped with Sequoia+. After the preprocessing, it branches according to the detection task. If it is a detection task, the detection is performed by using the trained detection model. If it is not a detection task, the data labeling task for the detection of dead and browning trees is performed. After the data labeling, the data is split into training, validation, and test dataset. The training is performed by using the pre-processed and augmented training data, and validation loss is calculated at each epoch. The model with best validation loss is saved during training. Finally, the model with best validation loss is used for the detection of dead and browning trees and the final accuracy is calculated by using test dataset which is not considered in training. The detailed explanation of the process will be given in this section.

### A. Image acquisition

#### 1) Study area

This study was conducted in the area of Daegi-ri, Yeongcheon-si, Gyeongsangbuk-do, and Republic of Korea. This area is one of the areas affected by PWD in Republic of Korea over the past five years. The total area is 180, 663 m<sup>2</sup> and the circumference is about 1.67 km as shown in Figure 3. Most of the forest species in the study area consist of *Pinus densiflora*, and *Pinus koraiensis*. Most of these species are vulnerable to PWD. As a result, the area is classified as a region that is more likely to be damaged by PWD. Therefore, it was considered to be a suitable area to acquire image data of suspected trees infected with PWD and was selected as this study area.

#### 2) Study equipment

This study used *DJI Inspire 2* of rotorcraft drone and *Parrot Sequoia+*, which is a multi-spectral camera to collect drone aerial multi-spectral photography. First, *DJI Inspire 2* consists of four rotating vanes [30]. It is relatively convenient to operate compared to fixed-wing drones and has the advantage of maintaining flight at a fixed point, which is easy for aerial image shooting. Secondly, the multi-spectral camera *Parrot Sequoia+* consists of five optical lenses and a sunlight sensor [31]. Five optical lenses absorb RGB, Green, Red, NIR and Red Edge wavelengths for each lens and output high resolution images. In addition, the frame was built by 3D printing to mount the *Parrot Sequoia+* multi-spectral camera on the *DJI Inspire 2*.

#### 3) Drone aerial multi-spectral image

To collect the drone aerial multi-spectral imagery, a multi-spectral camera *Parrot Sequoia+* was attached to *DJI Inspire 2* rotorcraft to shoot at an altitude of 150 m. The ground sample distance (GSD) of the photographed picture is about 8 cm/pixels, which represents a higher spatial resolution than the satellite image. The *Parrot Sequoia+* of multi-spectral camera supports the RGB, Green, Red, NIR and Red Edge band channel, so that five multi-spectral aerial images are captured per shooting point. As a result, the drone aerial multi-spectral

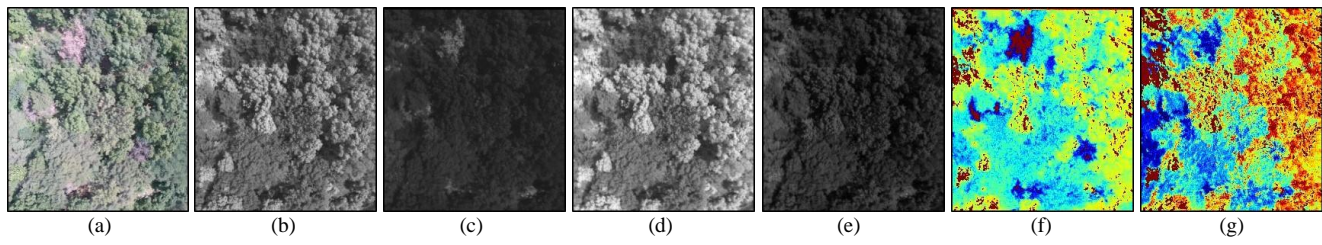


Fig. 4. Images of RGB, multi-spectral and vegetation index. (a) RGB; (b) Green; (c) Red; (d) NIR; (e) Red Edge; (f) NDVI; (g) NDRE.

photography captured 4,608 by 3456(RGB band), 1,280 by 960 pixels(Green, Red, NIR and Red Edge band) at 940 images at 90 points and 98 points on April 11, 2019 and June 12, 2019. Thereafter, the collected 940 images were classified into 188 photographing points for each of five multi-spectral images.

### B. Image preprocessing and augmentation

Drone aerial multi-spectral imagery performed six stages of image processing. Through the image processing, the hardware limitations such as radiation distortion, distance error of each camera lenses and resolution difference between RGB and mono-band images can be overcome. Also a large amount of data set was secured through data augmentation after edge enhancement to highlight the characteristics of the target object and image cropping for model optimization.

#### 1) Image calibration

The first step is image calibration. Radiation distortion caused by the refractive index of the camera convex lens was corrected through image calibration [32]; converted pixel values in an image to absolute spectral emission values in  $W/m^2/sr/nm$ ; corrected sensor black level, sensor sensitivity, sensor gain and exposure settings, and lens vignetting effects. All parameters used in the model can be read from XML metadata in TIFF files saved by multi spectral cameras. Following is the (1) for image calibration.

$$L_i(x, y) = V_i(x, y) \frac{a_{1,i}}{g_i} \frac{p_i(x,y) - P_{BL,i}}{t_{e,i} + a_{2,i}y - a_{3,i}t_{e,i}y} \quad (1)$$

The following describes the variable arguments of (1), where  $V_i(x, y)$  is the vignette correction function map;  $a_{1,i}$ ,  $a_{2,i}$ , and  $a_{3,i}$  are radiometric calibration coefficients;  $P_i(x, y)$  is the raw digital count image;  $P_{BL,i}$  is the dark level;  $g_i$  is the gain;  $t_{e,i}$  is the exposure time[s]; and  $y$  is the pixel number. The vignette map,  $V_i(x, y)$ , can be represented as

$$V_i(x, y) = \frac{1}{K_i} \quad (2)$$

where  $K_i$  is a correction factor.

$$K_i = 1 + k_{0,i}R_i + k_{1,i}R_i^2 + k_{2,i}R_i^3 + k_{3,i}R_i^4 + k_{4,i}R_i^5 + k_{5,i}R_i^6 \quad (3)$$

where  $k_{0,i}$  through  $k_{5,i}$  are polynomial correction coefficients

and  $R_i$  is the distance of the pixel to the vignette centers.

$$R_i = \sqrt{(x - c_{x,i})^2 + (y - c_{y,i})^2} \quad (4)$$

where  $c_{x,i}$  and  $c_{y,i}$  represent the vignette center and  $i$  denotes the spectral band number. Now, by combining all the above equations (Equations 1-4), the radiance image,  $L_i(x, y)$ , can be computed.

#### 2) Image normalization and resizing

The second step is pixel normalization. The raw DN values of all channels except the RGB channel are from 0 to 65535 in 16-bit format. Therefore, to match the RGB channel, which is the DN value in the range of 0 to 255, the DN values of the Red, Red Edge, NIR, and Green channel were normalized to the same range of RGB image. Because the pre-trained initial weights are trained based on RGB images converted to 32-bit float, the input image was normalized equal to the range of the RGB image. The pre-trained feature extractor used in this study, VGG, subtract mean values for RGB channels. The values -122.68, 116.779, and 103.939 for RGB respectively are calculated from ImageNet dataset. This subtraction is same for RGB and normalized multispectral images of the proposed model. For the vegetation index, we multiply 100 because the value of vegetation index can have a range from -1 to 1. By the multiplication, the range of vegetation index channel becomes similar with the other channels. Note that, the loss of information due to normalization is small because the DN values of multi-spectral channel are converted from 16-bit unsigned integer to 32-bit float to feed the CNN network. To feed the CNN network, the DN values of RGB image are also converted from 8-bit unsigned integer to 32-bit float. In addition, the RGB image was resized from 4096 by 3456 to 1280 by 960.

#### 3) Image alignment.

The third step is image alignment. The position error in the image caused by the difference in the distance between the camera lenses for each spectrometer of the *Parrot Sequoia+* of multi-spectral camera was corrected through the SIFT method [33]. The detailed method is that after extracting the feature points of each spectral image through the SIFT method, each spectral image was matched based on the Green band image. As a method for matching, as shown in (5) below. Homography matrix was calculated through RANSAC estimation technique based on each feature point to match the feature points between

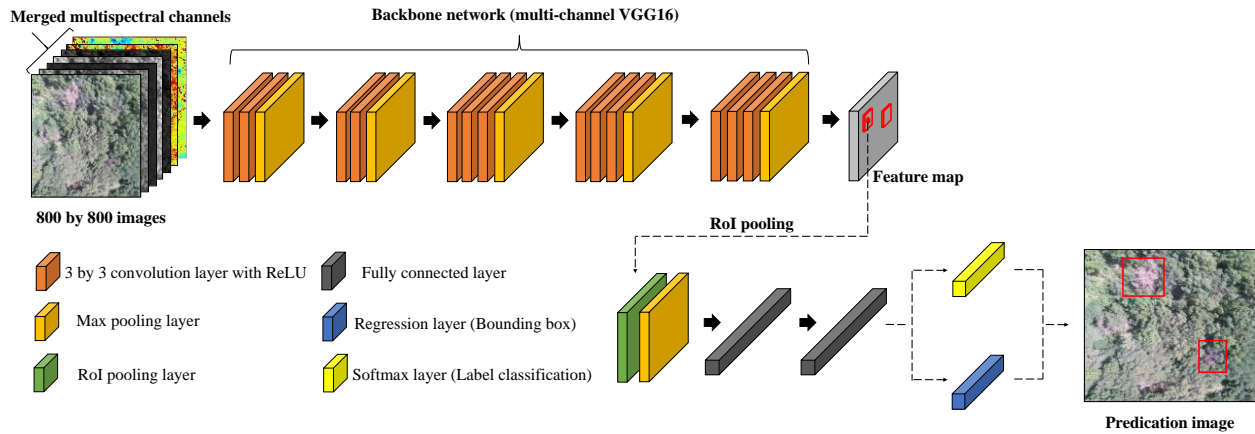


Fig. 5. Multi-channel CNN-based object detection model for detecting suspected trees infected with PWD. A structure that combines a multi-channel VGG16-based RPN model to create a region proposal and an R-CNN model to detect infectious object.

the two images. When accurate homography matrix is calculated, all pixels of the image are converted and mapped by warping based on the green band image.

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = H \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} \quad (5)$$

The following describes the homography arguments of (5), where  $x_1$  and  $y_1$  are the pixel value after warping. And  $H$  is the homography matrix. And  $x_2$  and  $y_2$  are the pixel value before warping.

#### 4) Image enhancement

The fourth step is image enhancement. For the corrected image, the image enhancement process using the  $3 \times 3$  sharpening filter highlights the edges and corners that are the feature points of the objects in the image [34]. The following (6) performs image enhancement through the sharpening mask filter using laplacian.

$$I_{enhancement}(i, j) = M * I(i, j), M = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (6)$$

#### 5) Image cropping

The fifth step is image cropping. In the existing  $1280 \times 960$  pixel size image, the image was cropped to  $800 \times 800$  pixel size from the center of the image. Equation (7) is a formula for cropping to  $800 \times 800$  pixels based on the center of the image.

$$I_{crop}(i, j) = I\left(\frac{i}{2} - 400, \frac{j}{2} - 400\right) \quad (7)$$

#### 6) Image augmentation

The entire 940 images was divided into a dataset for training, validation, and testing at a ratio of around 8:1:1. Subsequently, the training set was augmented. From the initial 750 training data images, 6000 images were obtained through flipping, rotation [35].

#### C. Labeling of suspicious trees

Symptoms of infection with PWD mainly cause browning and death. As a result, when observing aerial photographs using drones, trees infected with PWD represent trees covered with brown leaves and barren trees. Therefore, we mark suspected trees infected of PWD in aerial photographs collected to create training data to be used for deep learning-based object detection model. The marked trees are labeled in a rectangular box, separated into two class types (browning and dead). As a result, the xmin, xmax, ymin, and ymax coordinate values are extracted in XML file format with class information.

#### D. Vegetation index calculation

The vegetation index of the image according to each shooting location was calculated using the preprocessed multispectral image data set. Equations 8 to 9 below are formulas for calculating NDVI(Normalized Difference Vegetation Index) and NDRE(Normalized Difference Red-Edge Index) [36], [37].

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (8)$$

$$NDRE = \frac{NIR-RED\ EDGE}{NIR+RED\ EDGE} \quad (9)$$

#### E. Multi-channel CNN-based object detection

The model used to detect suspected trees infected with PWD is CNN-based object detection model and consists of 2 stage network structure that merges region proposal network (RPN) which extracts region proposal and R-CNN which performs classification and detection prediction, is shown in Figure 5. Loss function, which is a function that indicates the difference between the predicted value and the ground truth value, is shown in (10) below and converges to the minimum value through training. Adam Optimizer, which is used to converge the Loss function to the minimum value, performs an optimization process through the stochastic gradient descent based on the gradient amount of the past [38].

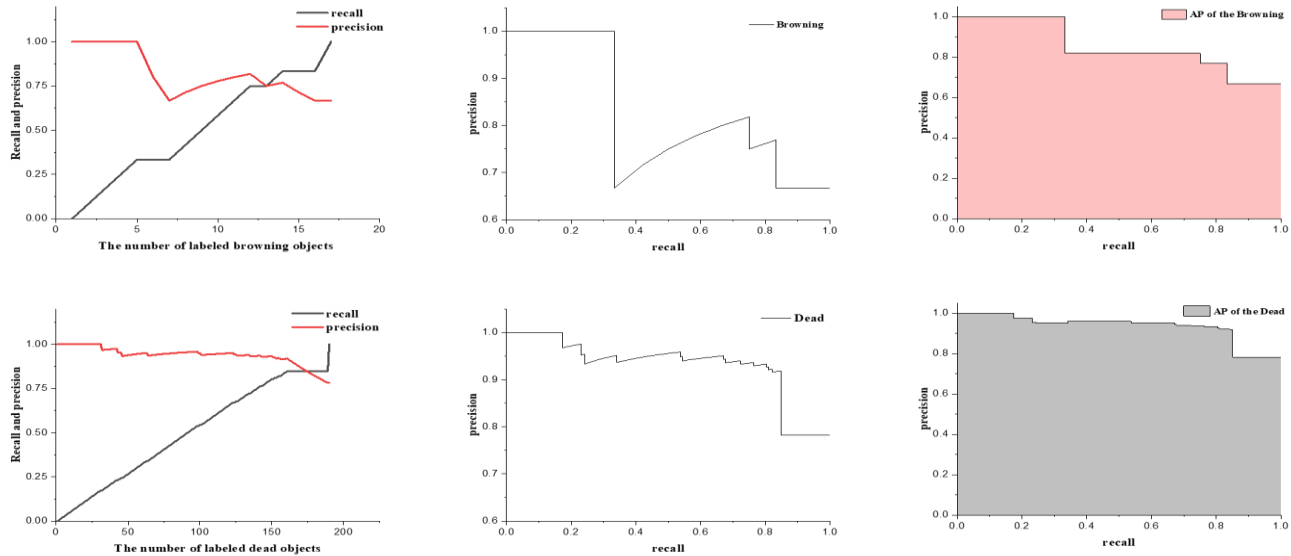


Fig. 6. AP graph of each classes (Dead, Browning). The graph in the first row is graphs related to the browning class, and the graph in the second row is graphs related to the dead class. Also, the graphs in the last column are graphs for calculating the AP, and the area of the colored area is the AP result value for each class.

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (10)$$

The following describes the variable arguments of (10). Here,  $i$  is the index of the anchor within the mini batch;  $P_i$  is the prediction whether anchor;  $i$  is an object or background;  $p_i^*$  is a ground truth label of 1, where 1 means that the anchor is positive and 0 means that the anchor is negative;  $t_i$  is the vector containing four coordinates of the bounding box;  $t_i^*$  is ground truth box, associated with a positive anchor;  $L_{cls}$  is log loss (object loss or not),  $L_{reg}$  is smooth L1 loss function (only used in positive anchors  $p_i^* = 1$ ); and  $N_{cls}$  is normalization. This is same as mini batch size, where  $L_{reg}$  is normalization and also same as the number of anchor locations, where  $\lambda$  is 10 by default. Equation 11 below describes the smooth L1 function in described in (10).

$$Smooth_{L1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (11)$$

The feature extractor used in this study is based on VGG16 [39]. Because the proposed method uses multi-channel input, it is necessary to increase the number of input channels of a feature extractor that generally uses 3 channels. Therefore, the number of channels in the front layer is adjusted according to the number of multi-channels. RPN consists of thirteen convolution layers and R-CNN consists of three fully connected layers. The final layer of the RPN consists of a classification layer that predicts the presence or absence of an object for a region proposal, and a regression layer that predicts the location of the region proposal. The final layer of the R-CNN consists of a softmax layer that predicts the class of the object and a

regression layer that predicts the location of the object. In addition, it prevented overfitting of training data through dropout when moving fully connected layer in R-CNN [40].

The initial parameter (e.g., Weight, bias) of the first convolution layer of VGG, which is feature extractor of the proposed model, extended by concatenation from 3 channel to multi-channel. Then, the weights of extended pre-trained model were set as the initial value of the multi-channel CNN-based object detector. The pre-trained initial weights generally give higher accuracy than the randomly set initial value parameter [41], [42]. In addition, in order to build a training model that shows high detection performance, hyper-parameters were manually determined through a number of experiments. In particular, the decision value for the threshold for extracting region proposals from CNN was determined from the low value to the high value through sequential enhancement to determine the optimal value. As a result, when the overlapping area of anchor box and label ground truth is 0.5 or more to make region proposals, the region proposal of the object class is made. When 0.3 or less, a background class region proposal is created. The ratio of object and background region proposal was set at 1:1 ratio. As a result, the maximum number of region proposals created was 12000. In order to filter out a large number of region proposal, the overlapping area between each region proposal created through non-maximum suppression (NMS) was filtered out. As a result, the final region proposal is 2000. After that, the learning rate was set to 0.0005, the batch size was set to 10, and epoch was set to 100.

### III. RESULTS AND DISCUSSION

This section presents the results of training and test based on the data set refined by the image preprocessing and the multi-channel CNN-based object detection. The evaluation index, mAP, was used to evaluate the performance of the trained

TABLE I  
RESULT OF EVALUATION INDEX VALUE ACCORDING TO COMBINATIONS OF INPUT DATA

Input data (The number of channel)	mAP			Dead AP			Browning AP		
	Mean	Best	Worst	Mean	Best	Worst	Mean	Best	Worst
RGB (3 channel)	77.06	80.16	73.34	93.12	95.02	91.98	60.99	68.33	51.66
RGB + Green + Red + NIR + Red Edge (7 channel)	81.06	84.32	79.49	91.23	91.97	90.34	70.89	76.66	67.36
RGB + NIR + Red Edge + NDVI (6 channel)	81.91	86.75	72.36	88.95	92.01	84.41	74.87	83.05	60.31
RGB + NIR + Red Edge + NDRE (6 channel)	82.70	86.63	84.84	93.96	95.48	93.58	76.66	77.77	76.11

model. As a result, we can confirm the performance of multi-channel CNN-based object detection on suspicion multi-spectral aerial photography.

#### A. Analysis of the training model

The language used to develop the learning model is Python, and the development environment used is the Tensorflow 1.12 version of the framework. TensorFlow is an open source software library that utilizes distributed graphics processing unit (GPU) resources to deliver high learning performance through high-performance numerical computation. Hardware resources for training and testing used I7-8086K CPUs and accelerated by using GeForce GTX 1080 Ti GPUs.

There are 6,190 image patches: 6000 for training 95 for validation, and 95 for test refined through image processing and augmentation consist of five types of multi-spectral images(RGB, Green, Red, NIR and Red Edge). The final dataset, consisting of RGB, Green, Red, NIR and Red Edge band image patches, was divided into each spectral band dataset consisting of 1, 238 image patches (1, 200 types of images are used for training, 19 types of images are used for remaining validation and testing) of different capture points and data augmentation types. After that, based on the divided training data set, RGB (3 channels), Green, Red, NIR, Red Edge, NDVI, NDRE (1 channel each) were merged into various combinations. Although the green and red channels of RGB have almost similar values to the green and red channels of a multi-spectral camera, they were input together because of the difference in scale and resolution between two lenses. In Table 1, 4 types of multi-channel combination were compared through cross-validation of 3 random sampling. As a result of training, the maximum performance of the training model showed the lowest loss function value in the 99 epochs. Also, the loss value of the validation data set shows a low value similar to that of the training data set. Therefore, the overfitting of the training data set did not occur.

#### B. Evaluation index

The performance verification of the proposed method is evaluated by the mAP value. mAP is an average value of AP for each class and is used as an evaluation index for analyzing the performance of an object detection model. The AP for each

class is calculated by predicting the ground-truth of each labeled object for all test images. The correct answer must satisfy both of the following conditions. First, the intersection over union (IoU), which is the intersection of the prediction object and the ground-truth object, is calculated and satisfied if the value is 0.5 or more. Second, the class of prediction object and ground-truth object should same to satisfy. After that, the ground-truth object of each labeled object that is identified as correct is sequentially listed to calculate the precision and recall values. Equation (13) and (14) are recall and precision. In (13) and (14), TP is true positive, FP is false positive, and FN is False negative. True is the correct answer of ground-truth, and False means the incorrect answer of ground true. In addition, positive is the result predicted by the correct answer, and negative means the result predicted by the incorrect answer. After that, the precision and recall values are set to each axis of the two-dimensional coordinate plane, and the calculated area for each class is displayed on the coordinate plane, and the corresponding points are connected to calculate the area. The equation for calculating the area is the same as (15), and when the  $r$  (is recall) value increases sequentially,  $P$  (is precision) becomes the maximum value, it becomes  $P_{interp}$  value, and this value multiply the recall value. The AP values are derived as a result of summing all the multiplied values by repeating this method. Then, the average value of the calculated AP for all classes is mAP, which is shown in (16).

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (12)$$

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

$$Recall = \frac{TP}{TP+FN} \quad (14)$$

$$AP = \sum (r_{n+1} - r_n) P_{interp}(r_{n+1}),$$

$$P_{interp}(r_{n+1}) = \max_{\tilde{r} \geq r_{n+1}} p(\tilde{r}) \quad (15)$$

$$mAP = \frac{\sum_{k=1}^n AP_k}{n} \quad (16)$$

In (15),  $n$  is a number that indicates an increase sequentially when recall value that is from 0 to 1,  $r_n$  is the recall value when



Fig. 7. Detection visualization of RGB images based on browning and dead trees. The blue box is the labeled ground true area, and the red box is the area predicted by the trained model.

$n$  is 0 and  $r_{n+1}$  is the recall value when  $n$  is 1;  $P_{interp}$  is the point where precision is maximized by continuously inputting the recall value; and  $\tilde{r}$  is the predicted recall value as it increases continuously. And  $P(\tilde{r})$  is the precision when the predicted recall value increases continuously.

### C. Inference result

Prediction and evaluation results of the trained model (6 channels : RGB + NIR + Red Edge + NDRE) using the test dataset are shown in table 1. The graphs of the model AP results in two classes for browning and dead, which is a symptom of PWD infection, are shown in Figure 6. Figure 6 shows the results of achieving the average performance for three crossover experiments, the AP value for the dead symptom is 95.48% and the AP value for the browning symptom is 77.77%. Therefore, the performance of multi-channel CNN-based object detection through data augmentation is mAP 86.63%, which is the average value of AP for browning and dead classes. The following Figure 7 is a visualized figure of the inferred class and bounding box, which will be an RGB image.

## IV. CONCLUSIONS

Recently, most deep learning-based tree detection models using aerial photography are based on RGB images. However, it is difficult to expect high detection performance because the detection model using the RGB image does not simply consider vegetation information of various band spectrums. In this paper, we propose a multi-channel CNN-based object detection to detect trees suspected of being infected with PWD. The proposed model detects suspected trees of PWD after training based on multispectral aerial photography in RGB, Green, Red, NIR and Red Edge bands acquired through a rotary wing drone equipped with a multispectral camera. In addition, in order to improve detection accuracy, image preprocessing was performed through image calibration, image alignment, and image enhancement. As a result of the study, the multi-channel CNN-based object detection model using RGB, NIR, red edge, and NDRE index was the best, and the final detection performance was mAP 86.63%. The proposed detection model can be applied to more detailed tree detection for dead and browning symptoms. Therefore, the proposed detection model is expected to be useful for detecting trees with symptoms of

various diseases.

## REFERENCES

- [1] Rutherford. T. A, "Nematode-induced pine wilt disease: factors influencing its occurrence and distribution," *Forest Science*, vol. 36, no. 1, pp. 145-155, 1990
- [2] Futai. K, "Pine wood nematode, *Bursaphelenchus xylophilus*," *Annual Review of Phytopathology*, vol. 51, pp. 61-83, 2013.
- [3] D.-S. Kim et al., "Escape of pine wood nematode, *Bursaphelenchus xylophilus*, through feeding and oviposition behavior of *Monochamus alternatus* and *M. saltuarius* (Coleoptera: Cerambycidae) adults," *Korean journal of applied entomology*, vol. 48, no. 4, pp. 527-533, 2009.
- [4] Bergdahl. D. R, "Impact of pinewood nematode in North America: present and future. *Journal of Nematology*," vol. 20, no. 2, pp. 260, 1988.
- [5] Mamiya. Y, "History of pine wilt disease in Japan. *Journal of nematology*," vol. 20, no. 2, pp. 219, 1988.
- [6] Mota. M. M. and Vieira P, "VIEIRA, Paulo (ed.). *Pine wilt disease: a worldwide threat to forest ecosystems*. Heidelberg, Germany," Springer, 2008.
- [7] S.-C. Shin, "Pine wilt disease in Korea. In: *Pine wilt disease*," Springer, Tokyo, pp. 26-32, 2008.
- [8] S.-C. Sin, "Current Status of Damage, Control and Research of Pine Barley Nematode Disease," *Proceedings of the Korean Society of Applied Entomology Conference*. Korean Society of Applied Entomology, pp. 33-39, 2005.
- [9] D.-S. Kim et al., "The allegorical ecology of the adult imago of the sea salt of pine tree nematode," *Korean journal of applied entomology*, vol. 42, no. 4, pp. 307-313, 2003.
- [10] J.-B. Lee et al., "Spectroscopic Characterization of Pine Tree Nematode Disease Infected Trees Using a Ground Hyperspectral Camera," *Korean Journal of Remote Sensing*, vol. 30, no. 5, pp. 665-675, 2014.
- [11] S.-R. Kim et al., "Hyperspectral analysis of pine wilt disease to determine an optimal detection index," *Forests*, vol. 9, no. 3, pp. 115, 2018.
- [12] J.-B. Kim et al., "Development of an aerial precision forecasting techniques for the pine wilt disease damaged area based on GIS and GPS," *Journal of the Korean Association of Geographic Information Studies*, vol. 13, no. 1, pp. 28-34, 2010.
- [13] J.-M. Park et al., "Detection of Trees with Pine Wilt Disease Using Object-based Classification Method," *Journal of forest and environmental science*, vol. 32, no. 4, pp. 384-391, 2016.
- [14] Everaerts. J, "The use of unmanned aerial vehicles (UAVs) for remote sensing and mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 37, no. 2008, pp. 1187-1192, 2008.
- [15] Tang. Lina, "Drone remote sensing for forestry research and practices," *Journal of Forestry Research*, vol. 26, no.4, pp. 791-797, 2015.
- [16] Frankf. Jonas, "Multi-temporal wheat disease detection by multi-spectral remote sensing," *Precision Agriculture*, vol. 8, no. 3, pp. 161-172, 2007.
- [17] Laliberte, A.S et al., "Multispectral remote sensing from unmanned aircraft: Image processing workflows and applications for rangeland environments," *Remote Sensing*, vol. 3, no. 11, pp. 2529-2551, 2011.
- [18] Puliti. S, "Tree-stump detection, segmentation, classification, and measurement using unmanned aerial vehicle (UAV) imagery," *Forests*, vol. 9, no. 3, pp. 102, 2018.



- [19] Feduck, C, "CASTILLA, Guillermo. Detection of coniferous seedlings in UAV imagery," *Forests*, vol. 9, no. 7, pp. 432, 2018.
- [20] Trichon, V, "Crown typology and the identification of rain forest trees on large-scale aerial photographs," In: *Tropical Forest Canopies: Ecology and Management*. Springer, Dordrecht, pp. 301-312, 2001.
- [21] Al Mansoori, S, "Hussain. Automatic palm trees detection from multispectral UAV data using normalized difference vegetation index and circular Hough transform," In: *High-Performance Computing in Geoscience and Remote Sensing VIII. International Society for Optics and Photonics*, pp. 1079203, 2018.
- [22] Hoshikawa, Takeshi, Kazukiyo. Yamamoto, "Individual Tree Detection and Classification for Mapping Pine Wilt Disease Using Multispectral and Visible Color Imagery Acquired from Unmanned Aerial Vehicle," *Journal of The Remote Sensing Society of Japan*, vol. 40, no. 1, pp. 13-19, 2020.
- [23] Syifa, Mutiara, S.-J. Park, C.-W. Lee, "Detection of the Pine Wilt Disease Tree Candidates for Drone Remote Sensing Using Artificial Intelligence Techniques," *Engineering*, vol. 6, no. 8, pp. 919-926, 2020.
- [24] Morales, G, "Automatic segmentation of *Mauritia flexuosa* in unmanned aerial vehicle (UAV) imagery using deep learning," *Forests*, vol. 9, no. 12, pp. 736, 2018.
- [25] Safonova, A, "Detection of fir trees (*Abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning," *Remote Sensing*, vol. 11, no. 6, pp. 643, 2019.
- [26] Deng, Xiaoling et al., "Detection and Location of Dead Trees with Pine Wilt Disease Based on Deep Learning and UAV Remote Sensing," *AgriEngineering*, vol. 2, no. 2, pp. 294-307, 2020.
- [27] Jiao, L, "A Survey of Deep Learning-Based Object Detection," *IEEE Access*, vol. 7, pp. 128837-128868, 2019.
- [28] Han, J, "Advanced deep-learning techniques for salient and category-specific object detection: a survey," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 84-100, 2018.
- [29] Ren, S, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, pp. 91-99, 2015.
- [30] Corp. DJI, China [Online]. Available: <https://dji.com/kr/inspire-2/info>, Accessed on: 2020
- [31] Pix4D, Switzerland [Online]. Available : <https://www.pix4d.com/product/sequoia/faq>, Accessed on: 2020
- [32] Yahyanejad, S, Misomy, J, Rinner, B, "Lens distortion correction for thermal cameras to improve aerial imaging with small-scale UAVs," 2011 IEEE International Symposium on Robotic and Sensors Environments (ROSE). IEEE, pp. 231-236, 2011.
- [33] Karami, E, Prasad, S, Shehata, M, "Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images," arXiv preprint arXiv:1710.02726, 2017.
- [34] Tai, H. T, "Pixel interpolator with edge sharpening," U.S. Patent, no. 5, 054,100, 1991.
- [35] Mikolajcayk, A, Grochowki, M, "Data augmentation for improving deep learning in image classification problem," 2018 international interdisciplinary PhD workshop (IIPhDW). IEEE, pp. 117-122, 2018.
- [36] Carlson, Toby N, David A. Ripley, "On the relation between NDVI, fractional vegetation cover, and leaf area index," *Remote sensing of Environment*, vol. 62, no. 3, pp. 241-252, 1997.
- [37] Handique, B. K et al., "Crop discrimination using multispectral sensor onboard unmanned aerial vehicle," *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*, vol. 87, no. 4, pp. 713-719, 2017.
- [38] Kingma, D. P, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
- [39] Simonyan, K, Zisserman, A, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
- [40] Srivastava, N, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929-1958, 2014.
- [41] Zeiler, Matthew D, Fergus Rob, "Visualizing and understanding convolutional networks," *European conference on computer vision*. Springer, Cham, pp. 818-833, 2014.
- [42] Phillips, M. C, Stein, R, Park T, "Analyzing Pre-Trained Neural Network Behavior with Layer Activation Optimization," 2020 Systems and Information Engineering Design Symposium (SIEDS). IEEE, pp. 1-6, 2020.



**Hae Gwang Park** received the B.S degree in industrial system engineering from the Keimyung University, Republic of Korea, in 2019, and, is currently pursuing the master's course in school of electrical engineering at the Kyungpook National University, Republic of Korea. His research interest includes computer vision and remote sensing

Currently, He is working as a student researcher at the Korea Institute of Industrial Technology's Safety System Research Group. In 2019, He was a recipient of best paper award in the Fall Conference of the Korean Geographical Information Society & ICSANE 2019 and Korean Society of Mechanical Engineers Fall Conference.



**Seung Hyun Jeong** was born in Daegu, Republic of Korea in 1985. He received the B.S., Ph.D. degrees in mechanical engineering from the Hanyang University, Seoul, in 2008, 2014, respectively.

From 2014 to 2016, he was a senior researcher in the Hyundai Heavy Industries, Construction Equipment Research Department. Since 2016, he has been a senior researcher at the Korea Institute of Industrial Technology's Safety System Research Group. His research interests include optimization algorithms, structural optimization, classification, object detection, semantic segmentation through deep learning, and innovative application of optimization and deep learning.



neural networks.

**Jong Pil Yun** received the Ph.D. degree in electrical engineering from the Pohang University of Science and Technology, in 2009. He is currently a Principal Researcher with the Korea Institute of Industrial Technology. His research interests include defect detection, classification using machine vision, image processing, and deep



**Min Young Kim** received his B.S., M.S. and Ph.D. degrees at Korea Advanced Institute of Science and Technology, Korea, in 1996, 1998 and 2004, respectively. He worked as a senior researcher for Mirae Corp. in 2004–2005, and as a chief research engineer for Kohyoung Corp. in 2005–2009 in the research field of artificial vision systems

for intelligent machines and robots. Since 2009, he has joined Kyungpook National University as an assistant professor of School of Electrical Engineering and Computer Science. Currently, he is an associate professor of School of Electronics Engineering in the same university and is the deputy director of KNU-LG Convergence Research Center and the director of Research Center for Neurosurgical Robotic Systems. He was a visiting associate professor of Dept. Electrical and Computer Engineering and School of Medicine in Johns Hopkins University in 2014–2016. His research interest is the visual intelligence for robotic perception and recognition of autonomous unmanned ground and aerial vehicles.