

A Study on Visual Saliency Detection in Infrared Images Using Boolean Map Approach

Mai Thanh Nhat Truong* and Sanghoon Kim*

Abstract

Visual saliency detection is an essential task because it is an important part of various vision-based applications. There are many techniques for saliency detection in color images. However, the number of methods for saliency detection in infrared images is limited. In this paper, we introduce a simple approach for saliency detection in infrared images based on the thresholding technique. The input image is thresholded into several Boolean maps, and an initial saliency map is calculated as a weighted sum of the created Boolean maps. The initial map is further refined by using thresholding, morphology operation, and a Gaussian filter to produce the final, high-quality saliency map. The experiment showed that the proposed method has high performance when applied to real-life data.

Keywords

Boolean Map, Infrared Image, Saliency Detection

1. Introduction

Visual saliency is a subjective perceptual quality that makes some objects of an image appear distinct from their surroundings and attract viewers' attention. Visual saliency detection is a classification task on the pixel level, in which each pixel is assigned a saliency score to determine whether it belongs to the salient objects or to the background. As visual saliency models seek to represent the human attention mechanism, visual saliency research plays an important role in computer vision. Moreover, visual saliency detection is an essential part in various vision-based applications such as content-aware image manipulation [1], video compression [2], image segmentation [3], and object recognition [4].

Saliency detection techniques can be classified into two major categories: bottom-up and top-down. Bottom-up approaches utilize low-level features such as color, texture, orientation; while top-down approaches depend on high-level factors such as prior knowledge regarding tasks or events. In the beginning, saliency detection algorithms were mainly based on bottom-up approaches. Recently, machine learning techniques have been extensively used in computer vision [5,6], furthermore, they become an important part in saliency detection systems that use top-down approaches. Borji et al. [7] performed evaluations on 41 state-of-the-art models using seven data sets. The report provided an in-depth analysis for the development of saliency research over the last few years. Another widely used benchmark for the

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received October 1, 2018; accepted June 29, 2019.

Corresponding Author: Sanghoon Kim (kimsh@hknu.ac.kr)

* Dept. of Electrical, Electronic, and Control Engineering, Hankyong National University, Anseong, Korea (mntnuong@toantin.org, kimsh@hknu.ac.kr)

evaluation of saliency detection methods is the Massachusetts Institute of Technology (MIT) Saliency Benchmark [8], which contains the results from over 70 models.

In addition to color or grayscale images, which are acquired from conventional imaging devices, there are infrared images representing specialized visual data. These images have various applications in critical fields such as security, surveillance, military [9], or medicine [10]. Despite the fact the number of saliency detection algorithms is high and continues to grow, techniques applicable specifically for saliency detection in infrared images are not as common as those applicable to color images. Saliency detection algorithms that work well with color images may result in poor performance when applied to infrared images. This is because infrared images are usually highly noisy and are of low contrast. Moreover, they have low resolution and insufficient features. As shown in Fig. 1, the detail of an infrared image (acquired from [11]) is insufficient even for humans to recognize salient objects. In Fig. 2, a high-performance saliency detection algorithm [12] was applied to an infrared image. This algorithm was originally designed for three-channel color images. As we can see from the result, this algorithm was not able to achieve acceptable result when being applied to the infrared image.



Fig. 1. The lack of visual detail in infrared images. Adapted from [11].

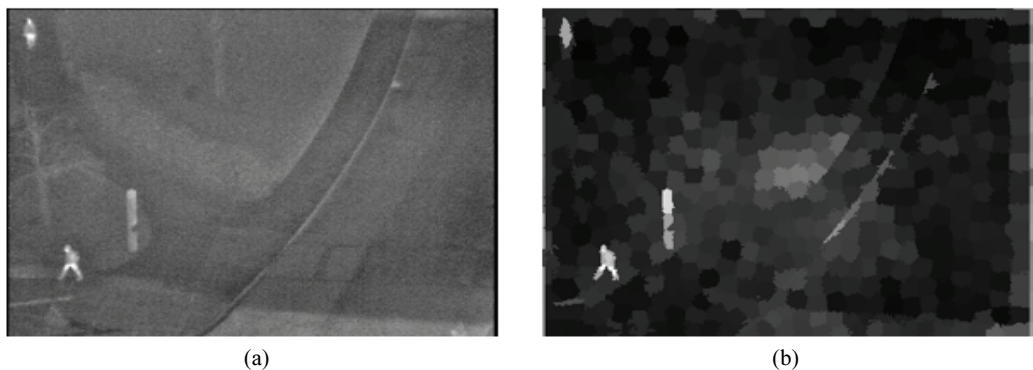


Fig. 2. Algorithm designed for color images is not effective when being applied to infrared images. (a) Input infrared image (b) Result from algorithm [12].

Several attempts have been made to overcome the difficulties discussed above. Li et al. [13] used the Gaussian scale-space representation to compute two types of saliency maps, namely, a dark saliency map, and a bright one. Then signal theory was used to compute phase information for the purpose of detecting

both bright and dark regions of an image in question. These regions were finally grouped, and a shape matching algorithm was applied for salient object detection. In [14], the authors first improved the quality and contrast of an image in a frequency domain; then, they combined luminance distribution and the gradient feature to expose the objects that have large gradient and compact distribution. The resulting saliency map for the considered infrared image was constructed by integrating these two features. Qin et al. [15] took advantage of the human attention mechanism and the information theory. They segmented the input image into regions, then calculated the contrast of each region and the spectral residual of the whole image. The final saliency image was obtained by combining the spatial correlation and spectral residual factors retrieved using spatial distance information.

In this research, we utilize Boolean maps [16] to overcome the problems of high noise and low contrast in infrared images. A Boolean map is a spatial representation that divides an image into two complementary components: the component that is selected and the component that is not selected. Zhang and Sclaroff [17] proposed a saliency detection algorithm based on this concept (namely, Boolean Map Saliency [BMS]). However, it produced imprecise results when applied to infrared images (Fig. 3). More details on the experiment and analysis will be discussed in the following sections.

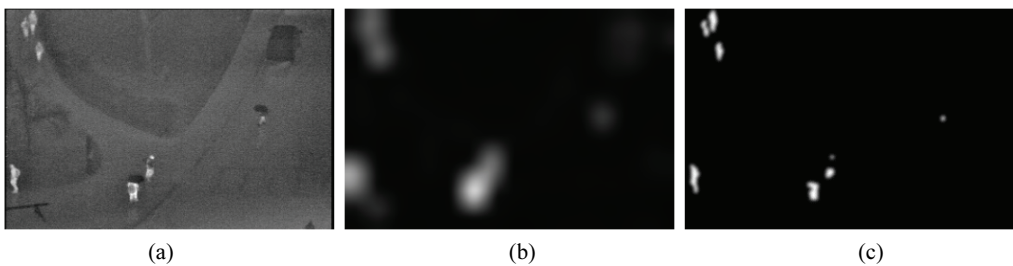


Fig. 3. Saliency detection results: (a) the input image, (b) the resulting image obtained using BMS, and (c) the resulting image obtained through the proposed method.

Saliency detection in infrared images is a challenging problem due to the nature of infrared images. In our proposed method, we utilized several techniques to overcome the existing difficulties. By constructing multiple Boolean maps, we aimed to expose image features in different threshold levels to solve the low-contrast problem. Combining this method with an adaptive thresholding algorithm, we suppressed the noise in infrared images while preserving their visual details. The complete algorithm is described in detail in Section 2. Section 3 demonstrates the performance of the proposed method and provides the results of comparison against the BMS approach. This paper is concluded with Section 4.

2. Methodology

A Boolean map is a spatial representation that divides a visual scene, usually an image, into two complementary components: the component that is selected and the component that is not selected. This theory is an interpretation of the human attention mechanism proposed by Huang and Pashler [16]. The key principle of this theory is that the visual attention of human is limited and it can only capture one object at a time. As illustrated in Fig. 4, an image with two objects is segmented into two Boolean maps showing the mechanisms of human attention.

Boolean maps were first applied for the purpose of saliency detection by Zhang and Sclaroff [17]. In this algorithm, a set of binary images is generated from a given image by thresholding different color channels of the image randomly. Then a saliency map is computed by discovering surrounded regions via the topological analysis of Boolean maps (Fig. 5). The advantage of this algorithm is the ability to achieve superior performance while retaining simplicity. However, as shown previously in Fig. 3, this algorithm does not produce acceptable results when being applied to infrared images. In the following sections we describe our approach to utilize Boolean maps theory for saliency detection in infrared images. The overall flowchart of the proposed system is illustrated in Fig. 6.

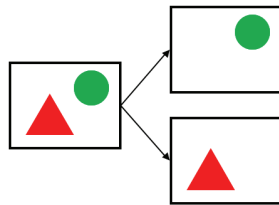


Fig. 4. An illustration of Boolean maps.



Fig. 5. Saliency detection using Boolean maps approach. Adapted from [17].

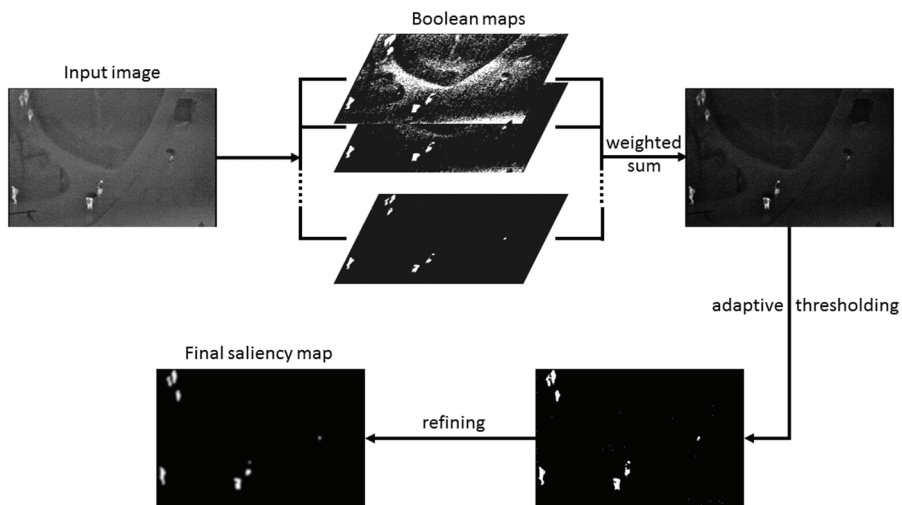


Fig. 6. The flow chart of the proposed method.

Given an infrared image I in which pixel values lie in the interval $[0, 255]$, in the first step, $N - 1$ Boolean maps B_i are calculated as follows:

$$B_i = Th(I, \theta_i), \quad i \in \{1, 2, \dots, N - 1\} \quad (1)$$

where the function $Th(\cdot)$ assigns 1 to pixels in I for those values that are greater than θ_i and 0 otherwise. The value of θ_i is defined as:

$$\theta_i = \left\lfloor \frac{i}{N} \times 255 \right\rfloor \quad (2)$$

We do not perform thresholding at $i = 0$ (minimum gray level) and $i = N$ (maximum gray level) because the thresholding results in these cases are meaningless, the image will be black or white entirely. Fig. 7 illustrates the Boolean maps created at this step with $N = 10$. The topmost image is the original infrared image, the below nine images are Boolean maps generated from this image. The threshold levels increase from left to right and from top to bottom. As can be seen in Fig. 7, by using multiple threshold values, we successfully exposed visual features from the infrared image.

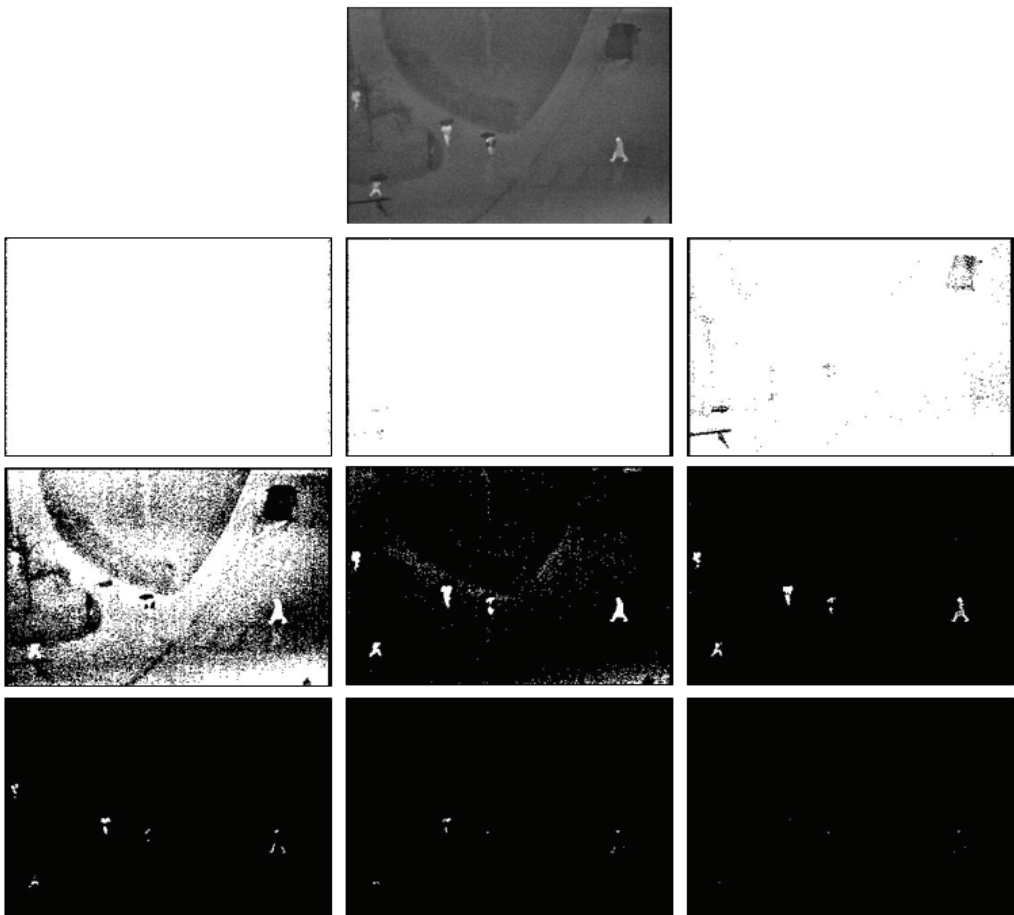


Fig. 7. Generated Boolean maps.

After obtaining the Boolean maps, the initial saliency map is calculated as a weighted sum of Boolean maps as follows:

$$S = \sum_i \omega_i B_i \tag{3}$$

where:

$$\omega_i = \frac{\theta_i}{255} \tag{4}$$

which means that Boolean maps created using higher thresholds have higher weights. This weighting scheme is based on the observation that infrared imaging captures the temperature of objects. Hence, salient objects in infrared images are usually small and bright indicating heat sources such as humans or vehicles. Fig. 8 shows that the weighted sum enables combining features from multiple Boolean maps exposing potential salient objects. However, the initial saliency map remains unclear at this step.

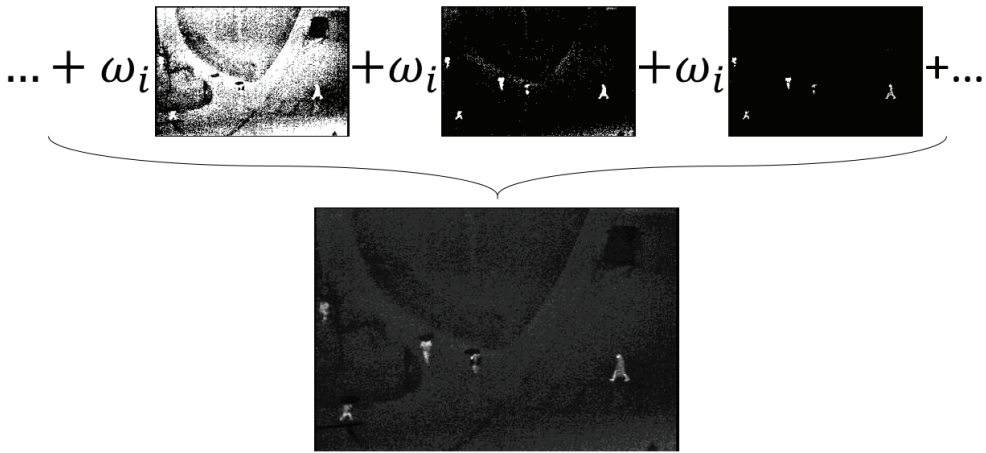


Fig. 8. The initial saliency map computed by weighted sum.

The initial saliency map S may contain noise due to the nature of infrared imaging. To eliminate the noise, we threshold the initial map by using a modified version of Otsu’s method proposed in [18]. This technique uses entropy information as an additional weighting scheme for Otsu’s method. It is worth noting that values of pixels in B_i are 0 and 1. Before going to the next step, the values of pixels in S are rescaled to the range $[0, 255]$ (rounded down). For the rescaled initial saliency map S , let the number of pixels with value i be m_i and M be the total number of pixels in S , a posteriori entropy of the map is then defined as:

$$H'_n = -P_k \ln P_k - (1 - P_k) \ln(1 - P_k) \tag{5}$$

Where

$$P_k = \sum_{i=0}^k p_i, \quad 1 - P_k = \sum_{i=k+1}^{255} p_i, \quad p_i = \frac{m_i}{M}, \tag{6}$$

If we maximize H'_n , we get a trivial result:

$$P_k = 1 - P_k = \frac{1}{2} \quad (7)$$

This problem is solved by converting (5) into an objective function:

$$\psi(k) = \ln P_k(1 - P_k) + \frac{H_k}{P_k} + \frac{H_n - H_k}{1 - P_k} \quad (8)$$

where

$$H_k = - \sum_{i=0}^k p_i \ln p_i, \quad H_n = - \sum_{i=0}^{255} p_i \ln p_i \quad (9)$$

Originally, the objective function of Otsu's method is written as follows:

$$k^* = \underset{0 \leq k \leq 255}{\operatorname{argmax}} \{(\varphi_1(k)\mu_1^2(k) + \varphi_2(k)\mu_2^2(k))\} \quad (10)$$

where φ and μ are the probability of class occurrence and the average pixel values of each class (foreground and background in this case) respectively, and the value k^* is the optimal value for thresholding the image. The entropy calculation ψ is combined with Otsu's method to create a new objective function:

$$k^* = \underset{0 \leq k \leq 255}{\operatorname{argmax}} \{\psi(k)(\varphi_1(k)\mu_1^2(k) + \varphi_2(k)\mu_2^2(k))\} \quad (11)$$

The modified version is better than the original method in term of highlighting small salient regions and is not affected by noise, as shown in Fig. 9.

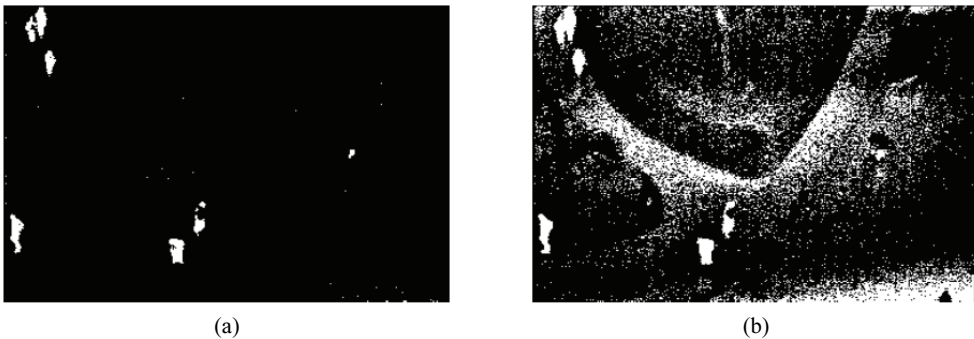


Fig. 9. Results from thresholding the initial saliency map using different methods: (a) the image obtained through modified Otsu's method and (b) the image obtained through original Otsu's method.

After thresholding, the initial saliency map becomes clearer, exposing salient objects. However, it is still cluttered by small leftover noise particles. In the final step, the saliency map is further refined with the use of opening, a morphological operation to eliminate noise particles. Thereafter, we apply a Gaussian filter to create the final saliency map (Fig. 10).

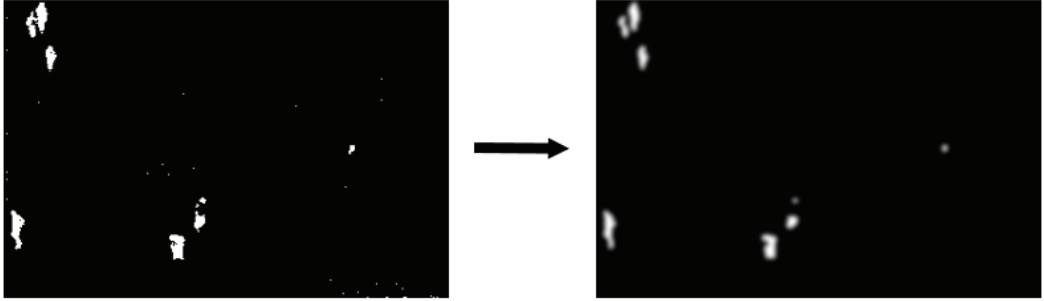


Fig. 10. Resulting saliency maps: (left) the map resulting from thresholding and (right) the final saliency map after applying morphological operation and Gaussian filter.

3. Experimental Results

We evaluated the performance of the proposed method using Ohio State University (OSU) Thermal Pedestrian Database from [19]. This data set contains infrared image sequences from a camera mounted on a rooftop of a building. The main objects of interest in this data set are pedestrians. The images are in gray scale and have a resolution of 360×240 pixels. Because this data set is used for pedestrian detection in the first instance, only bounding boxes around pedestrians are provided as the ground truth. We considered pedestrians as salient objects; hence, we used the bounding boxes as the indicators for salient regions. Fig. 11 shows several sample images from the data set: the first row provides original infrared images and in the second row, the locations of pedestrians in the corresponding images from the first row are indicated by yellow bounding boxes.

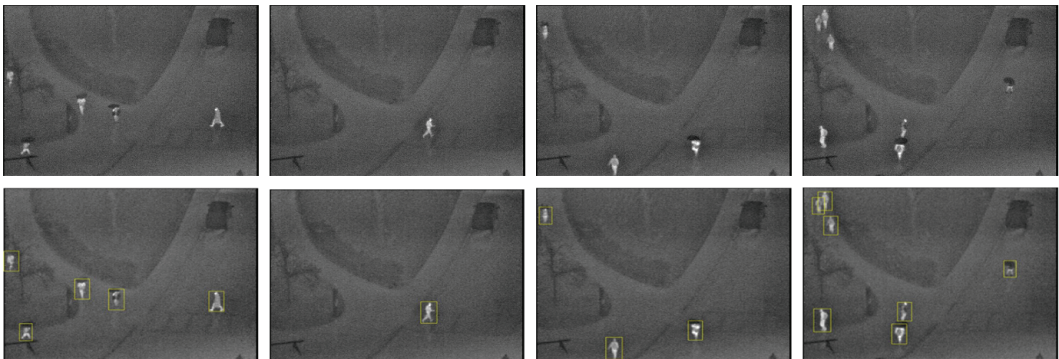


Fig. 11. Sample images from dataset with marked pedestrian.

The main parameter in the proposed method is N , the number of thresholding levels. If N is small, the number of exposed features is insufficient. If N is large, some Boolean maps will become redundant because the threshold values are very close to each other, which leads to unnecessary computation. In Fig. 12, we demonstrated the calculations of initial saliency maps with different values of N . The visual details of initial map created with $N = 5$ are insufficient, while the maps created with $N = 20$ and $N = 40$ are nearly identical. In the experiments, we selected $N = 20$ to achieve the best results while retaining low computational cost.

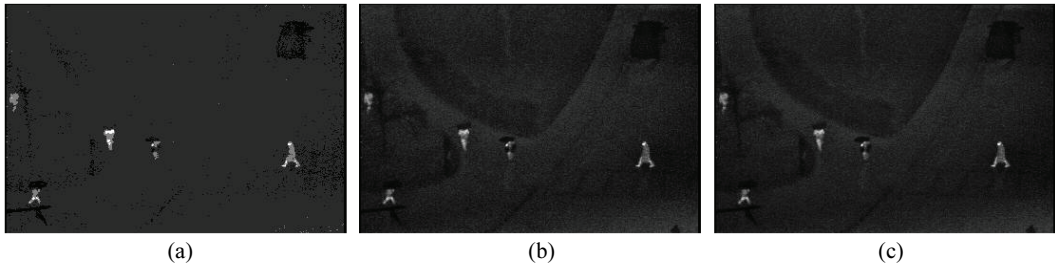


Fig. 12. Saliency map with different value of N . (a) $N = 5$, (b) $N = 20$, and (c) $N = 40$.

We also compared our method with the Boolean Map Saliency approach by Zhang and Sclaroff [17], which is also a saliency detection technique utilizing Boolean maps. The BMS can be considered as the first attempt to apply the Boolean map theory in saliency detection. Originally, BMS was designed to detect salient objects in three-channel color images. In this experiment we examine the performance of BMS when being applied to infrared images. The implementation of BMS was obtained from the author's website. It is also worth noting that for BMS we used the maps of eye fixation predictions instead of salient object detection results, because the saliency maps were very imprecise. The saliency detection results are illustrated in Fig. 13.

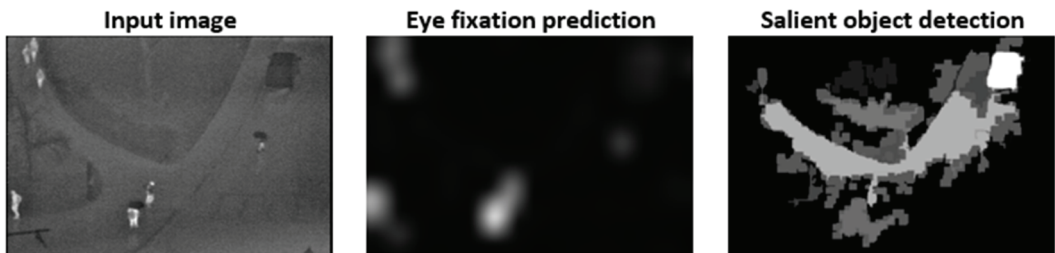


Fig. 13. Eye fixation prediction and salient object detection results using BMS, the images are obtained using BMS author's implementation.

In Fig. 14, the original infrared images are shown in the left column (Fig. 14(a), (d), (g), (j), and (m)), in which the pedestrians are highlighted by yellow bounding boxes. Each row shows the detection results obtained through BMS and through the proposed method for the respective infrared image. In the first row, there are seven pedestrians in the image (Fig. 14(a)). The result from BMS indicated that there are two large salient regions (Fig. 14(b)), while our algorithm produced six distinct salient regions indicating locations of the pedestrians (Fig. 14(c)); the remaining pedestrian on the right side of the image is small and hard to detect. Detection results are similar for other tests. BMS could only produce unclear large salient regions, while the proposed method was able to produce precise, distinct salient regions that indicated pedestrians' locations clearly. As can be seen, the proposed method produced better and clearer saliency maps for this data set. In particular, it captured the general shapes of the pedestrians. The detected salient regions locate accurately inside the bounding boxes. In its turn, BMS produced unclear salient regions, especially for regions that are close to image boundaries (in Fig. 14(e), (k), and (m)), the pedestrians at top left corners are unclear). To a certain extent, the precisely detected salient regions from the proposed method can be used as a feature in pedestrian tracking systems employing infrared video recorders.

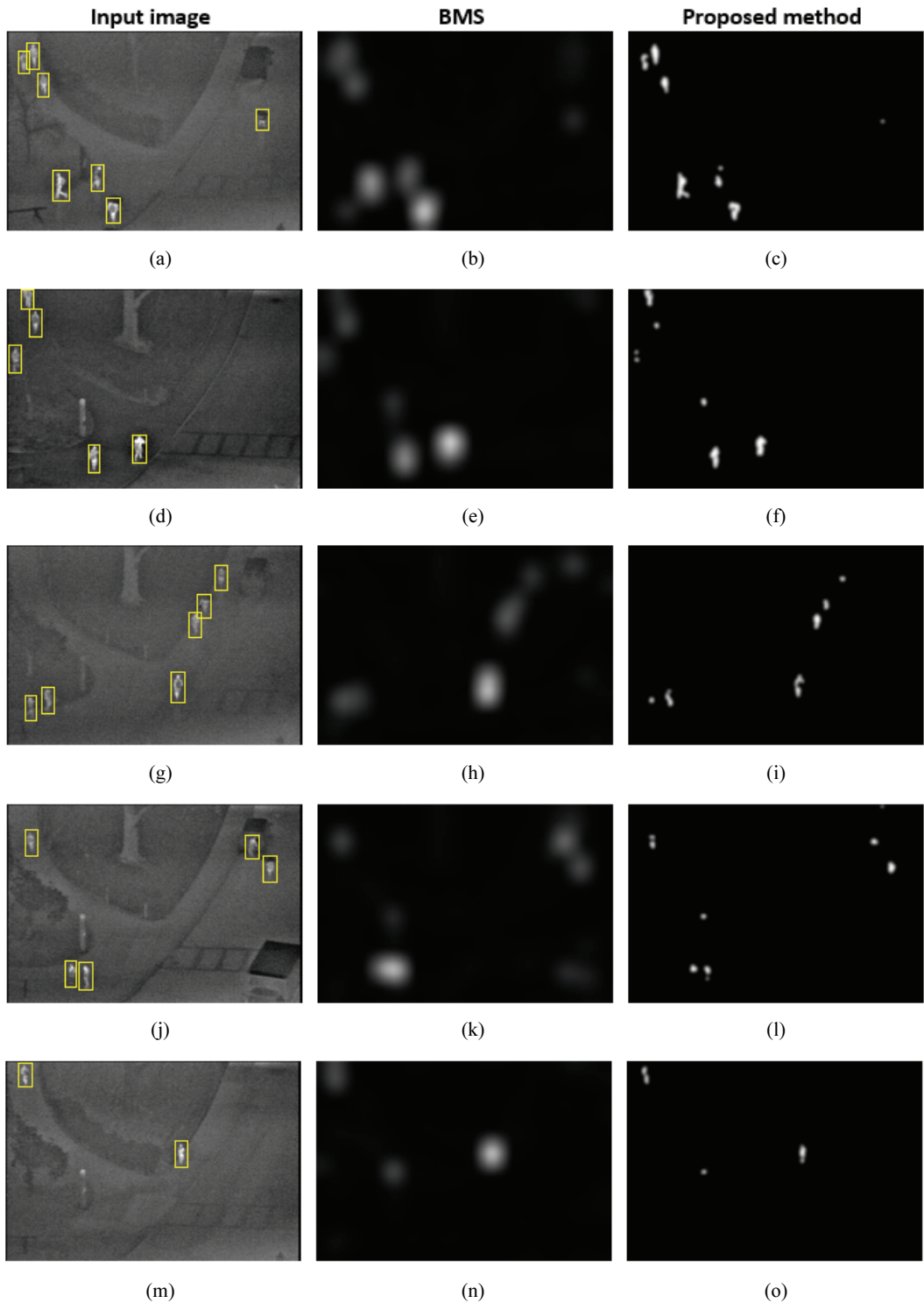


Fig. 14. Comparison between saliency detection results obtained using BMS and the proposed method: (a, d, g, j, m) infrared images with pedestrians inside yellow bounding boxes, (b, e, h, k, n) resulting images obtained through BMS, and (c, f, i, l, o) resulting images obtained through the proposed method.

4. Conclusion

In computer vision, visual saliency detection is a pixel-level classification task, in which each pixel is assigned a saliency score to determine whether it belongs to the salient objects or to the background. Research on visual saliency plays an important role in computer vision as it helps other vision-based algorithms achieve remarkable results because visual saliency models are representations of the human attention mechanism. Among various types of visual data, infrared images are related to a specialized visual data, which has various applications in critical fields. Even though the number of saliency detection algorithms is high and continues growing, techniques for saliency detection in infrared images are not as well developed as the ones in color images. Saliency detection algorithms that work well with color images may show unsatisfactory performance when being applied to infrared images. It is because infrared images are usually highly noisy and low-contrast, besides they have low resolution and insufficient visual features. Saliency detection in infrared images remains a challenging problem due to the nature of input data.

In this paper, we introduced a straightforward approach for saliency detection in infrared images. In this approach, first the input image is thresholded into several Boolean maps, and the number of maps is selected based on observation for the best results. Thereafter, an initial saliency map is calculated as a weighted sum of created Boolean maps. Finally, a high-quality saliency map is constructed by further refining the initial map through thresholding, morphology operation, and Gaussian filter. By constructing multiple Boolean maps, we exposed image features in different threshold levels, solving the low-contrast problem. Combining it with the adaptive thresholding algorithm, we suppressed the noise in infrared images while preserving details of the images. The experiment showed that the proposed method produced high performance results when applied to real-life infrared image sequences. For future work, we plan to utilize Boolean maps further, for such purposes as extracting more features or performing image enhancement to improve and to increase the performance of the proposed method of saliency detection in infrared images system.

Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2020R1F1A1067496).

References

- [1] R. Achanta and S. Susstrunk, "Saliency detection for content-aware image resizing," in *Proceedings of 2009 16th IEEE International Conference on Image Processing (ICIP)*, Cairo, Egypt, 2009, pp. 1005-1008.
- [2] S. Wulf and U. Zolzer, "Visual saliency guided mode decision in video compression based on Laplace distribution of DCT coefficients," in *Proceedings of 2014 IEEE Visual Communications and Image Processing Conference*, Valletta, Malta, 2014, pp. 490-493.

- [3] P. Mukherjee and B. Lall, "Saliency and KAZE features assisted object segmentation," *Image and Vision Computing*, vol. 61, pp. 82-97, 2017.
- [4] R. G. Mesquita and C. A. Mello, "Object recognition using saliency guided searching," *Integrated Computer-Aided Engineering*, vol. 23, no. 4, pp. 385-400, 2016.
- [5] K. M. Koo and E. Y. Cha, "Image recognition performance enhancements using image normalization," *Human-centric Computing and Information Sciences*, vol. 7, article no. 33, 2017.
- [6] C. Yuan, X. Li, Q. J. Wu, J. Li, and X. Sun, "Fingerprint liveness detection from different fingerprint materials using convolutional neural network and principal component analysis," *Computers, Materials & Continua*, vol. 53, no. 3, pp. 357-371, 2017.
- [7] A. Borji, M. M. Cheng, H. Jiang, and J. Li, "Salient object detection: a benchmark," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5706-5722, 2015.
- [8] Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba, "MIT saliency benchmark," [Online]. Available: http://saliency.mit.edu/results_mit300.html.
- [9] T. Tsukamoto, M. Esashi, and S. Tanaka, "Infrared-to-visible transducer using temperature sensitive Eu (TTA)3 on self-suspended thin film for inexpensive thermal imaging device," in *Proceedings of 2013 IEEE 26th International Conference on Micro Electro Mechanical Systems (MEMS)*, Taipei, Taiwan, 2013, pp. 421-424.
- [10] L. Chen, H. C. Chen, Z. Li, and Y. Wu, "A fusion approach based on infrared finger vein transmitting model by using multi-light-intensity imaging," *Human-centric Computing and Information Sciences*, vol. 7, article no. 35, 2017.
- [11] C. Liu, I. Cheng, and A. Basu, "Real-time runway detection for infrared aerial image using synthetic vision and an ROI based level set method," *Remote Sensing*, vol. 10, article no. 1544, 2018.
- [12] J. Kim, D. Han, Y. W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform and local spatial support," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 9-23, 2015.
- [13] W. Li, C. Pan, and L. X. Liu, "Saliency-based automatic target detection in forward looking infrared images," in *Proceedings of 2009 16th IEEE International Conference on Image Processing (ICIP)*, Cairo, Egypt, 2009, pp. 957-960.
- [14] L. Li, Y. Zheng, and F. Zhou, "Contrast and distribution based saliency detection in infrared images," in *Proceedings of 2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP)*, Xiamen, China, 2015, pp. 1-6.
- [15] S. Qin, L. Wang, H. Cheng, Q. Feng, M. Zhang, and C. Gao, "Infrared image saliency detection based on human vision and information theory," in *Proceedings of 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Datong, China, 2016, pp. 484-488.
- [16] L. Huang and H. Pashler, "A Boolean map theory of visual attention," *Psychological Review*, vol. 114, no. 3, pp. 599-631, 2007.
- [17] J. Zhang and S. Sclaroff, "Exploiting surroundedness for saliency detection: a Boolean map approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 5, pp. 889-902, 2016.
- [18] M. T. N. Truong and S. Kim, "Automatic image thresholding using Otsu's method and entropy weighting scheme for surface defect detection," *Soft Computing*, vol. 22, no. 13, pp. 4197-4203, 2018.
- [19] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery," in *Proceedings of 2005 7th IEEE Workshops on Applications of Computer Vision (WACV/MOTION)*, Breckenridge, CO, pp. 364-369.



Mai Thanh Nhat Truong <https://orcid.org/0000-0002-6448-7837>

He received his B.Sc. degree in Mathematics and Computer Science from the Ho Chi Minh City University of Science, Vietnam, in 2014, and M.Sc. degree in Electrical, Electronic, and Control Engineering from Hankyong National University, Korea, in 2017. Since September 2017, he has been with the Department of Electrical, Electronic, and Control Engineering at Hankyong National University, Korea, as a PhD candidate. His research interests are machine vision and image analysis.



Sanghoon Kim <https://orcid.org/0000-0001-5351-8215>

He received his B.Sc., M.Sc., and Ph.D. degrees in Electronic Engineering from Korea University, Seoul, in 1987, 1989, and 1999, respectively. From 1989 to 1994, he was a Research Engineer with the LG Semiconductor Company, where he was engaged in the research and development of PC chipset design. From January 2004 to January 2005, he was a Visiting Scholar with the University of Maryland, College Park, MD, USA. Since September 1999, he has been with Hankyong National University, Anseong, Korea, where he is currently a Professor. His current research interests are in the areas of image processing, object detection, and robot vision. Prof. Kim is a member of IEEE and the Korean Information Processing Society.