Journal of KIIT. Vol. 16, No. 9, pp. 85-94, Sep. 30, 2018. pISSN 1598-8619, eISSN 2093-7571 85
http://dx.doi.org/10.14801/jkiit.2018.16.9.85

Multi-scale Pedestrian Detection in Thermal Imaging Using Deep Convolutional Neural Network and Adaptive NMS

Tan Dat Trinh*[1], Jin Young Kim*[2]

"This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-R2718-16-0011) supervised by the IITP(Institute for Information & communications Technology Promotion)"

Abstract

In this paper, we propose a new method for pedestrian detection in thermal image/video by improving the non-maxima suppression(NMS) algorithm. We apply the sliding window detector based deep convolutional neural networks(DNN) to extract pedestrian candidates. A sliding window combined with image pyramids is used to identify pedestrians at varying scales and locations in the image via Convolutional Neural Network(CNN) based binary classification. To improve the performance, we propose an adaptive NMS algorithm to remove false alarms. The proposed NMS uses adaptive overlap-thresholds to overcome the drawback of the standard NMS and improve performance of detection system. It is automatically adjusting the overlap-thresholds based on the density of overlapping windows to improve the accuracy of system. Pedestrian detection experiment results with our thermal database and OSU thermal pedestrian database confirm that the proposed method outperforms the baseline method.

요 약

본 논문에서는 비최대억제(NMS) 알고리듬의 성능 개선을 통해 열화상(동영상)에서 보행자 탐지를 위한 새로운 방법을 제안한다. 보행자 후보들을 추출 하기 위하여 슬라이딩 창기반의 딥 컨볼루션 신경망을 사용한다. 영상 피라미드와 결합된 슬라이딩 창이 CNN 기반 이진분류를 통해 영상내 다양한 크기와 위치의 보행자를 식별한다. 보행자탐지 성능 향상을 위하여, 오경보를 제거하기 위해 적응적 NMS 방법을 제안한다. 제안한 NMS는 표준 NMS의 단점을 극복하고 탐지 시스템의 성능 향상을 위하여 적응적 중첩-임계값을 이용한다. 시스템의 정확도를 향상하기 위해 중첩된 창들의 밀도에 기반하여 자동적으로 중첩-임계값을 조절한다. 자체 열화상 영상 데이터베이스와 OSU 열상 보행자 데이터베이스를 대상으로 한 실험결과 제안한 방법이 기존의 베이스라인 방법에 비해 우수함을 확인하였다.

Keywords
human/pedestrian detection, thermal imaging, sliding window, DNN, adaptive NMS

* Dept. of Electronics and Computer Eng. Chonnam National University
- ORCID[1]: http://orcid.org/0000-0003-0043-612X
- ORCID[2]: http://orcid.org/0000-0002-4986-8980

· Received: Jun. 23, 2018, Revised: Sep. 04, 2018, Accepted: Sep. 07, 2018
· Corresponding Author: Jin Young Kim
Dept. of Electronics and Computer Eng, Chonnam National University
Tel.: +82-62-530-1757, Email: beyondi@jnu.ac.kr

Ⅰ. Introduction

Human/pedestrian detection is one most the important task in computer vision, surveillance and monitoring systems[1][2][3]. The most difficult problem is to detect and identify pedestrian activity in continuous 24/7 operability[4]. Visual camera is able to set up during daytime operation. But it has obvious limitations during night-time or dark environments. Otherwise, the visual camera is highly sensitive to illumination changing.

Thermal camera can help solving the illumination problems, and thermal imaging using infrared camera can be able to enhance the visibility of pedestrians in dark environments by detecting the amount of infrared radiation emitted/ reflected from pedestrians in the scene[5]. So, the thermal camera is expected to operate at all times. Thermal imaging systems for real-time applications are applied not only in military but also in industrial, and commercial applications. Thermal imaging processing still is difficult task due to some challenging problems: lower signal-to-noise ratio, polarity inversion, reflection, and halo effects[6].

Human/pedestrian detection can be considered as a two-stage approach[6][7]. Human candidate extraction is the first stage, the second is pedestrian classification. For each stage, various methods have been proposed [1]-[9]. Sliding window detector, human segmentation and background subtraction can be applied for the first stage. The sliding window detector is simple and fast, but it requires huge computation cost. The segmentation approach can be faster by reducing the number of analyzed regions but it depends on the assumption that pedestrian object is much warmer/hotter than the background and the contrast between the pedestrian and the background is much high. So, this technique is unable to detect a cooler pedestrian than the background (i.e., in summer/ day-time). The background subtraction technique requires a fixed-camera and is very sensitive to background or non-pedestrian changing (i.e., tree leaf waving, rippling water, extreme sudden illumination changes or car moving). It doesn't consider the background hot spots as foreground regions. However, it needs substantial computational and memory resources.

The sliding window detector is one of the most popular methods for pedestrian detection in both visual and thermal camera image processing[10-12]. In this approach, the multi-scale sliding window based on pyramid images is first applied to get all image patches. Then, the classifier independently identify all image patches as pedestrian or non-pedestrian. In [8], authors used histograms of oriented gradients (HOG) as features that were extracted from each window, and support vector machine (SVM) was applied as a classifier. Finally, they used a non-maxima suppression (NMS) approach to reduce the redundant windows and obtained detection results. I. Riaz, et. al. [11] applied census transform histogram features and SVM to detect human in thermal image. J. Baek, et. al. [12] developed a new feature called the thermal-position-intensity-histogram of oriented gradients and used kernel SVM for pedestrian detection at nighttime. Their method performs better than the baseline method in terms of detection performance.

In this study, we apply multi-scale sliding window detector based on deep convolutional neural network (DNN)[13][14] for pedestrian detection in thermal image. First, a sliding window combined with image pyramids is applied extract all candidates at varying scales and locations. Second, the DNN independently classifies all candidates as human or non-human. Finally, we apply a post processing approach based on kd-tree searching and the proposed non-maxima suppression to remove false alarms and enhance the performance of detection system. Our NMS method is developed by automatically adjusting the overlap-thresholds based on the density of overlapping windows to improve the accuracy of system.

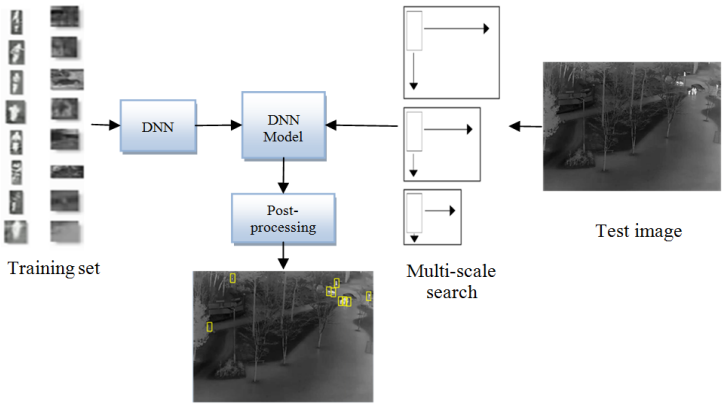


Fig. 1. Pedestrian detection system using sliding window detector based the DNN

II. Overview of Proposed System

We apply the multi-scale sliding window detector based on DNN for pedestrian detection in thermal image/video. The method includes two processes: training and testing. In the training stage, we train a binary classifier via the DNN using positive and negative samples. In the test stage, the multi-scale sliding window is first applied to get all image patches. Then, the DNN independently classifies all image patches as being pedestrian or non-pedestrian. Fig. 1 shows an overview of the pedestrian detection system.

2.1 Sliding Window Detector

In this study, a sliding window combined with image pyramids is applied at varying scales and locations to detect pedestrians in thermal image.

The size of the window is set as 32×16 pixels; it scans given image with stride of 4×4. Once scanning is done, the input image is shrunk in size with a fixed scale of 1/2, and the detector scans through with a shrunk image in the same way as above.

In a real application, pedestrians have different positions and sizes. Therefore, if we use only one detection window with fixed size, the detector would fail to find all pedestrians in an frame. One way to solve this problem is scanning the image with a detection window at all positions and scales. So, a sliding window combined with image pyramids can be applied recognize pedestrians at varying scales and locations. Fig. 2 shows an example of sliding window detector. In experiments, there are totally three scales of images using our dataset.

2.2 Binary Classifier via Deep Convolutional Neural Network

A binary classification of the pedestrians was performed using a trained by the DNN[14]. The DNN can jointly learn the discriminative features and the binary classification model. The input layer consists of thermal image pixels with the size 32×16. Training samples of different sizes are resized before being given as the input to the DNN. The first layer consists of a convolutional layer with 32 filters of size 3×3, stride of 1×1 with relu activation function, maximum pooling with filter size 2×2. The second

layer consists of a convolutional layer with 64 filters of size 3×3, stride of 1×1 with relu activation function, maximum pooling with filter size 2×2. The third layer consists of a convolutional layer with 128 filters of size 3×3, stride of 1×1 with RELU activation function and a maximum pooling with filter size 2×2. The fourth layer consists of the fully connected network with 1024 neurons with relu activation function and dropout ratio of 0.5. The output layer consists of two output neurons, with softmax function, providing a score for each class. Using the class scores, each sample is classified as either pedestrian or non-pedestrian, by assigning the label of the class with the highest score. The DNN architecture is shown in Fig. 3.

In this study, we train the network using 186400 samples that includes 22290 positive samples and 164110 negative samples. We apply batch size of 100 and 15 epochs to train our network based on Adam optimizer.

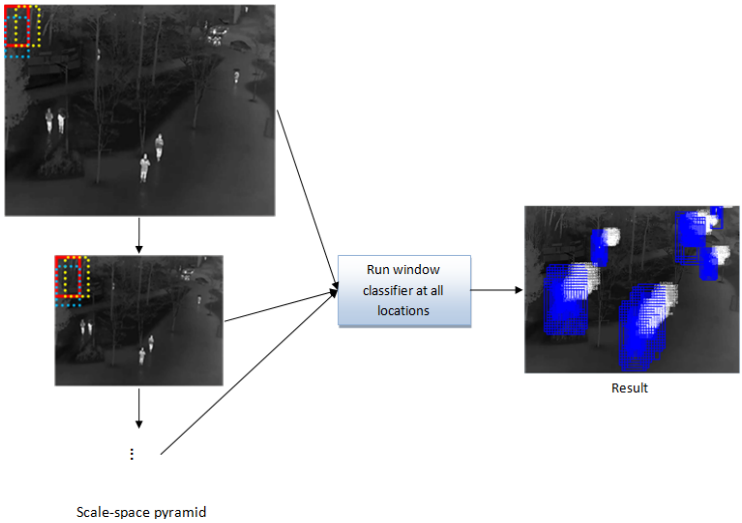


Fig. 2. Sliding window detector

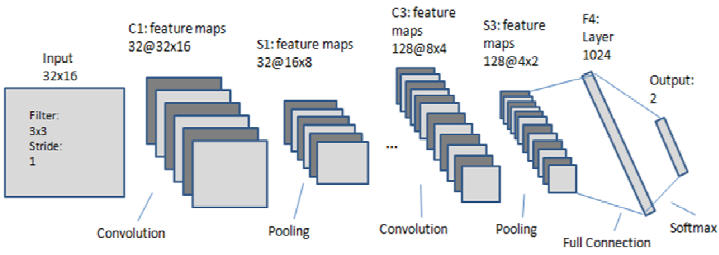


Fig. 3. DNN architecture

2.3 Post Processing and Baseline NMS

We propose a post-processing approach to remove miss-classification via sliding window detector by using clustering based on kd-tree searching shown in Algorithm 1. Goal of this method is reduce miss classification while preserving windows that cover object regions.

Algorithm 1. Reducing miss classification based on kd-tree searching

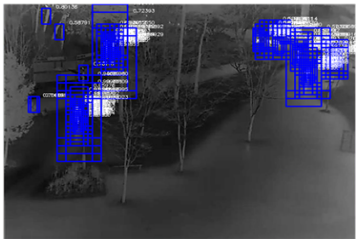
Input: Given a set S_C of windows (bounding boxes) after using sliding window detector.

- Step 1: clustering bounding boxes S_C based on their centers using kd-tree searching. A cluster includes points that lie into a sphere with radius of 8.
- Step 2: remove clusters if number of points in the clusters is smaller than 8.

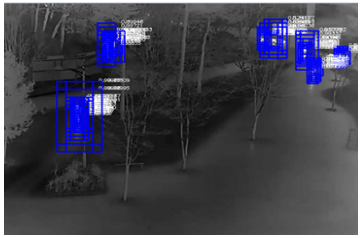
Fig. 4 shows an example of reducing miss classification based on kd-tree searching. Finally, we apply the non-maxima suppression (NMS) [10] to remove redundant windows and get detection results.

The approach based on sliding windows typically produces multiple windows with high scores close to the correct location of objects. The number of window hypotheses at this step is simply un-correlated with the real number of objects in the image. The goal of non-maxima suppression [10] is therefore to retain only one window per group, corresponding to the precise local maximum of the response function, ideally obtaining only one detection per object. The brief procedure of standard NMS is described in Algorithm 2.

The standard NMS is popular because it is conceptually simple, fast, and for most tasks results in satisfactory detection quality. Despite its popularity, the standard NMS has a drawback. It trades off precision vs. recall. If the overlap threshold is too big (too strict), then not enough surrounding detections are suppressed, high scoring false positives are introduced and precision suffers; if the overlap threshold is too low (too loose) then multiple true positives are merged together and the recall suffers (shown in Fig. 5).



(a) Results using sliding window detector

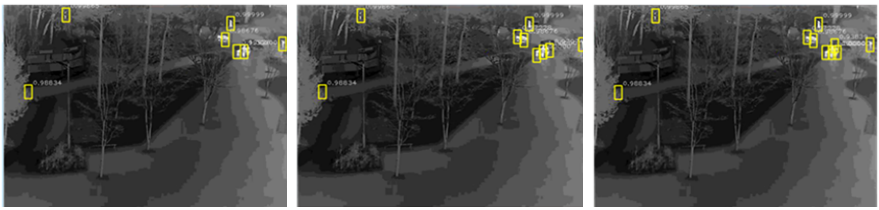


(b) Results after post processing

Fig. 4. Example of reducing false alarm using clustering based on kd-tree searching

Algorithm 2. Standard NMS

- Step 1: Sort all detection results based on DNN scores.
- Step 2: For detection result with the highest score, find overlapped results of the selected detection result. Overlapped detection results are found when the ratio of intersection area ($S_1 \cap S_2$) over area of union ($S_1 \cup S_2$) is larger than a fixed overlap threshold value.
- Step 3: Among overlapped detection results, select the one with the highest score and remove others.
- Step 4: Repeat step 1-3.



(a) T=0.02 (b) T=0.06 (c) T=0.2

Fig. 5. Example of results of standard NMS with different threshold T.

For any overlap threshold, standard NMS is sacrificing precision or recall. The drawback of standard NMS is shown in Fig. 5. When we select small threshold T of 0.02, there is a miss detection window. Otherwise, we seletect large threshold T of 0.2, there are more false alarms. In our experiments, we choose T of 0.06 that shows better performance.

III. Proposed Adaptive NMS

In this study, we propose an adaptive NMS using adaptive overlap threshold to overcome the drawback of the standard NMS and improve performance of detection system. The proposed NMS is automatically adjusting the overlap-thresholds based on k-means clustering and the density of overlapping windows in each cluster to improve the accuracy of system.

Let S_B be a set of M bounding boxes (windows) after post processing (after reducing miss classification using kd-tree searching) and N_B be a set of N bounding boxes. Let L be number of overlapping regions between bounding boxes in N_B. The proposed adaptive NMS method is shown in Algorithm 3.

To cluster the set of bounding boxes S_B using k-means, we first to estimate number of cluster k. In this study, we propose a heuristic-based method to determine k based on number of bounding boxes and number of overlapping regions between bounding boxes as shown in Step 1(c). Based on experiments, we realize that after applying the standard NMS using threshold of 0.06 in S_B, there are overlapped bounding boxes that cover the same object in N_B. So, to determine a suitable number of clusters for k-means algorithm, we need to subtract a compensation factor from number of overlapping regions L in N_B.

Algorithm 3. Proposed adaptive NMS.

Input: Given a set of S_B.

Step 1: Clustering bounding boxes into S_B using k-means.

a) Apply the standard NMS to S_B to obtain N_B.

b) Find number of overlapping regions L in N_B.

c) Number of cluster k is determined by

$$\begin{cases} k = N - 2L, & \text{if } N > 2L \\ k = N - L, & \text{if } N \le 2L \wedge N \neq L \\ k = N - L + 1, & \text{if } N = L \end{cases}$$

d) Apply k-means to S_B to get k groups of bounding boxes $\{G_1, \ldots, G_k\}$.

Step 2: Find adaptive overlap threshold T_o and apply NMS for each group G_i.

For each group $G_i (i = 1, \ldots, k)$

{ a) Let n_i be number of windows in group G_i. Calculate density of windows, $d_b = \frac{n_i}{M}$.

b) If $(d_b > 0.3)$ $T_o = 0.06 + \frac{d_b}{10}$

else $T_o = 0.06$

c) Apply the standard NMS to G_i using overlap-threshold T_o.

}

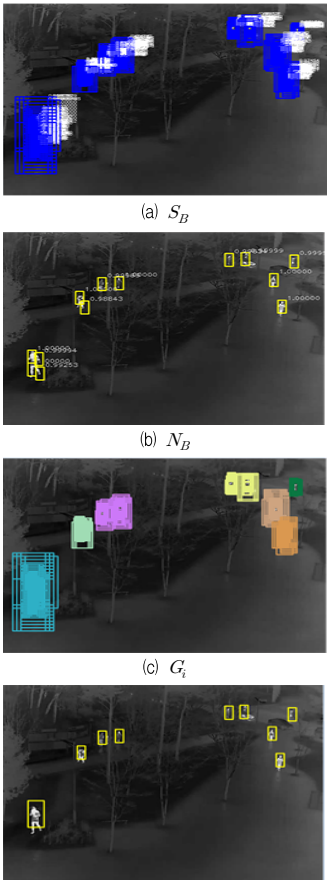
We use density of bounding boxes in each group to estimate adaptive overlap threshold T_o. We realize that if the overlap threshold is too large (too strict) then not enough surrounding detections are suppressed, high scoring false positives are introduced and precision suffers; if the overlap threshold is too low (too loose) then multiple true positives are merged together and the recall suffers. Based on experiments, we set overlap threshold T_o equal to 0.06 if the density of windows in each group d_b is small enough (< 0.3) to suppress enough surrounding bounding boxes. Otherwise, we increase the threshold T_o by a ratio $\frac{d_b}{10}$ to avoid missing objects or several objects can be considered one. Fig. 6 shows an example of results of the adaptive NMS method.

IV. Experiments and Results

4.1 Dataset

We evaluate the performance of pedestrian detection using our dataset and public dataset (OSU thermal pedestrian database [4]). Our dataset is captured by a Argo S Thermal Imaging Camera with 640×480 size thermal videos at the frame rate of 31 per second. In this study, we assume that the camera position is fixed. Our camera is placed at the height of about 10~12m from the ground (in the third floor of the building). We recorded the database at different locations within Chonnam National University (CNU), Gwangju campus. The size of moving objects is different. Some objects is very small, the other is larger. All videos are recorded in outdoor environments, in summer (July, August, September), when the temperature on average ranges between 18°C ~ 30°C. The setup is shown in Fig. 7.

We recorded 10 thermal videos in various conditions. Based on video frames, we manually segmented pedestrians and background to generate the training dataset. In this study, we train the network using 186400 samples that includes 22290 positive samples (pedestrians) and 164110 negative samples. (background).



(a) S_B

(b) N_B

(c) G_i

(d) Result

Fig. 6. Example of results of the adaptive NMS method

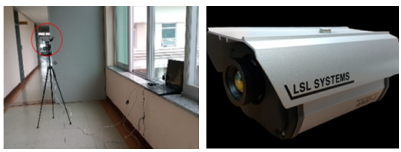


Fig. 7. Camera is used for the experiments.

Table 1. Performance comparison of pedestrian detection on our database

Data		TP	FA	MD	Precision	Recall
Seq1	Baseline	138	97	22	58.72%	86.25%
	Proposed	144	27	16	84.21%	90%
Seq2	Baseline	116	207	3	35.91%	97.47%
	Proposed	117	45	2	72.22%	98.31%

(a) Baseline NMS

(b) Proposed NMS

Fig. 8. Example of results on our dataset

In our experiment, we randomly select two test sequences from a test thermal video. Each sequence include 20 frames to evaluate the performance. Sequence 1 (Seq1) includes 160 pedestrians and sequence 2 (Seq2) includes 119 pedestrians. Fig. 8 shows an example of results on our dataset.

Table 1 shows the performance comparison of proposed method and baseline method on our database.

From this table, we realize that our post-processing approach and the proposed adaptive NMS can reduce lots of false alarms and improve the performance of detection system.

We also evaluate the performance of our method on the OSU thermal pedestrian database. We do experiments on 10 sequences from this data. Here, one-scale sliding window detector is applied. Table 2 shows the performance on the OSU dataset using the baseline NMS. Totally, the average precision is of 78.48% and average recall is of 95%. The performance on the OSU dataset using the proposed adaptive NMS is described in Table 3. Our method can obtain the average precision is of 84.55% and average recall is of 94.67%. The proposed method can significantly reduce a large of false alarms and improve the precision rate of 84.55% (on average) in comparison to the baseline method (78.48% on average). So, the proposed method can enhance the performance of pedestrian detection system.

Table 2. Performance of pedestrian detection on OSU thermal pedestrian database using the baseline NMS

Data	TP	FA	MD	Precision	Recall
seq1	81	0	10	100%	89.01%
seq2	100	27	0	78.74%	100%
seq3	88	54	13	61.97%	87.12%
seq4	109	24	0	81.95%	100%
seq5	100	88	1	53.19%	99.01%
seq6	91	12	6	88.34%	93.81%
seq7	78	0	16	100%	82.97%
seq8	99	4	0	96.11%	100%
seq9	95	9	0	91.34%	100%
seq10	92	6	5	93.87%	94.84%

Table 3. Performance of pedestrian detection on OSU thermal pedestrian database using the proposed approach

Data	TP	FA	MD	Precision	Recall
seq1	83	7	9	92.22%	90.02%
seq2	100	29	0	77.51%	100%
seq3	90	87	11	50.84%	89.10%
seq4	109	41	0	72.66%	100%
seq5	101	93	0	52.66%	100%
seq6	93	16	4	85.32%	95.87%
seq7	79	0	14	100%	84.94%
seq8	98	7	1	93.33%	98.98%
seq9	95	23	0	80.50%	100%
seq10	90	22	8	80.35%	91.83%

However, our method still has some drawbacks. When the background objects have same intensity as pedestrians, they are classified as pedestrian label. That leads to reduce the precision rate. Otherwise, when the pedestrians have same intensity as background, they are considered as background. That leads to reduce the recall rate. These drawbacks will be considered as our future works. Fig. 9 shows examples of drawbacks of our method in shown in the ellipse regions

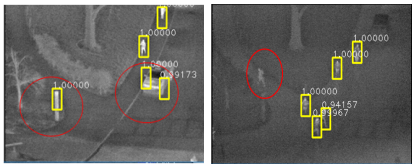


Fig. 9. Drawbacks of our method

V. Conclusion

In this paper, we applied a new approach for pedestrian detection from thermal image, where sliding window approach and deep convolutional neural networks were adopted. To improve the accuracy, we proposed a post processing approach with kd-tree searching and adaptive NMS method to remove false alarms. According to the human object experiments for thermal images, the precision and recall rates were enhanced by 30.9 and 2.3% respectively. In future, we will focus on fast implementation of the developed system.

References

[1] M. Paul, S. M. Haque, and S. Chakraborty, "Human detection in surveillance videos and its applications-a review", EURASIP Journal on Advances in Signal Processing, Vol. 1, No. 176, pp. 1-16, Dec. 2013.

[2] Ahra Jo, Jeong-Sik Park, Yong-Ho Seo, and Gil-Jin Jang, "Performance Improvement of Human Detection in Thermal Images using Principal Component Analysis and Blob Clustering", Journal of the Institute of Internet, Broadcasting and Communication, Vol. 12, No. 2, pp. 157- 162, Apr. 2013.

[3] Seok-Gyu Choi and Wenjie Xu, "A Study on Person Re-Identification System using Enhanced RNN", Journal of the Institute of Internet, Broadcasting and Communication, Vol. 17, No. 2, pp. 25-31, Apr. 2017.

[4] J. W. Davis and V. Sharma, "Background-subtraction in thermal imagery using contour saliency", International Journal of Computer Vision, Vol. 71, No. 2, pp. 161-181, Feb. 2007.

[5] J. W. Davis and V. Sharma, "Robust detection of people in thermal imagery", IEEE Proceedings of the 17th International Conference on Pattern Recognition, Vol. 4, pp. 713-716, Aug. 2004.

[6] E. Goubet, J. Katz, and F. Porikli, "Pedestrian tracking using thermal infrared imaging", In Proceedings of SPIE, Vol. 6206, pp. 797-808, Apr. 2006.

[7] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery", IEEE Workshops on Application of Computer Vision, Vol. 1, pp. 364-369, Jan. 2005.

[8] E. S. Jeon, et. al., "Human detection based on the generation of a background image by using a far-infrared light camera", Sensors, Vol. 15, pp. 6763-6788, Mar. 2015.

[9] M. Teutsch, T. Muller, M. Huber, and J. Beyerer, "Low resolution person detection with a moving thermal infrared camera by hot spot classification", IEEE Conference on Computer Vision and Pattern Recognition Workshops, Jun. 2014.

[10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 886-893, Jun. 2005.

[11] I. Riaz, J. Piao, and H. Shin, "Human detection by using centrist features for thermal images", International Journal on Computer Science and Information Systems, Vol. 8, No. 2, pp. 1-11, 2013.

[12] J. Baek, S. Hong, J. Kim, and E. Kim, "Efficient pedestrian detection at nighttime using a thermal camera", Sensors, Vol. 17, No. 8, pp. 1-20, Aug. 2017.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton "ImageNet classification with deep convolutional neural networks", Proceedings of the 25th International Conference on Neural Information Processing Systems, Vol. 1, pp. 1097-1105, Dec. 2012.

[14] K. Simonyan and A. Zisserman "Very deep convolutional networks for large-scale image recognition", Proceedings of ICLR, pp. 1-14, May 2015.

Authors

Tan Dat Trinh

2010 : B.S. degree in Department of Mathematics and Computer Sciences from HCMC University of Natural Sciences

2013 : M.S. degree in Department of ECE, Chonnam National University

2014 ~ now : Ph.D. course in Department of ECE, Chonnam National University

Research interests : Audio signal processing, Image Processing, Computer Vision and Pattern Recognition

Jin Young Kim

1986 : B.S., 1988 : M.S., 1994 Ph.D. degrees in Department of Electronic Engineering, Seoul National University

1995 ~ now : Professor in Department of Electronics and Computer Engineering, Chonnam National University

Research interests : Audio-Visual signal processing and embedded system.