J Electr Eng Technol.2018; 13(3): 1377-1382
http://doi.org/10.5370/JEET.2018.13.3.1377

ISSN(Print) 1975-0102
ISSN(Online) 2093-7423

A Study of Image Classification using HMC Method Applying CNN Ensemble in the Infrared Image

Ju-Young Lee*, Jae-Wan Lim* and Eun-Jin Koh†

Abstract – In the marine environment, many clutters have similar features with the marine targets due to the diverse changes of the air temperature, water temperature, various weather and seasons. Also, the clutters in the ground environment have similar features due to the same reason. In this paper, we proposed a robust Hybrid Machine Character (HMC) method to classify the targets from the clutters in the infrared images for the various environments. The proposed HMC method adopts human's multiple personality utilization and the CNN ensemble method to classify the targets in the ground and marine environments. This method uses an advantage of the each environmental training model. Experimental results demonstrate that the proposed method has better success rate to classify the targets and clutters than previously proposed CNN classification method.

Keywords: Infrared image, Convolutional neural network, Machine learning.

1. Introduction

An autonomous flight device equipped with an infrared (IR) imaging sensor need to classify the targets in the marine and ground environments during flight time. When they want to find some targets in each area, it is important that the autonomous flight device removes various clutters correctly to improve the target classification capacity in the IR images. Therefore, it is required to improve the performance of the classification algorithm.

Meanwhile, target detection algorithms for marine or ground environment usually adopt various methods to classify the targets or clutters. For example, SURF (Speeded Up Robust Features) [1], RDL (Robust Dictionary Learning) [2] and QFT (Quaternion Fourier Transform) [3] are the methods to classify the targets or clutters. But, these methods can be limited because the targets and clutters have many similar features due to the diverse changes of the air temperature, water temperature, diversity weathers.

For this reason, some researches demonstrated that the CNN (Convolutional Neural Network) method can benefit to the target classifications. Ødegaard's paper [4] showed that a CNN method can be used to classify images of ships in the harbor.

According to the Dr. Hiroshi Tasake's book, 'Everyone has multiple personality [5]', if we consciously can deal with the multiple personality in accordance with the given situations or environments, a variety of talents can be created and hidden possibilities can be founded. If this concept is applied to machine learning, it is possible to deal with the multiple machine characters consciously in accordance with the given situations or environments. Also, if we apply this method to the various subjects that need to be solved by deep learning, we can improve the success rate of classification by discovering the latent possibilities of the machine.

Therefore, we propose the novel method, Hybrid Machine Character (HMC) method, to classify the targets from the clutters in the marine and ground environments by applying a concept of the multiple personality. The HMC utilizes the three training models which are comprised of the marine, ground and circumstance models. The circumstance model is especially used as the preprocessing algorithm to recognize whether the given situation is the ground or marine environment. The ground and marine models are consciously selected according to the result of the circumstance model.

In Section 2.1, we describe the ground learning model to classify the targets from the clutters in the ground environment. In Section 2.2, we describe the marine learning model to classify the targets from the clutters in the marine environment. In Section 2.3, we describe the circumstance learning model that classifies the ground and marine environment. And the HMC is proposed with details of its process and structure. In Section 3, the proposed method compared with the previous research model [4] and others. It showed that the proposed method improved the performance of the classification.

2. Learning Model for HMC

The CNN (Convolutional Neural Network) is a type of feed-forward artificial neural network that is inspired by the organization of the animal visual cortex. Recently,

† Corresponding Author: Agency for Defense Development, Korea. (eikoda@gmail.com)
* Agency for Defense Development, Korea. (developer_y@naver.com, jwan@add.re.kr)
Received: May 15, 2017; Accepted: January 11, 2018

The CNN has been reported that it exhibits excellent performance with the wide applications in the image and video recognitions [6, 7]. Also, as a result of the GPU and computing system development, The CNN is appropriate to solve the various image recognition subjects because the CNN structure can conduct the complex and multiple layers more rapidly than before [8]. Thus, the HMC also adopt the CNN to utilize the successful benefits of the CNN.

HMC applies the CNN to the both marine and ground target recognition methods. As shown in previous studies, if CNN structure consisting of multiple layers is applied, it will be advantageous to increase accuracy. Therefore, in this paper, we propose a method that can improve the accuracy with the CNN layers.

2.1 Learning and classify on the ground environment

The CNN structure for the ground environment employs two convolution layers and two pooling layers. The size of the convolution filter is 5×5, the first convolution layer use 6 filters, the second convolution layer use 12 filters. Each filter and bias performs initialization with an arbitrary value between 0 and 1. When 28×28 image is input, the first convolution layer performs convolution operation using 6s convolution filters and $24 \times 24 \times 6$ images are output to the first pooling layer. In the first pooling layer, the maximum value of the 2×2 region is taken and $12 \times 12 \times 6$ images are output. The second convolution layer performs convolution operation using 12s convolution filters and $8 \times 8 \times 12$ images are output to second pooling layer. In the second pooling layer, the maximum value of the 2×2 region is taken and $4 \times 4 \times 12$ images are fully-connected to the output node. Each activation function is sigmoid-function.

Using this CNN structures, ground learning model trained 13,900 target and clutter images in the ground environment. As shown in fig. 1 is a part of trained images.

We used 1,090 images set to test the ground learning model. As shown in fig. 2 is a part of test images.

We optimized the weight and bias of the filter by setting the learning epoch value to 10. In order to simplify the matrix vector operation, 50 batch lists were computed simultaneously. The overall structure is shown in fig. 3 and loss-function for the learning result is shown in fig. 4. A loss-function represents the price paid for inaccuracy of predictions in classification problems.

This loss-function is the mean squared error of the estimated data. Mean squared error (E) is given by the expression.

$$E = \frac{1}{2}\left\| y(x) - a \right\|^2 \tag{1}$$

At this expression, $y(x)$ is the expected value, a is the output value of the ground learning model.

After performing 139,000 training sessions, the error rate was 0.05687, and 89 test images were incorrect, and success rate was about 91.83%.

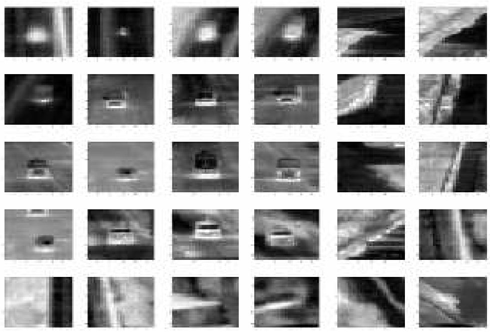


Fig. 1. A part of trained images on the ground environment

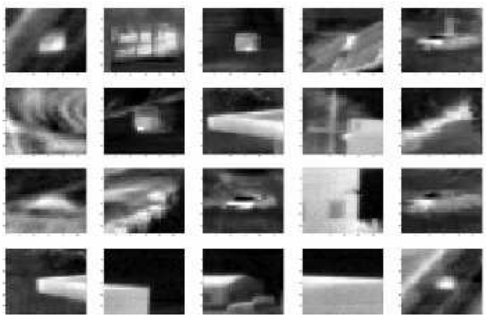


Fig. 2. A part of test images on the ground environment

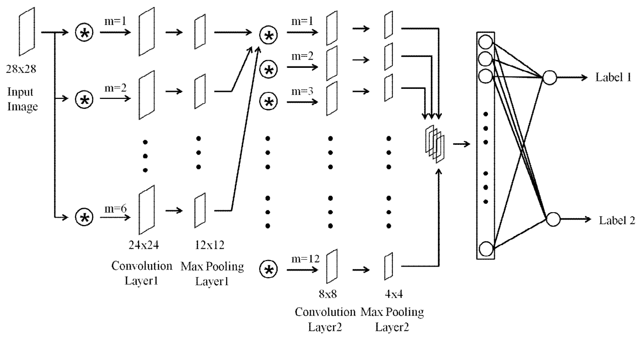


Fig. 3. A Structure of the ground learning model

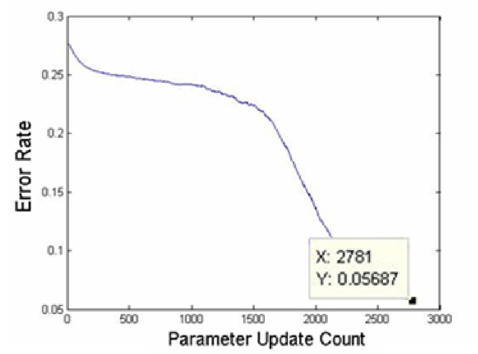


Fig. 4. A loss-function of the ground learning model

2.2 learning and classify on the marine environment

The CNN structure for the marine environment employs two convolution layers and two pooling layers. And unlike the ground learning model, we added a fully-connected layer between the last pooling layer and the output node. The size of the convolution filter is 5×5, the first convolution layer use 6 filters, the second convolution layer use 12 filters. Each filter and bias performs initialization with an arbitrary value between 0 and 1. When 28×28 image is input, the first convolution layer performs convolution operation using 6s convolution filters and $24 \times 24 \times 6$ images are output to the first pooling layer. In the first pooling layer, the maximum value of the 2×2 region is taken and $12 \times 12 \times 6$ images are output. The second convolution layer performs convolution operation using 12s convolution filters and $8 \times 8 \times 12$ images are output to second pooling layer. In the second pooling layer, the maximum value of the 2×2 region is taken and $4 \times 4 \times 12$ images are fully-connected to the output node through the next two fully-connected layers. Each activation function is sigmoid-function.

Using this CNN structure, marine learning model trained 10,000 target and clutter images in the marine environment. As shown in fig. 5 is a part of trained images.

We used 114 images set to test the marine learning model. As shown in fig. 6 is a part of test images.

The output nodes are classified into three categories: positive target, negative target, and clutter. The overall structure is shown in fig. 7 and loss-function for the learning result is shown in fig. 8.

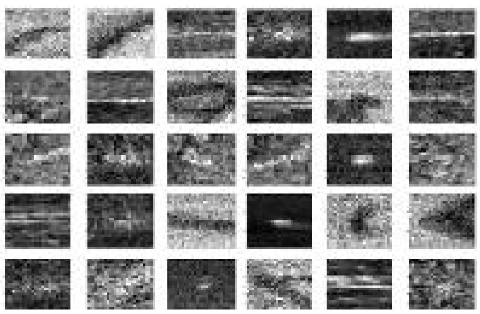


Fig. 5. A part of trained images on the marine environment

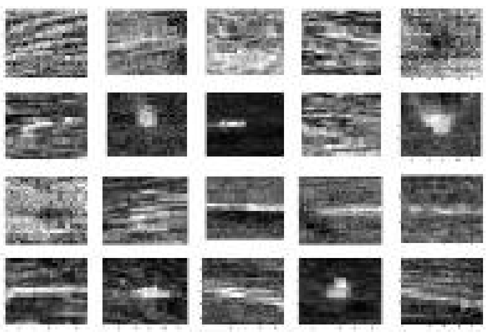


Fig. 6. A part of test images on the marine environment

After performing 100,000 training sessions, the error rate was 0.03078, and 11 test images were incorrect, and success rate was about 90.4%.

2.3 Hybrid machine character method

In Section 2.1 and Section 2.2 demonstrated the classification result of targets and clutters as two learning models for the ground and marine environment. In this result, learning models trained for each environment can classify targets and clutters over 90%.

The control user can use the GPS information to inform the autonomous device whether the current position is the ground or marine environment. However, since it is possible that there is an error between the currently designated location information and the actual infrared image, it is necessary to recognize it automatically. If the autonomous device recognizes the situation opposite to the current environment, the classification rate decreases.

In this paper, two CNN learning model which is learned for the ground and marine environment is considered as multiple personality of machine. And as stated in the Tao Yang's paper [9], each pre-trained learning model will be referred to as the "machinality" of the machine. Also, the machinality suitable for the ground environment is M_g, the machinality suitable for the marine environment is M_s, the overall structure with two machinalities is called the Hybrid Machine Character(HMC) method.

To use the HMC method efficiently, we used another learning model to classify the current environment. The purpose of this learning model is to classify the marine and ground environments and this machinality is M_e .

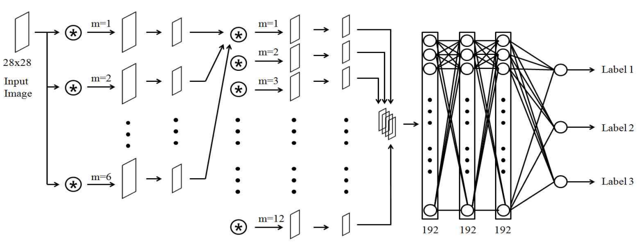


Fig. 7. A Structure of the marine learning model

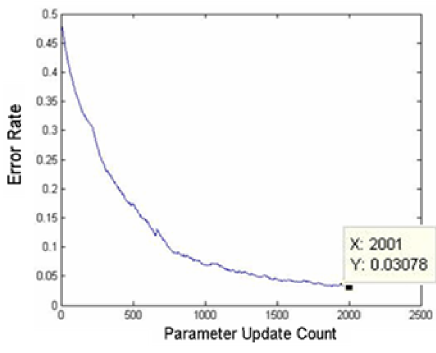


Fig. 8. A loss-function of the marine learning model

This learning model employs three convolution layers and four pooling layers. The size of the convolution filter is 5×5, the first convolution layer use 6 filters, the second convolution layer use 12 filters, the third convolution layer use 24 filters. Each filter and bias performs initialization with an arbitrary value between 0 and 1. And the input image was assumed to be the 120×120 ROI regions, where the target must be found. When the 120×120 images are processed, the first pooling layer is taken the maximum value of the 2×2 region and 60x60 images are output. The first convolution layer performs convolution operation using 6s convolution filters and $56 \times 56 \times 6$ images are output to the second pooling layer. In the second pooling layer, the maximum value of the 2×2 region is taken and $28\times 28\times 6$ images are output. The second convolution layer performs convolution operation using 12s convolution filters and $24\times 24\times 12$ images are output to third pooling layer. In the third pooling layer, the maximum value of the 2×2 region is taken and $12\times 12\times 12$ images are output. The third convolution layer performs convolution operation using 24s convolution filters and $8\times 8\times 24$ images are output to fourth pooling layer. In the fourth pooling layer, the maximum value of the 2×2 region is taken and $4\times 4\times 24$ images are fully-connected to the output node. Each activation function is sigmoid-function.

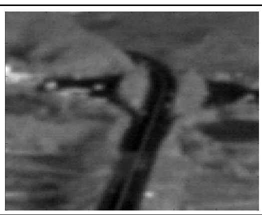
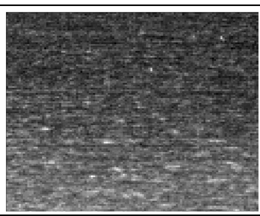
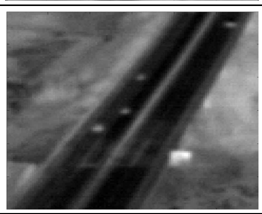
As shown in Table 1 is a part of the ground and marine environment images.

The overall structure is shown in fig. 9 and loss-function for the learning result is shown in fig. 10. After performing 6,400 training set with 15 epoch value, the error rate was 0.0363, and 26 test images were incorrect among 410 test set, and success rate was about 93.65%.

If this learning model classify current environment in the ROI image as the ground environment, HMC use the M_g model that trained for the ground environment. Otherwise, HMC use the M_s model that trained for the marine environment. HMC classify the current environment and then selects the appropriate machine character to classify the targets and clutters. The overall structure is shown in fig. 11.

Fig. 11 shows overall HMC architecture. The first 120×120 ROI images are processed to the M_e model. This machinality functions as to classify ground and marine environment. M_e is composed of 4 max pooling and 3 convolution layers. After pass through these layers, M_e determines whether the 120x120 ROI images are the ground or marine environment. If M_e model's decision is the ground environment, the 28×28 target image is processed to the M_g model. M_g is composed of 2 max pooling and 2 convolution layers. This machinality functions as to classify the targets from the clutters in the ground environment. If M_e model's decision is the marine environment, the 28×28 target image is processed to the M_s model. M_s is composed of 2 max pooling, 2 convolution and 2 fully-connected layers. This machinality functions as to classify the targets from the clutters in the marine environment.

Table 1. A part of the ground and marine environment

Ground environment	Marine environment
	
	

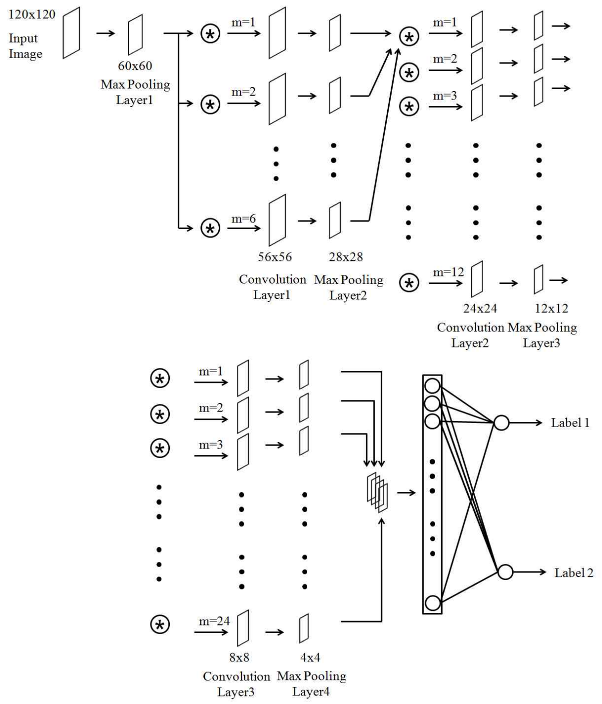


Fig. 9. A Structure of the circumstance learning model

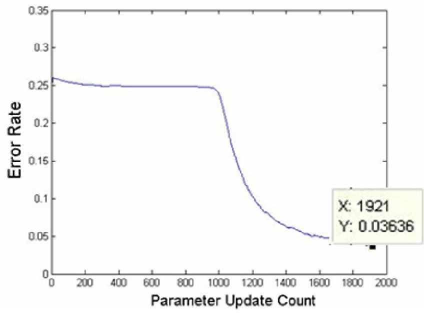


Fig. 10. A loss-function of the circumstance learning model

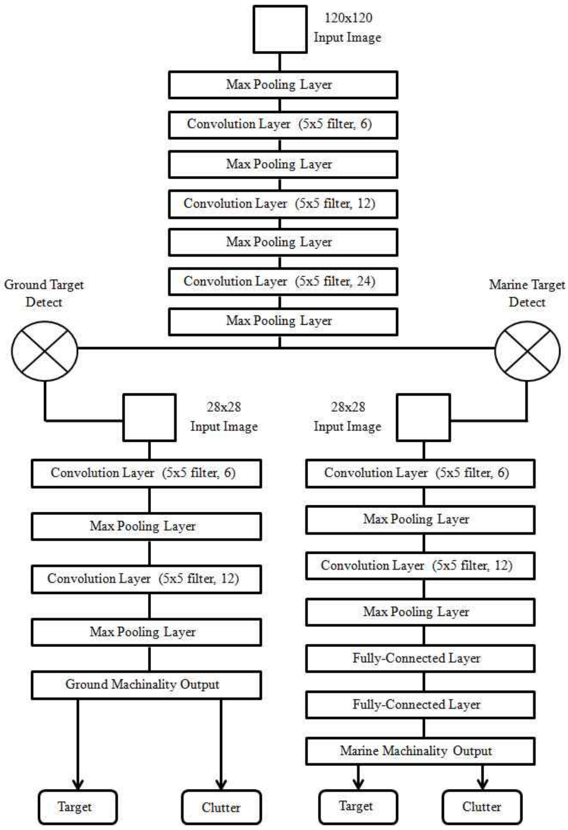


Fig. 11. A Structure of the proposed HMC method

3. Comparison of HMC Method with Others

To compare HMC method with others, first, the whole ground and marine images are only trained by M_g model. Second, the whole ground and marine images are only trained by M_s model. Next, previously proposed research is compared with proposed method. They use 3 convolution layers, 2 sub-sampling layers and 1 fully-connected layer. Finally, average ensemble and max ensemble is compared. After trained the whole images, the loss-function of only trained by M_g is fig.12 and only trained by M_s is fig. 13.

The epoch values of two models are 10. After performing 239,000 training sessions, the error rate of M_g was 0.0087, and success rate was about 64.70%. Also, the error rate of M_s was 0.0467, and success rate was about 72.59%. In the case of previous research, the error rate was 0.0119, and success rate was about 71.51%. To improve the success rate of each learning model, we use the model ensemble. M_g and M_s models have their own output values. Average ensemble calculates the average of each output values. For example, if the target's output value is 0.4220 in the M_g model and 0.1531 in the M_s model, the final output value is 0.2876. On the other hand, max ensemble calculates the max value of each output values. If the target's output value is 0.4220 in the M_g model and 0.1531 in the M_s model, the final output value is 0.4220.

Table 2. A result of comparisons with HMC method

Epoch	Training Set	Machinality	Success Rate
10	23,900	Marine	64.70%
10	23,900	Ground	72.59%
25	23,900	Previous research[4]	71.51%
10	23,900	Average-Ensemble	71.84%
10	23,900	Max-Ensemble	73.09%
(15) + 10	(6,400)+23,900	HMC (no M_e errors)	91.69%
(15) + 10	(6,400)+23,900	HMC (apply M_e errors)	85.42%

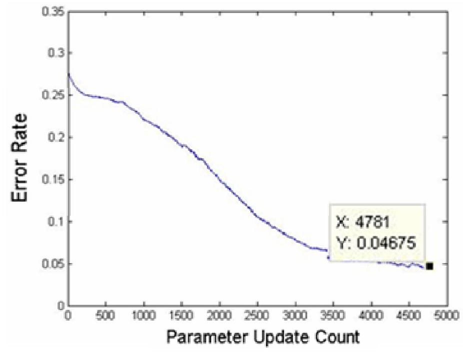


Fig. 12. A loss-functions of the M_g trained whole images

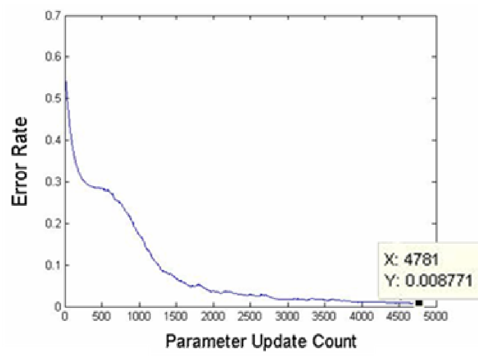


Fig. 13. A loss-functions of the M_s trained whole images

As a result of average ensemble, the success rate was about 71.84% that is lower than the success rate of the M_s solely. Also, as a result of max ensemble, the success rate was about 73.09% that is higher than the each success rate of the M_g and M_s . On the other hand, the proposed HMC method improves the success rate to about 91.69%, assuming that the ground and marine environment are precisely classified by M_e. Even considering the error of M_e , the success rate was about 85.42%, which was better than the other methods.

Table 2 shows the result of comparisons. If the autonomous device apply the proposed HMC method to classify targets and clutters in the infrared image, the performance is better than using the single learning model, previous research and model ensemble.

4. Conclusion

In this paper, we proposed the HMC method that is a new method of applying the advantage of human recognition and management of multiple personality to machine learning. Experiments have shown that autonomous flight devices can improve performance when they applied the proposed HMC method to classify the targets from the clutters in the ground or marine environment. Especially, the proposed HMC method is better than using the single learning model, previous research and model ensemble. Also, it is possible to apply HMC method to various image classification techniques. It is possible to obtain more useful results by changing two machinality consciously according to the current situation or position such as the weather, surrounding obstacles and altitude. If we continuously develop the proposed HMC method, we can expect the extension of the diversity and expertise of the machine learning algorithm.

References

[1] H. Zhang, Z. Zhao and F. Xiao, "Robust Detection Method of Small Targets in Sea-Clutter via Improved Fast Clustering Segmentation," 10.1109/IHMSC.2016.236, pp. 123-126, 2016

[2] C. Yang, H. Liu, S. Liao and S. Wang, "Small target detection in infrared video sequence using robust dictionary learning," *Infrared Physics & Technology*, vol. 68, pp. 1-9, 2015

[3] S. Qi, J. ma, H. Li, S. Zhang and J. Tian, "Infrared small target enhancement via phase spectrum of Quaternion Fourier Transform," *Infrared Physics & Technology*, vol. 62, pp. 50-58, 2014.

[4] Ødegaard, N., Knapskog, A. O., Cochin, C., Louvigne, J. C. "Classification of ships using real and simulated data in a convolutional neural network," Rader Conference, 2016.

[5] Hiroshi Tasaka, "Everyone has multiple personality," Kobunsha Co. Japan, 2015.

[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv, 2014.

[7] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolution Neural Networks," In Pro. NIPS, 2012.

[8] S. Bell, P. Upchurch, N. Snavely and K. Bala, "Material recognition in the wild with the materials in context database," CoRR, abs/1412.0623, 2014.

[9] Tao Yang, "Application of Computational Verbs to Feeling Retrieval from Texts," *International Journal of Computational Cognition*, vol. 4, no. 3, pp. 28-45, 2006.

Ju-Young Lee He received B.S degree in computer science from Seokyeong University, S. Korea, in 2011, the M.S. degree in IT Media Engineering from Seoul National University of Science and Technology, S. Korea, in 2013. Since 2013, he has been a member of Agency for Defense Development, S. Korea. His main research interests include image processing, IT media convergence technology, machine learning, and pattern classification.

Jae-Wan Lim He received the B.S. degree in electronics in 1987, the M.S. degree in electronics in 1989, all from ChungNam National University, S. Korea. Now he is a responsible researcher in Agency for Defense Development, S. Korea. His research interests include image processing, object tracking, pattern classification and system design for automatic target recognition & tracking.

Eun-Jin Koh He received his BS and MS degree in computer science from Inha University, S. Korea, in 2005 and 2007, respectively and received his PhD degree from University of Science and Technology, S. Korea, in 2011. Since 2011, he has been a member of Agency for Defense Development, S. Korea. His main research interests include image processing, human-computer interaction, object tracking, and pattern classification.