IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 9, 2019, accepted July 29, 2019, date of publication August 5, 2019, date of current version August 28, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2933310

Mutual Guidance-Based Saliency Propagation for Infrared Pedestrian Images

YU ZHENG[1], FUGEN ZHOU[1], (Member, IEEE), LU LI[1], (Member, IEEE), XIANGZHI BAI[1,2,3], (Member, IEEE), AND CHANGMING SUN[4]

[1]Image Processing Center, Beihang University, Beijing 100191, China
[2]State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China
[3]Advanced Innovation Center for Biomedical Engineering, Beihang University, Beijing 100083, China
[4]CSIRO Data61, Sydney, NSW 1710, Australia

Corresponding author: Xiangzhi Bai (e-mail: jackybxz@buaa.edu.cn)

This work was supported by the National Nature Science Foundation of China under Grant U1736217.

ABSTRACT Saliency detection is important in computer vision. However, most of the existing saliency models are designed for visible images. It is still a challenging problem to apply saliency detection algorithms on infrared images. In this paper, an effective propagation based saliency detection method for infrared pedestrian images is proposed. Firstly, based on the thermal characteristics of infrared images and thermal radiation models, a thermal analysis based saliency (TAS) is introduced. TAS measures the stableness of pedestrians based on maximally stable extremal regions, which is further improved by an intensity filter. Then, by taking into account the appearance characteristic of pedestrians, an appearance analysis weighted saliency (AAS) is proposed which combines the intensity and shape features of pedestrians to improve the intensity contrast. Finally, besides the commonly used intra-scale neighborhood, an inter-scale neighborhood is introduced to jointly construct a mutual guidance-based saliency propagation model. This model could simultaneously integrate the saliency features and improve the saliency performance. Two datasets DIP and IMS with 600 infrared pedestrian images are published. Then, extensive experiments and comparisons with state-of-the-art methods demonstrate the effectiveness of the proposed saliency method for infrared pedestrian images.

INDEX TERMS Infrared images, pedestrian, saliency, propagation.

I. INTRODUCTION

Human vision has the ability to effectively select relevant information out of irrelevant noises and to locate the highly relevant subjects in a scene. As a fundamental issue in computer vision, saliency detection has been applied as a pre-processing procedure to a wide range of computer vision tasks, such as object segmentation [1], image compression [2], object detection [3], and image retrieval [4]. As saliency detection is capable of finding the most important and distinctive region in an image, we apply saliency detection to infrared pedestrian detection, which is an essential and important task for driving assistants and intelligent transportation systems. However, constrained by the characteristics of infrared imaging, it is still challenging to accurately detect saliency in infrared pedestrian images.

The associate editor coordinating the review of this article and approving it for publication was Avishek Guha.

The development of saliency detection methods can be roughly divided into two stages. The first stage focuses on exploring low-level cues of salient objects, such as color [5], orientation [6], and texture [7]. Because of the uniqueness and rareness of salient objects, contrast prior has been widely used as a computational mechanism to measure the difference between foreground and background. Contrast could be investigated from both local and global perspectives according to the scale of pixel neighborhoods. Local contrast [8]–[10] assumes that the more distinctive an object is compared with its neighborhoods, the more salient this object will be. However, contrast with only local cues always results in wrongly suppressed internal regions of salient objects. To alleviate these problems, global contrast [11] is proposed, which assigns higher saliency scores to objects with more unique features in the whole image. Global contrast is useful to highlight the whole object, but it may fail to thoroughly suppress the background. Previous contrast

mechanisms usually take pixels as processing units, which may suffer from a boundary blurring problem. To obtain saliency maps with well-defined boundaries, contrast based on segments is exploited. It could suppress noises in background and reduce the computational load, as used in methods such as simple linear iterative clustering (SLIC) [12], mean shift [13], and Gaussian mixture model [14].

The second stage of saliency detection is the propagation based saliency detection. Recently, propagation algorithms attract increasing attention in saliency detection and have achieved state-of-the-art performances. Markov chains [15], random walks [16], and manifold ranking [17] are the most frequently used propagation methods, which are all based on graphs. Harel *et al.* [18] first put forward the graph based visual saliency, which employs an ergodic Markov chain to produce feature maps. Li *et al.* [19] propose a novel regularized random walk, which suggests a fitting constraint to take into account the local image data and prior estimation. Later, Zhang *et al.* [17] infer the saliency score of each region via graph-based manifold ranking which ranks the similarity of superpixels with foreground or background seeds. In addition to these classic methods, various new patterns of saliency propagation are proposed. Li *et al.* [20] define the saliency value using a co-transduction algorithm, which fuses both boundary and objectness labels through an inter propagation scheme. Qin *et al.* [21] present a cellular automata based saliency propagation method exploiting the intrinsic relevance between neighboring cells to improve the saliency performance. Qin *et al.* [22] further propose the Cuboid Cellular Automata to integrate multiple saliency maps in a Bayesian framework, which incorporates the low-level image features as well as high-level semantic information. Nevertheless, these saliency propagation methods still cannot perform well with challenging images, especially when the salient objects are similar to backgrounds.

Even though various saliency models have been proposed recently, most of them are designed for visible images. Some works directly apply these state-of-the-art models on infrared images as pre-processing to locate salient objects [23], [24]. But they could only obtain coarse results or even fail in saliency detection. Compared with visible images, infrared images have unique advantages. They are less sensitive to lighting conditions, and this makes it possible to eliminate the influence of illumination variations, so it could be used in both day and night and other difficult situations. Additionally, benefiting from the insensitivity to color, texture, and other appearance features, infrared images can be used to separate objects with similar appearances by their thermal radiation differences. With infrared pedestrian images, more challenges exist. Firstly, due to the limitation of infrared thermal imaging, infrared pedestrian images have low clarity, low SNR, and low contrast. Secondly, there is no color or little texture information in infrared images, which makes it difficult to extract saliency features of objects in infrared images. And this is also the primary reason why most of the existing saliency models fail with infrared images. Thirdly, high image intensity is a crucial characteristics for pedestrians in infrared images, but non-human objects, such as light poles, vehicles, and tree trunks, may also produce additional bright areas. These interferences increase the difficulty of saliency detection in infrared pedestrian images.

To apply saliency detection to infrared pedestrian images, some researches have been carried out. Ko *et al.* [25] calculate the luminance saliency map by estimating the luminance contrast using a center-surrounded scheme. Zhang *et al.* [26] propose an associative saliency, generated from both region and edge contrasts. Li *et al.* [27] apply the gradient information on pedestrians to enhance the uniqueness of intensity, and combine it with multi-scale contrasts to obtain the final saliency. Wang *et al.* [28] exploit a mutual consistency guided fusion strategy to adaptively combine the luminance contrast saliency map and contour saliency map for infrared images. Li *et al.* [1] first calculate the background likelihood with background prior, and then use a Bayesian model to obtain the object prior based saliency. The final saliency of this method is an integration of background prior and object prior.

However, previous saliency models designed for infrared images mainly use low-level features, such as gradient and intensity to describe salient objects, and employ weighted summation or multiplication to integrate these features. Thus, these features only fit simple images, and they perform poorly for complex infrared scenes, which have diverse composition of backgrounds, including trees, buildings, roads, skies, street lamps, brushwood, and other objects. And taking the above problems into consideration, our work proposes two unique saliency features from both thermal characteristics and appearance characteristics to describe pedestrians in infrared images. These two features have better ability to represent the saliency of pedestrians in infrared images. Also, our algorithm introduces saliency propagation to integrate features and optimize the saliency performance simultaneously. The proposed method consists of three parts: Firstly, the thermal analysis based saliency (TAS) is proposed based on the thermal characteristics of pedestrians and radiation models; Secondly, taking into account the appearance features, the appearance analysis-weighted saliency (AAS) is introduced as a complement; At last, the mutual guidance-based saliency propagation method is proposed in this paper to mutually facilitate the two features and improve the final saliency.

Thus, the main contributions of this paper are as follows:

- A novel propagation based saliency model is proposed to adaptively detect pedestrians from complex infrared images. The proposed method advances state-of-the-art saliency detection methods on both public datasets and a more complex dataset constructed in this work.
- Two features are explored from both an infrared imaging mechanism and the actual performance to describe the saliency of pedestrians in infrared images, including TAS and AAS. These features are able to distinguish pedestrians from complex backgrounds.

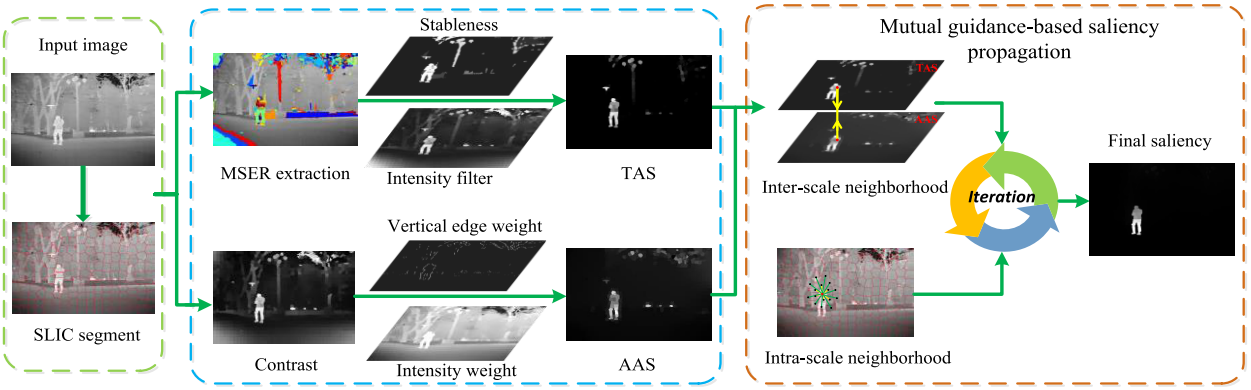


FIGURE 1. The diagram for the proposed saliency detection method.

- A mutual guidance based saliency detection method is developed in this paper, which puts forward the concepts of intra-scale and inter-scale neighborhoods. This propagation method can not only integrate the two saliency features but also correct any mistakes in initial saliency maps to improve the final saliency.
- Two datasets IMS and DIP are constructed including 600 infrared pedestrian images with more than 33 scenes. We publish the dataset and the source code of this work at <https://github.com/zhxtu/SP_IR>.

II. PROPOSED METHOD

Fig. 1 shows the diagram of the proposed saliency detection method for infrared pedestrian images. Firstly, SLIC [29] is used to segment the input infrared image into homogeneous superpixels. Secondly, the maximally stable extremal region (MSER) [30] is extracted to measure the stableness of pedestrians, which is further improved by an intensity filter to obtain the thermal analysis based saliency (TAS). Thirdly, the intensity contrast is calculated and further enhanced by the vertical edge weight and intensity weight to obtain the appearance analysis-weighted saliency (AAS). Finally, a mutual guidance based propagation method, which combines the intra-scale and inter-scale neighborhoods, is introduced to integrate the two features and improve the final saliency.

A. THERMAL ANALYSIS BASED SALIENCY (TAS)

Infrared images are generated from the translation of thermal radiation through thermographic cameras. Thus, infrared images are the products of the complex interaction among factors such as temperature, emissivity, and atmosphere effect. Besides, the intensity of each object is determined not only by the thermal radiation of the object itself, but also by the reflection of other objects and the atmosphere [31]. Calculating the saliency of pedestrians is actually suppressing the radiation from background and obtaining the radiation of pedestrians themselves.

Based on the thermal analysis, we first introduce the MSER-based local stableness, which is further improved by the intensity filter to obtain the TAS.

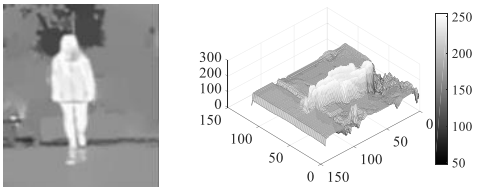


FIGURE 2. An example of a local region in an infrared pedestrian image, and the corresponding 3D intensity plot.

1) MSER-BASED LOCAL STABLENESS

Fig. 2 shows an infrared image with a pedestrian and its corresponding 3D intensity plot. Obviously, the intensity on the pedestrian differ greatly from that of its surrounding regions. This phenomenon results from the thermal imaging principle [32] that stronger thermal radiations generate higher intensities. As temperature increases, the atomic and molecular activity would be enhanced. This would produce more heat and stronger thermal radiation. Thus, pedestrians with higher temperatures are usually brighter than the background.

Besides, object emissivity serving as a decisive factor of infrared radiation is closely related to the material property of the object [31]. Thus, regions composed of different materials differ in intensity accordingly. Then, pedestrian regions are different from their surrounding regions and are completely surrounded by regions with lower intensities.

Following the principle that areas surrounded by others tend to be more salient, infrared pedestrian regions could be described by the capacity of the MSER for detecting the surrounded regions with a homogeneous intensity. Thus, the MSER-based local stableness is proposed. Although MSER is an existing approach, it is mostly applied in text localization and has not been used to measure accurate saliency yet. MSER is defined by an extremal property of its intensity function in the region and on its outer boundary. To calculate MSER in an image $\boldsymbol{I}_m$, the extremal regions are defined as $\boldsymbol{R}_l$:

$$\forall p \in \boldsymbol{R}_l, \quad \forall q \in boundary(\boldsymbol{R}_l) \rightarrow I_m(p) \geq I_m(q), \qquad (1)$$

FIGURE 3. (a) Pixel based stableness F; (b) Superpixel based stableness without an intensity filter; (c) Superpixel based stableness with an intensity filter.

where $I_m(p)$ is the intensity of pixel p in the region and $I_m(q)$ is the intensity of pixel q on its outer boundary. The extremal regions are identified as connected regions within the binary threshold images $\boldsymbol{I}_{bim}^{g}$:

$$\boldsymbol{I}_{bim}^{g} = \begin{cases} 1 & \boldsymbol{I}_m \geq g \\ 0 & \text{otherwise} \end{cases} g \in [\min(\boldsymbol{I}_m), \max(\boldsymbol{I}_m)], \quad (2)$$

where threshold g is a series of integers from the lowest intensity value to the highest intensity value of the input image $\boldsymbol{I}_m$. To generate MSER from $\boldsymbol{R}_l$, the stableness value $\boldsymbol{\Psi}$ is calculated for each connected region as follows:

$$\Psi(\boldsymbol{R}_l^g) = (|\boldsymbol{R}_l^{g+\delta} - \boldsymbol{R}_l^{g-\delta}|)/|\boldsymbol{R}_l^g|, \quad (3)$$

where $\boldsymbol{R}_l^g$ is the l-th region in image $\boldsymbol{I}_{bim}^{g}$, and δ is a stability range. If $\Psi(\boldsymbol{R}_l^g)$ is lower than threshold T_M, $\boldsymbol{R}_l^g$ would be taken as MSER. Thus, the final MSER contains K stable regions $\boldsymbol{SR} = \{\boldsymbol{sr}_1, \boldsymbol{sr}_2, \ldots, \boldsymbol{sr}_K\}$.

To measure the stableness $\boldsymbol{F}$ of each pixel, their probability of belonging to stable regions is calculated. For each stable region in $\boldsymbol{SR}$, pixels inside the region are set as 1 while other pixels are set as 0 to obtain the score matrix $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_K\}$. Thereafter, the number of stable regions which overlap each other in the same pixel is accumulated to measure the stableness of the corresponding pixel. The more stable a pixel is, the higher its probability of belonging to a pedestrian will be:

$$F(p) = \sum_{k=1}^{k} e_k(p) \quad e_k(p) = \begin{cases} 1 & p \in \boldsymbol{sr}_k \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where $e_k(p)$ indicates whether pixel p belongs to the k-th stable region $\boldsymbol{sr}_k$. And $F(p)$ is the stableness for pixel p, which is shown in Fig. 3(a).

As thermal radiation from parts of a human body is hampered by clothes, there are generally noises inside pedestrian regions. In order to smooth the internal distribution of intensities inside pedestrian regions, the image is segmented into N homogeneous superpixels $\boldsymbol{SP} = \{\boldsymbol{sp}_1, \boldsymbol{sp}_2, \ldots, \boldsymbol{sp}_N\}$ by SLIC. And then, the saliency value $\boldsymbol{F}_s$ of each superpixel is calculated by mapping the pixel-wise stableness $\boldsymbol{F}$ into its corresponding superpixel:

$$F_s(i) = \frac{\sum_{p \in \boldsymbol{sp}_i} F(p)}{|\boldsymbol{sp}_i|}. \quad (5)$$

With the accumulation on each superpixel, stable regions could be enhanced and backgrounds are suppressed, while the accurate contour information could also be preserved. Fig. 3(b) shows that superpixel based stableness can reduce the inhomogeneous saliency distribution inside human body regions and partly reduce noises in background.

2) INTENSITY FILTER-ENHANCED SALIENCY

With only TAS, some objects, such as street lamps and tree trunks, may be wrongly assigned with high saliency values. To distinguish pedestrians from other objects, the principle that pedestrians always produce stronger thermal radiation is used. For pedestrians in the scene, the thermal radiation received by an infrared camera is not only from pedestrians themselves, but also from the radiation reflected from other objects onto pedestrians and the thermal radiation of atmosphere. According to the physics of radiation [33], emissivity and reflectivity are inversely proportional. And the reflectivity of the pedestrian is usually much lower than its emissivity because of its rough surface. Thus, radiation reflected from other objects could be ignored. As the radiation of atmosphere is directly received by a thermal sensor, the influence of atmosphere is significant. As a result, the total radiation composition E of an object is:

$$E = E_o + E_A, \quad (6)$$

where E_A is the radiation from atmosphere, and E_o is the radiation of the object itself. By subtracting E_A from E, E_o is obtained to measure the saliency.

Since the values of E, E_o, and E_A cannot be directly calculated, their corresponding contributions on intensities are employed. E corresponds to the intensity value in the image. E_o of pedestrians is much different from E_o of other objects, thus the contribution of E_o to intensity is obtained to show the saliency of pedestrians. Since atmosphere exists in the whole scene, the contribution of E_A on intensity is defined as the average intensity I_μ of the image $\boldsymbol{I}_m$. Corresponding to Eq. (6), the intensity filter $\boldsymbol{IF}$ is defined by subtracting I_μ from the average intensity of each superpixel to obtain the saliency of pedestrians:

$$IF(i) = \left| \frac{\sum_{p \in \boldsymbol{sp}_i} I_m(p)}{|\boldsymbol{sp}_i|} - I_\mu \right|^2, \quad (7)$$

where $|\boldsymbol{sp}_i|$ is the area of the i-th superpixel. By enhancing the stableness with an intensity filter, TAS is calculated as:

$$S_{TAS}(i) = F_s(i) \cdot IF(i). \quad (8)$$

By subtracting I_μ from the intensity of each superpixel, the radiation of atmosphere would be removed. The high intensity of pedestrian regions is the result of its strong radiation, which can produce a higher $\boldsymbol{IF}$ value. Also, with l_2-th calculation, the difference of saliency between pedestrians and other objects would be further enlarged. With the integration of the intensity filter and stableness, the saliency of pedestrians is effectively improved while the saliency of other objects is suppressed. It is obvious in Fig. 3(c) that

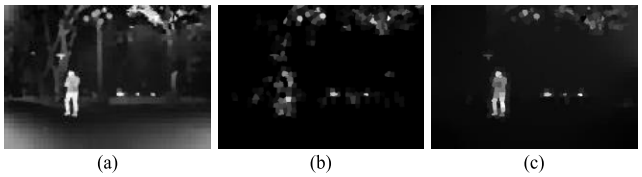


FIGURE 4. (a) Contrast without weight; (b) Contrast with vertical edge weight; (c) Contrast with both vertical edge weight and intensity weight.

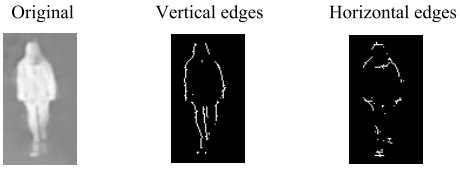


FIGURE 5. An example of a pedestrian in an infrared image, and its vertical edges and horizontal edges obtained by the Canny edge detection method.

the intensity filter has greatly suppressed the background and enhanced the performance of stableness.

B. APPEARANCE ANALYSIS-WEIGHTED SALIENCY (AAS)

Although TAS has a good ability to make pedestrian regions prominent, some targets which are too small or too similar to the background may be wrongly suppressed by TAS. As a supplement, the AAS is introduced. Contrast is a commonly used feature in saliency detection, which often measures the color difference. Observed from infrared pedestrian images, the intensity distribution of a pedestrian is obviously different from backgrounds. Therefore, the contrast can also be applied to infrared images to highlight pedestrians. And the contrast is defined as:

$$con(i) = \sum_{j=1}^{N} |v_i - v_j| \cdot \exp\left(d_i - d_j|\right), \tag{9}$$

where d_i and v_i are the coordinates and average intensity for $\boldsymbol{sp}_i$. And d_j and v_j are the corresponding values for $\boldsymbol{sp}_j$. As shown in Fig. 4(a), contrast has the ability to make pedestrians more prominent.

However, there are two shortcomings for the contrast feature. Firstly, low contrast is an inherent characteristic of infrared images, which makes it difficult to separate pedestrians from backgrounds with only the contrast feature. And then, tree, lamps, and other objects with high intensities may also have high values in the contrast map, which may affect the saliency detection of pedestrians. To handle these problems, the AAS is introduced which employs the appearance information of pedestrians to enhance the contrast, which is calculated as:

$$S_{AAS}(i) = w_i \cdot Con(i), \tag{10}$$

where w_i is the appearance weight for superpixel $\boldsymbol{sp}_i$, composed of the vertical edge weight and intensity weight.

Vertical shape is a distinct feature of pedestrians, which is widely used in pedestrian detection and recognition. Aspect ratio [34] is commonly used to describe the vertical feature of pedestrians, yet it is inaccurate and difficult to extract. In this paper, the vertical edge weight is used to describe the vertical feature of pedestrians. As objects usually contain more edge information than background, superpixels with more edge information are more likely to belong to the salient object. Also, the vertical edges of a pedestrian are much stronger than the horizontal edges and can better represent a pedestrian as shown in Fig. 5.

To calculate the vertical edge weight $\boldsymbol{w}^{ve}$, the probability of boundary (PB) [35] is used to detect the boundary map $\boldsymbol{M}^{pb}$ of the input image. Then, the vertical gradient $\boldsymbol{g}_v$ of $\boldsymbol{M}^{pb}$ is obtained to measure the vertical edge weight $\boldsymbol{w}^{ve}$:

$$w_i^{ve} = \frac{1}{|\boldsymbol{b}_i|} \sum_{p \in \boldsymbol{b}_i} g_v(p), \tag{11}$$

where w_i^{ve} and $\boldsymbol{b}_i$ represent the vertical edge weight and edge pixel set for superpixel $\boldsymbol{sp}_i$ respectively, and $g_v(p)$ is the vertical gradient value for pixel p. Seen from Fig. 4(b), the vertical edge weight is able to suppress backgrounds.

However, background regions along edges are wrongly highlighted, while the regions inside the pedestrian with only a few edges are also mistakenly suppressed by the vertical edge weight at the same time. As pedestrians have higher intensities than surrounding regions, the intensity of each superpixel is applied as the intensity weight $\boldsymbol{w}^{in}$ to distinguish inner regions of pedestrians from backgrounds:

$$w_i^{in} = \frac{1}{|\boldsymbol{sp}_i|} \sum_{p \in \boldsymbol{sp}_i} I_m(p). \tag{12}$$

Fig. 4(c) shows that the intensity weight is an effective complement to the vertical edge weight. It not only fills the holes caused by the edge weight, but also suppresses the surrounding regions of pedestrians. At last, by integrating the vertical edge weight and the intensity weight, the appearance weight $\boldsymbol{w}$ is defined as:

$$\boldsymbol{w} = \boldsymbol{w}^{ve} + \boldsymbol{w}^{in}. \tag{13}$$

This equation formulates the rule that superpixels with more vertical edges and higher intensity values have higher probabilities of belonging to pedestrians. The effectiveness of the appearance weight is demonstrated by Fig. 4. The vertical edge weight performs well to suppress backgrounds, and the intensity weight can better highlight foregrounds. Thus, the appearance weight improves the intensity contrast to achieve a better saliency detection performance.

C. MUTUAL GUIDANCE BASED SALIENCY PROPAGATION

Previous propagation based saliency models commonly integrate saliency features to obtain the initial saliency map before propagation via summation or multiplication [16], [21]. These integration methods always result in information loss and wrong saliency distribution. Thus, the proposed method introduces mutual guidance based propagation to integrate saliency features and optimize saliency

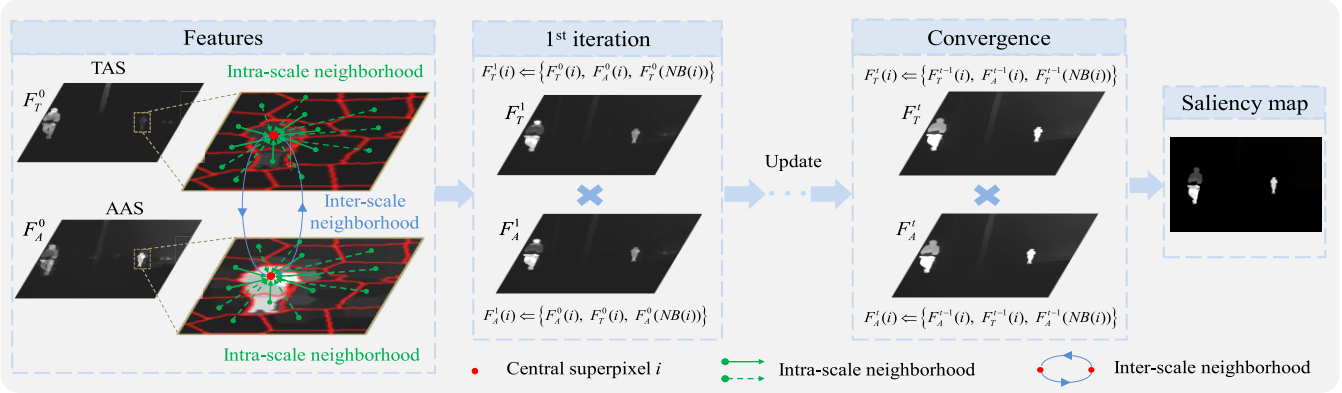


FIGURE 6. Demonstration of the proposed mutual guidance based propagation method: intra-scale neighborhood, inter-scale neighborhood and update rules.

performance simultaneously, which does not need integration before propagation. The propagation method is inspired by cellular automata which consists of three factors: cell, neighborhood and updating rules.

In this paper, each superpixel is taken as a cell. Previous propagation models only use surrounding superpixels to smooth and amend the initial saliency. Different from this, the proposed method propagates saliency scores between not only the neighboring superpixels (intra-scale neighborhood), but also the TAS and AAS feature maps (inter-scale neighborhood).

1) INTRA-SCALE NEIGHBORHOOD

Based on the intuition that neighboring cells are likely to share similar saliency values, the saliency of each cell should be determined by its neighborhood. As shown in Fig. 6, the intra-scale neighborhood of a cell (red dot) is defined as its direct neighboring cells (green dots connected by solid lines) and the direct neighborhood of these cells (green dots connected by dotted lines). Also, neighborhoods that have similar intensities to the central cell should be assigned a large weight on the central cell. Thus, the intensity similarity matrix $\boldsymbol{M} = [m_{ij}]_{N\times N}$ is defined to determine the impact strength of each cell on the central cell:

$$m_{ij} = \begin{cases} \exp(|v_i - v_j|/\sigma^2) & j \in \boldsymbol{NB}(i) \\ 0 & i = j \text{ or otherwise,} \end{cases} \tag{14}$$

where $\boldsymbol{NB}(i)$ is the intra-scale neighborhood of the i-th cell. Fig. 7 shows that pedestrians are continuously highlighted via intra-scale neighborhoods.

2) INTER-SCALE NEIGHBORHOOD

As intra-scale neighborhoods can assimilate neighboring cells, small targets may be wrongly suppressed, as shown in the second column of Fig. 7. With the smoothing effect of intra-scale neighborhoods, pedestrians with small sizes are likely to be assimilated by their surrounding backgrounds because of the low contrast between them. To solve this issue, an inter-scale neighborhood is proposed.

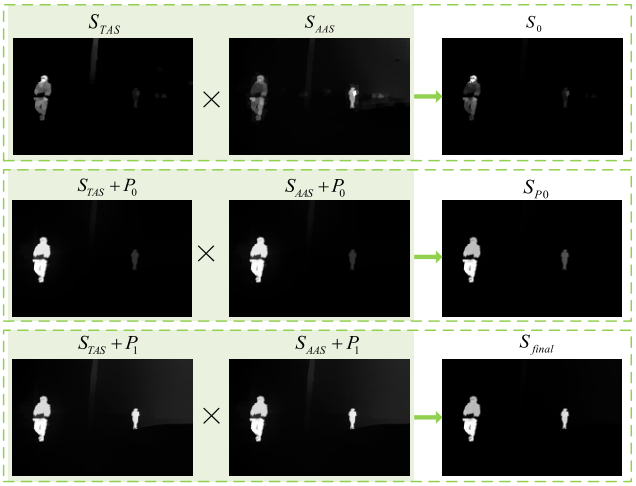


FIGURE 7. Effects of the proposed saliency propagation method. Top:the original results without propagatio; Middle: results of propagation P_0, which only concerns intra-scale neighborhood. Bottom: results of mutual guidance based saliency propagation P_1.

The principle is that the final saliency value of each cell should be approximately consistent with their corresponding values in TAS and AAS. TAS based on MSER can locate the salient regions and suppress backgrounds, but it sometimes unduly suppresses salient regions. AAS based on contrast can enhance the difference between foreground and background, but it cannot strongly suppress the background. Therefore, these two features always complement each other. Then, cells with the same coordinates in the TAS and AAS maps are defined as the inter-scale neighborhood of each other (red dots linked by blue arrows in Fig. 6 are a pair of neighborhoods), so they can amend each other as an aid of intra-scale neighborhoods. Therefore, the state of cell i at the t-th iteration is determined by three parts:

$$sf_1^t(i) \Leftarrow \{sf_1^{t-1}(i), sf_2^{t-1}(i), sf_1^{t-1}(\boldsymbol{NB}(i))\}, \tag{15}$$

$\boldsymbol{sf}_1$ and $\boldsymbol{sf}_2$ are used to represent the TAS and AAS feature maps respectively. Thus for cell i in feature map $\boldsymbol{sf}_1$, its current state is decided by its last state in the two feature maps, and also the last state of its surrounding cells in $\boldsymbol{sf}_2$.

As can be observed from Fig. 7, the pedestrian with a smaller size is wrongly suppressed by the intra-scale neighborhood based propagation, which is represented by P_0. But with the use of inter-scale neighborhoods, the wrongly suppressed pedestrian is recovered and highlighted by P_1. This result shows the effectiveness of inter-scale neighborhood, which makes TAS and AAS guide each other in the process of saliency propagation to further improve the final saliency.

3) UPDATING RULES

To balance the impact strengths of intra-scale and inter-scale neighborhoods on the propagation, a coherence matrix $\boldsymbol{C} = \text{diag}\{c_1^*, c_2^*, \ldots, c_N^*$ is defined in Algorithm 1 from lines 3 to 7. Thus, to propagate the saliency of each cell in TAS and AAS, the updating rules are defined as:

$$\begin{cases} \boldsymbol{F}_T^t = \boldsymbol{F}_T^{t-1} + \underbrace{(\boldsymbol{I}-\boldsymbol{C})\cdot\boldsymbol{M}\cdot\boldsymbol{F}_T^{t-1}}_{\text{Intra-scale}} + \underbrace{\boldsymbol{C}\cdot\boldsymbol{F}_A^{t-1}}_{\text{Intra-scale}} \\ \boldsymbol{F}_A^t = \boldsymbol{F}_A^{t-1} + \underbrace{(\boldsymbol{I}-\boldsymbol{C})\cdot\boldsymbol{M}\cdot\boldsymbol{F}_A^{t-1}}_{\text{Intra-scale}} + \underbrace{\boldsymbol{C}\cdot\boldsymbol{F}_T^{t-1}}_{\text{Intra-scale}} \\ \boldsymbol{F}_T^t = \frac{\boldsymbol{F}_T^t}{\|\boldsymbol{F}_T^t\|}, \boldsymbol{F}_A^t = \frac{\boldsymbol{F}_A^t}{\|\boldsymbol{F}_A^t\|} \\ \boldsymbol{S}^t = \boldsymbol{F}_T^t \cdot \boldsymbol{F}_A^t, \end{cases} \tag{16}$$

where $\boldsymbol{F}_T^t, \boldsymbol{F}_A^t$, and $\boldsymbol{S}^t$ represent the states of TAS, AAS, and the saliency respectively at the i-th iteration. As defined in Eq. (16), the intra-scale neighborhood encourages neighboring cells with higher similarities to take similar saliency scores. If a cell is surrounded by salient cells, the saliency scores of these neighborhoods will be accumulated by the calculation of $\boldsymbol{M}\cdot\boldsymbol{F}_T^{t-1}$. Thus, this cell will become more and more salient through propagation. On the contrary, the background cells will be suppressed. Thus, an intra-scale neighborhood has the ability to smooth and optimize saliency. However, if a salient cell is mostly surrounded by background cells, e.g., pedestrians with smaller sizes, the smoothing effect of intra-scale neighborhoods may wrongly suppress the salient cell. At this time, the proposed inter-scale neighborhood can be used to solve this problem. With the definition of $\boldsymbol{C}$, if a cell has a significant difference with its intra-scale neighborhood, the weight of its inter-scale neighborhood will be larger. Therefore, the saliency value in the next state should be more dependent on its inter-scale neighborhood. Due to the supplementary effect of the two features, the saliency can be amended and improved. To avoid the modulus of saliency features to become too large or too small, they are normalized in each iteration.

It is also important to decide when to stop the iteration. If there are not enough iterations, the propagation cannot achieve an ideal result. Otherwise, if it iterates too many times, there will be unnecessary increase in computational load. And sometimes excessive iterations may make the saliency result worse. Qin *et al.* [21] set the maximum iteration to a fixed value, which is simple but not always suitable for all the images. The complexity of images and the performances of TAS and AAS all affect the propagation,

Algorithm 1 Mutual Guidance Based Saliency Propagation

Input: The TAS and AAS. The intra-scale neighborhood similarity matrix $\boldsymbol{M} = [m_{ij}]_{N\times N}$. The balance parameter σ.

1: $t = 0$
2: Initialize: $\boldsymbol{F}_T^0 = \boldsymbol{S}_{TAS}, \boldsymbol{F}_A^0 = \boldsymbol{S}_{AAS}, check = 1$, $T_{max} = 15$
3: **For** $i \leftarrow 1$ to N **do**
4: $\quad c_i = \frac{1}{\max(m_{ij})} \quad j = 1, 2, \ldots, N$
5: **End for**
6: $\{c_1^*, c_2^*, \ldots, c_N^*\} = \text{Normalize}\{c_1, c_2, \ldots, c_N\}$
7: $\boldsymbol{C} = \text{diag}\{\{c_1^*, c_2^*, \ldots, c_N^*\}$
8: **For** $t \leftarrow 1$ to 3 **do**
9: $\quad \boldsymbol{F}_T^t = \boldsymbol{F}_T^{t-1} + (\boldsymbol{I}-\boldsymbol{C})\cdot\boldsymbol{M}\cdot\boldsymbol{F}_T^{t-1} + \boldsymbol{C}\cdot\boldsymbol{F}_A^{t-1}$
10: $\quad \boldsymbol{F}_T^t = \boldsymbol{F}_T^{t-1} + (\boldsymbol{I}-\boldsymbol{C})\cdot\boldsymbol{M}\cdot\boldsymbol{F}_T^{t-1} + \boldsymbol{C}\cdot\boldsymbol{F}_A^{t-1}$
11: $\quad \boldsymbol{F}_T^t = \frac{\boldsymbol{F}_T^t}{\|\boldsymbol{F}_T^t\|}, \boldsymbol{F}_A^t = \frac{\boldsymbol{F}_A^t}{\|\boldsymbol{F}_A^t\|}$
12: $\quad \boldsymbol{S}^t = \boldsymbol{F}_T^t \cdot \boldsymbol{F}_A^t$
13: **End for**
14: **While** $check > thresh$ to $t \le T_{max}$ **do**
15: $\quad \boldsymbol{F}_T^t = \boldsymbol{F}_T^{t-1} + (\boldsymbol{I}-\boldsymbol{C})\cdot\boldsymbol{M}\cdot\boldsymbol{F}_T^{t-1} + \boldsymbol{C}\cdot\boldsymbol{F}_A^{t-1}$
16: $\quad \boldsymbol{F}_T^t = \boldsymbol{F}_T^{t-1} + (\boldsymbol{I}-\boldsymbol{C})\cdot\boldsymbol{M}\cdot\boldsymbol{F}_T^{t-1} + \boldsymbol{C}\cdot\boldsymbol{F}_A^{t-1}$
17: $\quad \boldsymbol{F}_T^t = \text{Normalize}(\boldsymbol{F}_T^t), \boldsymbol{F}_A^t = \text{Normalize}(\boldsymbol{F}_A^t)$
18: $\quad \boldsymbol{S}^t = \boldsymbol{F}_T^t \cdot \boldsymbol{F}_A^t$
19: $\quad check = \text{var}(\boldsymbol{S}^{t-3}, \boldsymbol{S}^{t-2}, \boldsymbol{S}^{t-1}, \boldsymbol{S}^t)$
20: **End while**
21: $T = t, \boldsymbol{S}_{final} = \text{Normalize}(\boldsymbol{S}^T)$

Output: The final saliency scores $\boldsymbol{S}_{final}$ for each cell.

thus an adaptive termination condition of the iteration is necessary. In this work, the termination of iteration is decided by checking the average variance among the current state and its previous 3 iterations:

$$check = \text{var}(\boldsymbol{S}^{t-3}, \boldsymbol{S}^{t-2}, \boldsymbol{S}^{t-1}, \boldsymbol{S}^t). \tag{17}$$

Considering the propagation mechanism, the propagation would develop a steady local environment of results and come to convergence. Thus, when *check* reaches the threshold T_C, the iteration should stop. Whereas there are cases that the value of *check* is always larger than T_C, the maximum of iteration is empirically set as $T_{max} = 15$.

In summary, the stopping criterion of saliency propagation is defined by the following rules:

- When *chenk* has a value below threshold $T_C - = 10^{-5}$, the iteration will stop.
- When iterations reaches T_{max}, the iteration will stop, regardless of whether *check* has reached T_C.

And after the iteration stops, the final saliency of the proposed method will be $\boldsymbol{S}_{final} = \boldsymbol{S}^T$, where T is the number index of the last iteration.

4) CONVERGENCE ANALYSIS

Since the saliency score is propagated with the similarity estimation, salient parts with a similar appearance in the

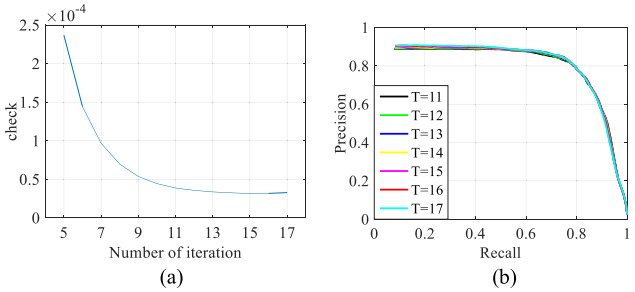


FIGURE 8. (a) Average variance trend in the propagation; (b) Performance evaluation for iterating certain times T on PR curves.

image would naturally merge and enhance each other due to the connectivity and compactness of the object. Moreover, the boundary between an object and the background would become more explicit according to the contrast between different components. Thus, saliency maps would not change any more once the system achieves stability. And the propagation would converge gradually.

To intuitively demonstrate the convergence, experiments are also conducted. As shown in Fig. 8(a), the variances of all images in the dataset DIP are recorded from the 5-th iteration to the 17-th iteration, and then variances are averaged at each iteration. The trend is declining and will gradually flatten to 0.4×10^{-4}, which indicates that the propagation would eventually converge and saliency results would barely change once the number of iterations reaches a sufficiently large value.

To further illustrate the convergence of the propagation, we set the iteration number T to a fixed value for the whole dataset and change it from 11 to 17 to see its influence on the final saliency via the precision-recall (PR) curves [15]. A PR curve is a commonly used evaluation metric which measures the similarity between saliency maps and the ground truth (GT). Fig. 8(b) shows that the performances for different values of T are very similar with each other. This result demonstrates that the saliency performance will eventually stabilize and converge with the growing number of iterations.

Furthermore, we use mathematical inference to prove the convergence of the proposed propagation method. Re-writing the update rules in Eq. (16) into a matrix form, we have

$$\begin{bmatrix} \boldsymbol{F}_T^t \\ \boldsymbol{A}_T^t \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} + (\boldsymbol{I} - \boldsymbol{C}) \cdot \boldsymbol{M} & \boldsymbol{C} \\ \boldsymbol{C} & \boldsymbol{I} + (\boldsymbol{I} - \boldsymbol{C}) \cdot \boldsymbol{M} \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_T^{t-1} \\ \boldsymbol{A}_T^{t-1} \end{bmatrix}. \tag{18}$$

The update rules can be rewritten as a linear recursive sequence:

$$\boldsymbol{u} = \boldsymbol{A}\boldsymbol{u}_{t-1} \quad (t = 1, 2, \ldots), \tag{19}$$

where $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I} + (\boldsymbol{I} - \boldsymbol{C}) \cdot \boldsymbol{M} & \boldsymbol{C} \\ \boldsymbol{C} & \boldsymbol{I} + (\boldsymbol{I} - \boldsymbol{C}) \cdot \boldsymbol{M} \end{bmatrix}_{n \times n}$, $n = 2N$, $\boldsymbol{u}_{t-1} = \begin{bmatrix} \boldsymbol{F}_T^{t-1} \\ \boldsymbol{A}_T^{t-1} \end{bmatrix}$, and $\boldsymbol{u}_t = \begin{bmatrix} \boldsymbol{F}_T^t \\ \boldsymbol{F}_A^t \end{bmatrix}$. And we have

$$\boldsymbol{u}_t = \boldsymbol{A}\boldsymbol{u}_{t-1} = \boldsymbol{A}^2\boldsymbol{u}_{t-2} = \cdots = \boldsymbol{A}^t\boldsymbol{u}_0. \tag{20}$$

As the coherence matrix $\boldsymbol{C}$ is a diagonal matrix, and the similarity matrix is a symmetric matrix, $\boldsymbol{A}$ is also a symmetric matrix. Thus, $\boldsymbol{A}$ has n linearly independent eigenvectors. Sorting the eigenvalues of $\boldsymbol{A}$ in descending order, we have eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and their corresponding eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, satisfying $\boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i (i = 1, 2, \ldots, n)$. If $\boldsymbol{A}$ has only one maximum eigenvalue, they will satisfy the inequality

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n| \tag{21}$$

If $\boldsymbol{A}$ has r multiple maximum eigenvalues, they will satisfy the inequality

$$|\lambda_1| = |\lambda_2| = \ldots = |\lambda_r| > |\lambda_{r+1}| \geq |\lambda_{r+2}| \geq \ldots \geq |\lambda_n| \tag{22}$$

As $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ are linearly independent, there exists only one array $\alpha_1, \alpha_2, \ldots, \alpha_n$ that is not all zero to rewrite $\boldsymbol{u}_0$ as:

$$\boldsymbol{u}_0 = \alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + \ldots + \alpha_n\boldsymbol{x}_n. \tag{23}$$

To avoid the modulus of $\boldsymbol{u}_t$ to become too large or too small, it is normalized in each iteration. Therefore, the normalization could be represented as

$$\begin{cases} \boldsymbol{y}_t = \dfrac{\boldsymbol{u}_t}{\|\boldsymbol{u}_t\|} \\ \boldsymbol{u}_t = \boldsymbol{y}_t \end{cases} \quad (t = 1, 2, \ldots), \tag{24}$$

Substituting Eq. (20) and Eq. (23) into Eq. (24), we have

$$\begin{aligned} \boldsymbol{y}_t &= \frac{\boldsymbol{A}^t\boldsymbol{u}_0}{\|\boldsymbol{A}^t\boldsymbol{u}_0\|} = \frac{\alpha_1\boldsymbol{A}^t\boldsymbol{x}_1 + \alpha_2\boldsymbol{A}^t\boldsymbol{x}_2 + \cdots + \alpha_n\boldsymbol{A}^t\boldsymbol{x}_n}{\|\alpha_1\boldsymbol{A}^t\boldsymbol{x}_1 + \alpha_2\boldsymbol{A}^t\boldsymbol{x}_2 + \cdots + \alpha_n\boldsymbol{A}^t\boldsymbol{x}_n\|} \\ &= \frac{\alpha_1\lambda_1^t\boldsymbol{x}_1 + \alpha_2\lambda_2^t\boldsymbol{x}_2 + \cdots + \alpha_n\lambda_n^t\boldsymbol{x}_n}{\|\alpha_1\lambda_1^t\boldsymbol{x}_1 + \alpha_2\lambda_2^t\boldsymbol{x}_2 + \cdots + \alpha_n\lambda_n^t\boldsymbol{x}_n\|} \\ &= \left(\frac{\lambda_1}{|\lambda_1|}\right)^t \frac{\alpha_1\boldsymbol{x}_1 + \alpha_2\left(\frac{\lambda_2}{\lambda_1}\right)^t\boldsymbol{x}_2 + \cdots + \alpha_n\left(\frac{\lambda_n}{\lambda_1}\right)^t\boldsymbol{x}_n}{\left\|\alpha_1\boldsymbol{x}_1 + \alpha_2\left(\frac{\lambda_2}{\lambda_1}\right)^t\boldsymbol{x}_2 + \cdots + \alpha_n\left(\frac{\lambda_n}{\lambda_1}\right)^t\boldsymbol{x}_n\right\|}. \end{aligned} \tag{25}$$

Considering Eq. (21), when $t \to \infty$, if $\boldsymbol{A}$ as only one maximum eigenvalue, we have

$$\boldsymbol{y}_t = \begin{bmatrix} \boldsymbol{F}_T^t \\ \boldsymbol{F}_A^t \end{bmatrix} \to \begin{cases} \dfrac{\alpha_1\boldsymbol{x}_1}{\|\alpha_1\boldsymbol{x}_1\|} & \lambda_1 > 0 \\ \pm\dfrac{\alpha_1\boldsymbol{x}_1}{\|\alpha_1\boldsymbol{x}_1\|} & \lambda_1 < 0. \end{cases} \tag{26}$$

Letting $\frac{\alpha_1\boldsymbol{x}_1}{\|\alpha_1\boldsymbol{x}_1\|} = \begin{bmatrix} \boldsymbol{s}_1 \\ \boldsymbol{s}_2 \end{bmatrix}$, where $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are both N dimension vectors. Then, we have:

With $\lambda_1 > 0$, we have $\boldsymbol{F}_T^t \to \boldsymbol{s}_1$, $\boldsymbol{F}_A^t \to \boldsymbol{s}_2$, and the final saliency $\boldsymbol{S}^t = \boldsymbol{F}_T^t \cdot \boldsymbol{F}_A^t \to \boldsymbol{s}_1 \cdot \boldsymbol{s}_2$.

With $\lambda_1 < 0$, we have $\boldsymbol{F}_T^t \to \pm\boldsymbol{s}_1$, $\boldsymbol{F}_A^t \to \pm\boldsymbol{s}_2$. As the signs in front of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are always the same, we have $\boldsymbol{S}^t = \boldsymbol{F}_T^t \cdot \boldsymbol{F}_A^t \to \boldsymbol{s}_1 \cdot \boldsymbol{s}_2$ or $\boldsymbol{S}^t = \boldsymbol{F}_T^t \cdot \boldsymbol{F}_A^t \to (-\boldsymbol{s}_1) \cdot (-\boldsymbol{s}_2) = \boldsymbol{s}_1 \cdot \boldsymbol{s}_2$.

Considering Eq. (27), when $t \leftarrow \infty$, if A has multiple maximum eigenvalues, we have

$$\boldsymbol{y}_t = \begin{bmatrix} \boldsymbol{F}_T^t \\ \boldsymbol{F}_A^t \end{bmatrix} \rightarrow \begin{cases} \dfrac{\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \ldots + \alpha_r \boldsymbol{x}_r}{\|\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \ldots + \alpha_r \boldsymbol{x}_r\|} & \lambda_1 > 0 \\ \pm \dfrac{\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \ldots + \alpha_r \boldsymbol{x}_r}{\|\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \ldots + \alpha_r \boldsymbol{x}_r\|} & \lambda_1 < 0 \end{cases} \tag{27}$$

As the linear combination of eigenvectors corresponding to the same eigenvalue is still an eigenvector of that eigenvalue, $\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \ldots + \alpha_r \boldsymbol{x}_r$ is also an eigenvalue of λ_1. Thus, similar to the case that $\boldsymbol{A}$ has only one maximum eigenvalue, we could also have $\boldsymbol{S}^t = \boldsymbol{F}_T^t \cdot \boldsymbol{F}_A^t \rightarrow \boldsymbol{s}_1 \cdot \boldsymbol{s}_2$ by defining $\frac{\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \ldots + \alpha_r \boldsymbol{x}_r}{\|\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2 + \ldots + \alpha_r \boldsymbol{x}_r\|} = \begin{bmatrix} \boldsymbol{s}_1 \\ \boldsymbol{s}_2 \end{bmatrix}$.

Consequently, the saliency score always has a certain limit, which proves the convergence of the proposed propagation method.

III. EXPERIMENTS

A. DATASET AND ANALYSIS

To evaluate the effectiveness of the proposed method, experiments are carried out on three datasets.

1) OSU

Sequences irw01 and irw06, from the public Terravic Motion IR Database in the OTCBVS Benchmark Dataset Collection [36] are used. There are totally 400 images in this dataset and each image contains two pedestrians. This dataset mainly focuses on the changing postures of pedestrians, which is much simple to handle because of its high contrast and flat backgrounds.

2) IMS

This dataset, which is provided by our collaborator, consists of 200 images. There are 39 images containing one pedestrian and other images containing two pedestrians. In this dataset, pedestrians either walk towards or walk away from the camera, so the sizes of pedestrians change greatly. With this dataset, the robustness of the proposed algorithm for pedestrians with different sizes could be verified. Moreover, the images in dataset IMS have a lower contrast than those in dataset OSU.

3) DIP

As the datasets above are relatively simple and only cover a small number of scenes, we construct a more comprehensive dataset to testify the effectiveness of our method. There are 400 infrared images with human-segmented GT in this dataset, which were obtained via the use of a Tau 2 LWIR camera. The complexity of the dataset DIP can be illustrated on the following aspects:

Complex Objects: with multiple pedestrians with diverse postures and sizes. There are totally 634 pedestrians in the dataset, which contains not only 220 images with a single pedestrian but also 180 images with multiple pedestrians. These pedestrians are enormously different from each other in clothing, somatotype, posture and size.

Complex Backgrounds: with diverse composition of background in multiple scenes. There are totally 31 scenes which differ greatly from each other in the dataset DIP. And these scenes have diverse background compositions, including road, sky, buildings, street lamps, trees, brushwood, and other objects.

Based on its complexity and comprehensiveness, dataset DIP is closer to actual scenes and can be better used to examine the robustness of saliency models for infrared pedestrian images.

B. EVALUATION METRICS

In order to evaluate the saliency models, the widely used PR curves [15], F-measure [8], and mean absolute error (MAE) [8] are employed to measure the correctly/wrongly assigned pixels between each image and its corresponding GT among the whole dataset. A good saliency map should achieve a higher PR curve and a larger F-measure value, meanwhile maintaining a low MAE value.

Firstly, to measure the similarity between saliency maps and the GT, precision and recall are defined as:

$$Precision(h) = \frac{|\boldsymbol{BM}(h) \cap \boldsymbol{GT}|}{|\boldsymbol{BM}(h)|}, \tag{28}$$

$$Recall(h) = \frac{|\boldsymbol{BM}(h) \cap \boldsymbol{GT}|}{|\boldsymbol{GT}|}, \tag{29}$$

where $\boldsymbol{BM}(h)$ is the binary mask obtained by binarizing the saliency map with threshold h, and h is a set of integer values from 0 to 255. Then under the same h, precision or recall values are averaged among the dataset to estimate the percentage of correctly assigned pixels.

Secondly, as precision and recall measure the saliency performance from different point of views, the F-measure is used to obtain the combination of them:

$$F - measure = \frac{(1 + \beta^2) \times Precision \times Recall}{\beta^2 \times Precision + Recall}. \tag{30}$$

For saliency detection, precision is a measurement of correctness, which evaluates the percentage of actual salient regions in detected regions, and recall is a measurement of coverage rate, which focuses on how many salient regions are detected. As saliency detection is usually used to automatically locate salient objects, it is more important to determine whether the salient regions are correctly located than whether each salient region is totally detected. Moreover, 100% recall can be easily achieved by setting the whole region to foreground [37]. Thus, precision is more important than recall in saliency detection. As suggested by existing saliency detection methods [8], [10], [12], β^2 is set to 0.3 to bias toward the precision rate.

Lastly, MAE is used as a complement of PR curves and F-measure to measure the pixel-wise error between the

saliency map and GT:

$$MAE = \frac{1}{|\boldsymbol{S}|}\sum_{p\in\boldsymbol{S}}|S(p) - GT(p)|\,, \tag{31}$$

where $S(p)$ denotes the saliency value of pixel p.

C. PARAMETER ANALYSIS

To choose the appropriate parameters for our model, we use a sub set (20%) of dataset DIP as the validation set to tune the parameters N, T_M, σ^2 and T_C.

1) PARAMETER N OF SLIC

N controls the number of superpixels. If N is too small, SLIC might wrongly merge targets and background into the same superpixel. If N is too large, objects would be segmented into many superpixels, which not only increases the computational load, but also loses the ability for noise suppression. To choose a suitable value for N, the experiment is conducted by varying N from 400 to 900. Fig. 9(a) shows that the proposed method performs the best in PR curves when N is set as 700. Hence, we used 700 as the optimal value for N in all subsequent experiments.

2) PARAMETER T_M OF TAS

T_M is the threshold for the generation of MSER. Extremal regions with stableness Ψ lower than T_M will be taken as MSER. To decide the value for T_M, we vary T_M from 0.05 to 0.3 in the experiment. Fig. 9(b) shows the variation of saliency performance for different values of T_M. We can see that the performance increases following the decrease of T_M until $T_M = 0.1$. Actually, if T_M is too large, many background regions will be taken as MSER. This would produce wrongly highlighted background regions. If T_M is too small, the number of MSER will reduce and parts of pedestrians may be missed. Thus, T_M is set as 0.1.

3) PARAMETERS σ^2 AND T_C OF THE MUTUAL GUIDANCE BASED SALIENCY PROPAGATION

σ^2 is the parameter in Eq. (14), which controls the similarity between neighboring cells. Fig. 9(c) shows the variation of the final saliency performance with different values of σ^2. Obviously, the final saliency obtains the best PR performance when σ^2 is approximately set as 0.1. Also, the performance becomes better with the increase of σ^2 when it is smaller than 0.1. Then, the performance becomes worse with the increase of σ^2. Consequently, we set $\sigma^2 = 0.1$ in this paper. T_C is a threshold parameter, which decides when to stop the saliency propagation. The smaller the T_C is, the larger number of iterations the propagation might need. This may lead to the unnecessary increase in computational load. If T_C is too large, the smaller number of iteration will result in low performance with propagation. It is shown in Fig. 9(d) that the method performs the best when $T_C = 1 \times 10^{-5}$ in the largest range of recall. Thus, T_C is set as 1×10^{-5}.

D. EVALUATIONS OF MODEL COMPONENTS

In this section, a series of experiments are presented to investigate the influence of various factors on the proposed saliency model.

1) INTENSITY FILTER

To eliminate the effect of atmosphere from the radiation of pedestrians, the intensity filter is introduced, which subtracts the average intensity of infrared images from each superpixel. To explore the best functional form of the intensity, we define the intensity filter as:

$$IF(i) = \Psi\left(\frac{\sum_{p\in\boldsymbol{sp}_i} I_m(p)}{|\boldsymbol{sp}_i|} - I_\mu\right). \tag{32}$$

Ψ is defined as $\Psi = |\cdot|^2$ in this paper, which performs better compared with other forms of the intensity filter. In the top row of Fig. 10, $\boldsymbol{F}_s$ represents the superpixel-based stableness without an intensity filter, which is defined in Eq. (5). $\Psi = \exp()$, $\Psi = |\cdot|^1$, and $\Psi = |\cdot|^2$ represent the exponential function, linear function, and quadratic function respectively. It is obvious that the saliency performance of stableness is effectively improved by an intensity filter. Moreover, the MAE, PR curves, and F-measure all demonstrate that the quadratic function is superior to the others. Hence, a quadratic function is selected for the intensity filter, which is shown in Eq. (7).

2) APPEARANCE WEIGHT

To measure the availability of each part of the appearance weight, the experiment is designed by comparing their corresponding saliency performances. In the middle row of Fig. 10, $\boldsymbol{Con}$ represents the intensity contrast of Eq. (8). $\boldsymbol{Con}\cdot\boldsymbol{w}^{ve}$ is the vertical edge weighted contrast. $\boldsymbol{Con}\cdot\boldsymbol{w}^{in}$ is the intensity weighted contrast. And contrast with both $\boldsymbol{w}^{ve}$ and $\boldsymbol{w}^{in}$ is the result $\boldsymbol{S}_{AAS}$ of Eq. (9). It can be found that $\boldsymbol{Con}\cdot\boldsymbol{w}^{in}$ performs better in all the evaluation metrics than $\boldsymbol{Con}$, which verifies the effect of intensity weight. $\boldsymbol{Con}\cdot\boldsymbol{w}^{ve}$ performs worse in PR curves and F-measure. That is because vertical edge weight aims at emphasizing the superpixels containing vertical edges, which suppresses the regions inside pedestrians. However, $\boldsymbol{Con}\cdot\boldsymbol{w}^{ve}$ performs the best in MAE, which demonstrates the effectiveness of vertical edge weight in suppressing backgrounds. Besides, $\boldsymbol{S}_{AAS}$ is better than both $\boldsymbol{Con}\cdot\boldsymbol{w}^{in}$ and $\boldsymbol{Con}\cdot\boldsymbol{w}^{ve}$. This shows the effectiveness of appearance weight and the complementary effect between vertical edge weight and intensity weight.

3) EFFECTIVENESS OF PROPAGATION

This experiment is designed to verify the effectiveness of the saliency propagation model and the contribution of inter-scale neighborhood on the saliency result. P_0 is the propagation without inter-scale neighborhood, and P_1 is the proposed mutual guidance based saliency propagation method. The original saliency is defined as $\boldsymbol{S}_0 = \boldsymbol{S}_{TAS} \times \boldsymbol{S}_{AAS}$. Thus, $\boldsymbol{S}_0$,

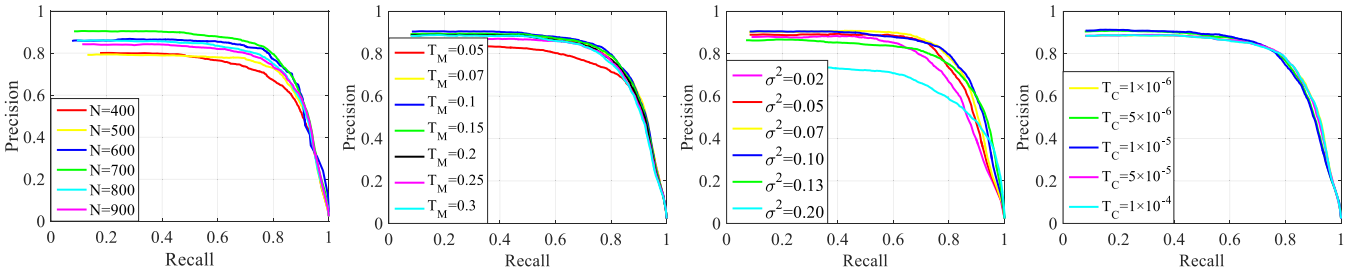


FIGURE 9. Saliency comparisons for parameter analysis with PR curves. (a). on N. (b) on T_M. (c) on σ^2. (d) on T_C.

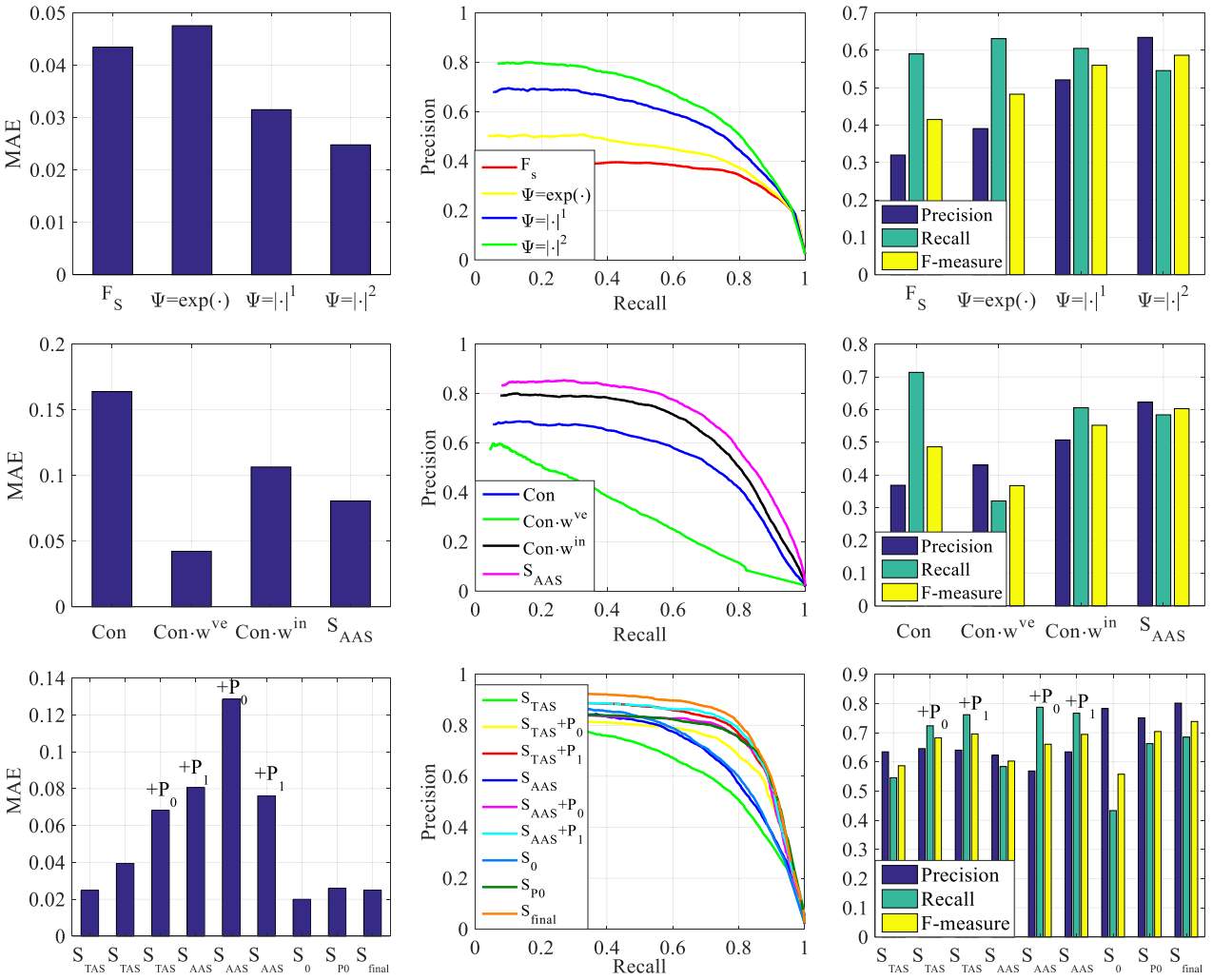


FIGURE 10. Performance evaluations of model components with MAE, PR curves and F-measure. Top: the effects of different intensity filters. Middle: the effectiveness of appearance and each part of it. Bottom: the mutual guidance based saliency propagation.

S_{P0}, and S_{final} represent the final saliency without propagation, with the propagation P_0 and with the propagation P_1 respectively.

Firstly, with the P_1 propagation, S_{TAS} and S_{AAS} are both improved to a better performance, which is shown in the bottom row of Fig. 10. It is worth noting that, although S_{AAS} performs better than S_{TAS}, P_1 could improve their saliency performance to a similar state. Furthermore, the final saliency S_{final} is also better than S_{TAS} and S_{AAS}. These results demonstrate the effectiveness of the mutual guidance based propagation on improving the saliency performance to a certain degree.

To show the contribution of inter-scale neighborhood through quantitative analysis, the performance of saliency propagation with only intra-scale neighborhood P_0 is compared with P_1 which concerns both intra-scale and

inter-scale neighborhoods. It is worth noting that with the promotion of P_1, TAS, AAS, and the final saliency are all improved to better performances than P_0, which is obvious in Fig. 10. These facts illustrate the contribution of the inter-scale neighborhood to complement the intra-scale neighborhood and improve the saliency performance.

Moreover, as can be seen from Fig. 10, the final saliency is better than the results of both TAS and AAS after propagation.

E. COMPARISON WITH STATE-OF-THE-ART SALIENCY MODELS

Following previous saliency models for infrared images [24]–[28], the proposed saliency detection method is first compared with 10 state-of-the-art saliency models: FT [8], CA [11], GS [38], BD [39], BSCA [21], MAP [40], MB+ [41], RS [17] and HCA [22]. And the experiments are carried out on three datasets, OSU, IMS, and DIP.

1) SUBJECTIVE COMPARISONS

Some saliency maps of the proposed method and state-of-the-art methods are shown in Fig. 11, which directly present the visual comparisons. For data set OSU, we can see that most saliency models effectively handle the second image. This is because the second image has a relatively simple background and a high contrast between pedestrians and backgrounds. However, it can be found that all the methods except FT and our method fail in the first image, because salient objects are defaulted to be close to the center of an image in most saliency detectors. And it is difficult to detect pedestrians near the border. Moreover, it is obvious that the proposed method better suppresses the background than FT.

For dataset IMS, most of the state-of-the-art methods perform badly, where pedestrians cannot even be recognized. That is because the ground and trees along the road have high intensities similar to pedestrians. HCA could accurately highlight the pedestrians, while parts of the background are also wrongly highlighted. CA could highlight the contour of pedestrians and partly suppress the background. But the blurring contour and wrongly highlighted background make CA a bad saliency detector. The proposed method could highlight the whole region of pedestrians and suppress the background at the same time.

For dataset DIP, saliency detection is more difficult than the two other datasets. Almost all these methods have the ability of separating pedestrians from background in the third image of Fig. 11, where the pedestrians have relatively lager sizes. But they fail in the other images with pedestrians of small sizes. Because the state-of-the-art saliency models are all tested on datasets with larger salient objects. BD and HCA can separate pedestrians from the background in most images, but the noises in background cannot be suppressed efficiently. Note that the proposed method could highlight pedestrians regardless of the size of pedestrians and has better saliency value distributions.

Therefore, the proposed method achieves good saliency detection performance for infrared pedestrian images superior to the other state-of-the-art methods.

2) OBJECTIVE COMPARISONS

We further objectively compare different saliency models using PR curves, F-measure, and MAE. For dataset OSU, it is obvious in Fig. 12 that the proposed method achieves similar performances to the BD method on PR curves and F-measure. However, it is noteworthy that our saliency model has the lowest MAE value 0.01, which is smaller than all other methods. This fact demonstrates the effectiveness of our method on background suppression.

For dataset IMS, the proposed method is superior to the other state-of-the-art methods in all the evaluation metrics. The proposed method achieves the highest precision in almost all the recall range [0, 1] up to 0.95, while the precisions of all the other methods are lower than 0.2. Also, the proposed method performs the best in MAE and F-measure. This is in accord with the visual performance that the dark sky region tends to be taken as a salient region, while the pedestrian regions are totally suppressed with these compared methods.

For dataset DIP, the proposed method achieves the best performance than other methods, which attains the highest precision in almost all the recall range [0, 1] up to 0.91, while all the other method are lower than 0.7. HCA could only obtain a high precision value when its recall is small, because the background noise cannot be well suppressed by HCA. The F-measure value of the proposed method is also higher than the others. And it achieves the lowest MAE value, which is close to 0.03. These results demonstrate the ability of the proposed method to highlight pedestrians and suppress backgrounds.

All the experiments illustrate the superiority of the proposed method and also its robustness to the complexity of images.

F. COMPARISON WITH SALIENCY MODELS OF INFRARED PEDESTRIAN IMAGES

Actually, the above state-of-the-art saliency models are designed for visible images, and saliency detection of infrared images has not been extensively studied. To comprehensively demonstrate the superiority of the proposed saliency model, four saliency models which are designed for infrared pedestrian images are applied as comparison. These methods include LSM [25], AS [26], CD [27], MCS [28], and BO [1] which have been introduced in Section I. The experiments are also conducted on all the three datasets, OSU, IMS, and DIP.

1) SUBJECTIVE COMPARISONS

Fig. 13 shows saliency maps of the proposed method and the other infrared saliency models mentioned above. It is obvious that the proposed method achieves the best performance. We can find that LSM CD, and MCS attempt to obtain edges of pedestrians, while AS, BO and the proposed method try to highlight the whole pedestrians.

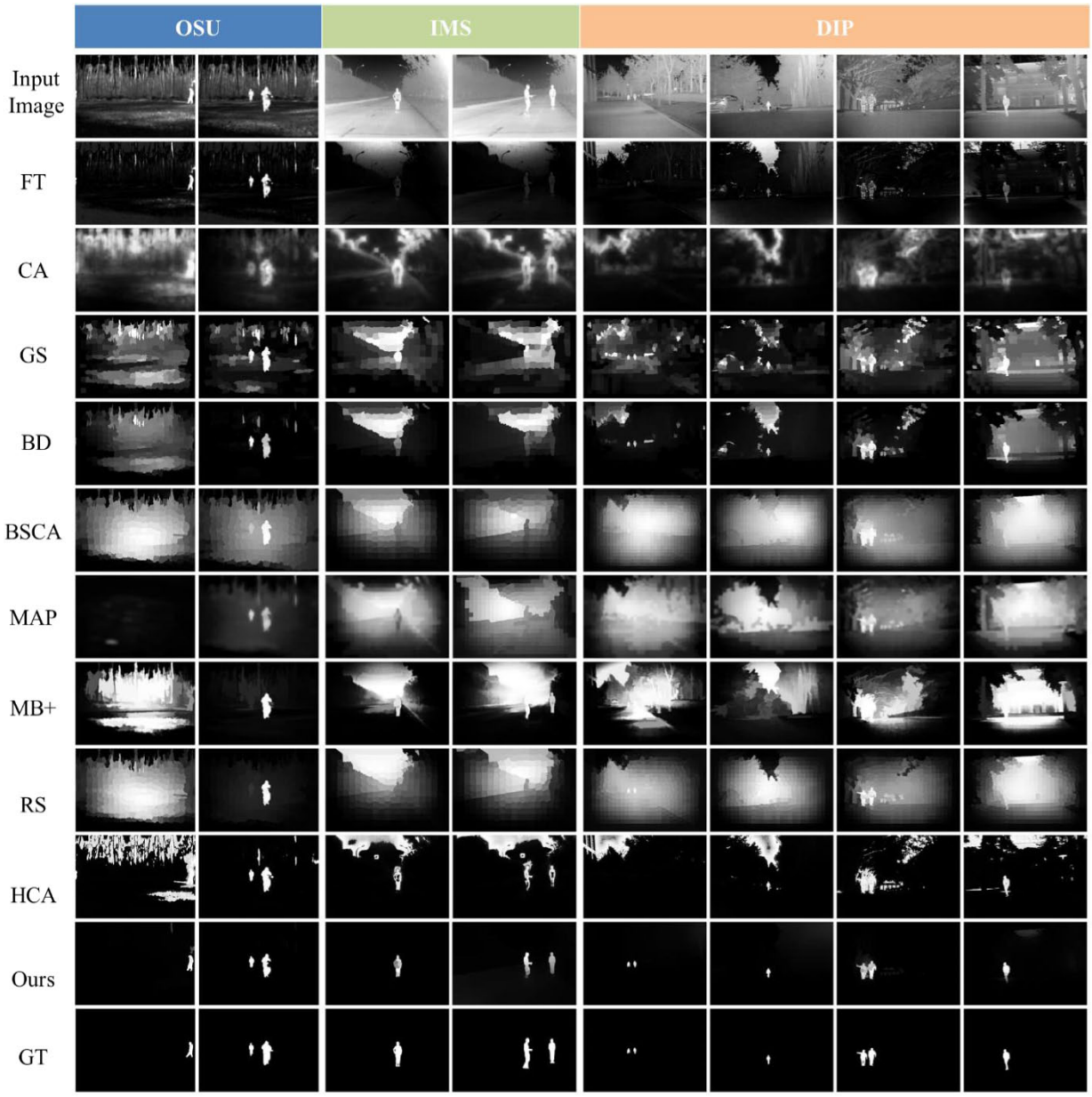


FIGURE 11. *Visual comparison on the three datasets OSU, IMS and DIP among the proposed method and 10 state-of-the-art saliency detection methods.*

For dataset OSU, as the background of images in this dataset consists of bushes, their abundant textures lead to the failure of LSM to separate pedestrians from the background. MCS has a better ability to suppress the noises of background, whereas CD performs better than LSM and MCS in highlighting pedestrians. Different from the above methods, AS, BO and the proposed method can highlight pedestrians as a complete region. And the proposed method suppresses background more effectively.

For dataset IMS, the background is more homogeneous than OSU, so edges in background are better suppressed in saliency maps for LSM, CD and MCS. But the wrongly highlighted edges in background and regions inside pedestrians all indicate the poor performance of these methods. AS could separate pedestrians from background and highlight each pedestrian as a whole region. However, the high intensity corners are wrongly distributed with high saliency values by AS. BO performs poorly in the second image, because BO first needs an object detection method to locate pedestrians, and then calculates the saliency of pedestrians in the detected regions marked by rectangle boxes. Its performance of saliency detection heavily depends on the accuracy of detection methods. The proposed method needs no pre-detection but accurately locates the pedestrians via saliency.

For dataset DIP, the complex background of this dataset still results in poor performance of LSM and MCS. CD and

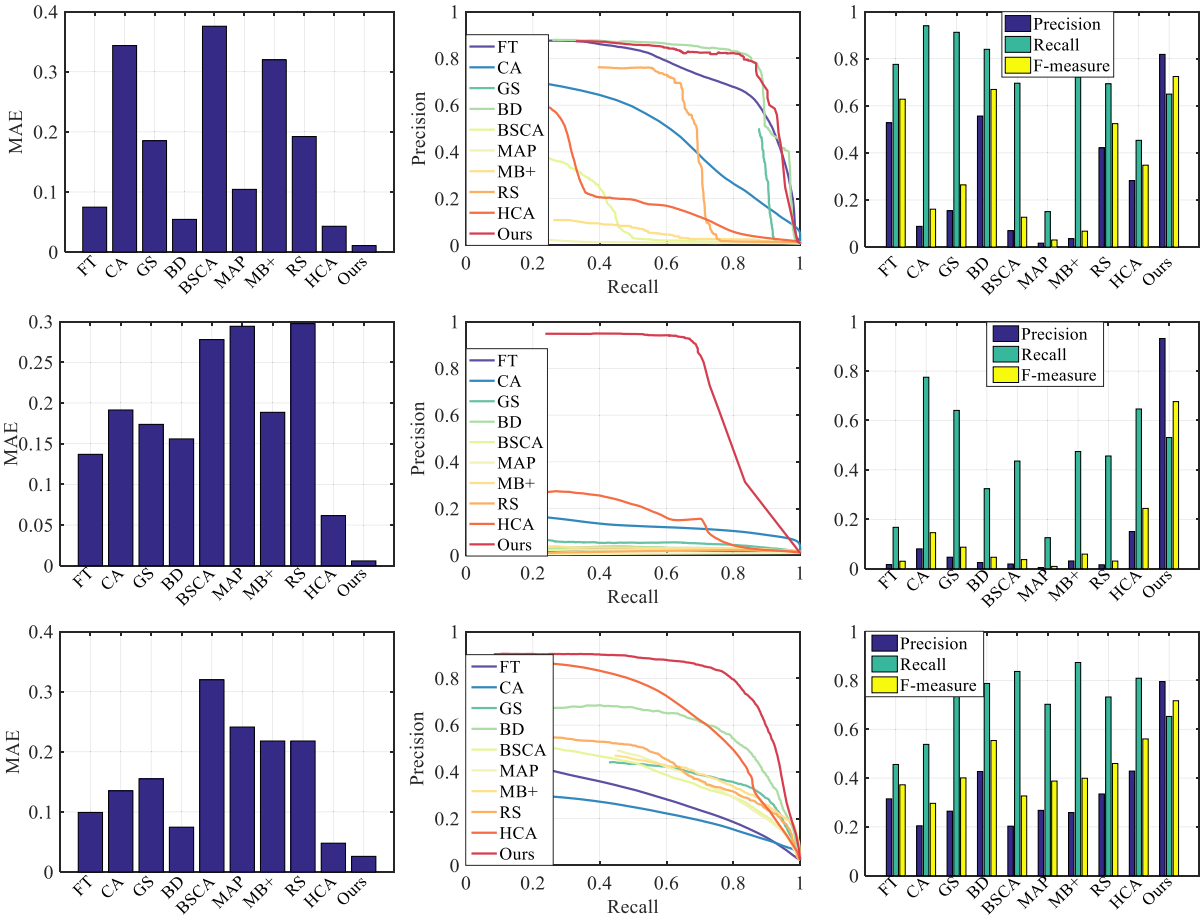


FIGURE 12. Objective comparison among the proposed method and 10 state-of-the-art methods with MAE, PR curves and F-measure. From top to the bottom are on datasets OSU, IMS, and DIP.

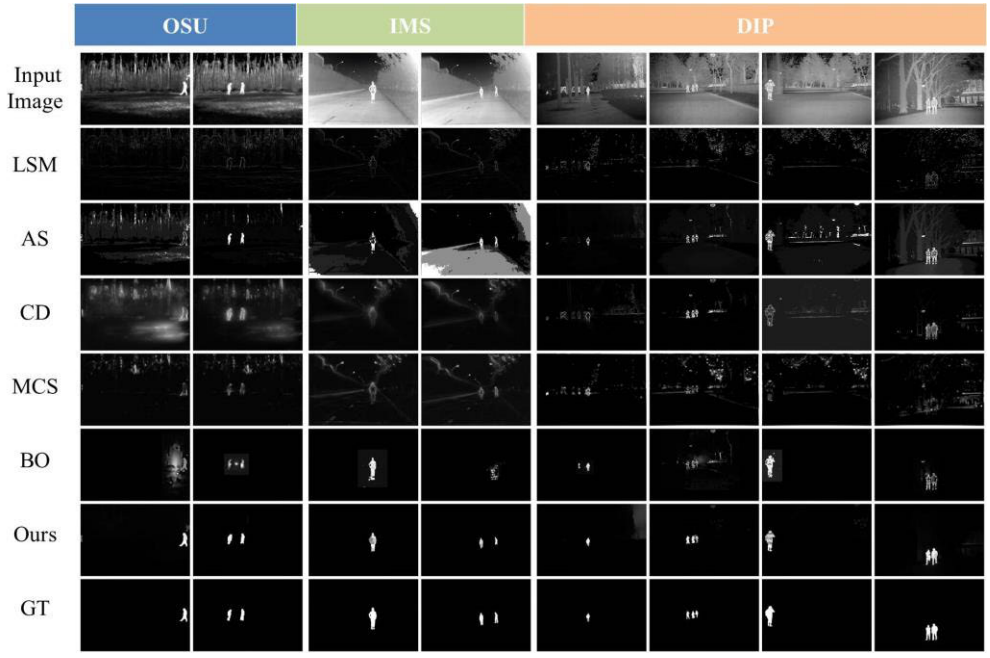


FIGURE 13. Visual comparison on the three datasets OSU, IMS and DIP between the proposed method and 4 saliency models of infrared pedestrian images.

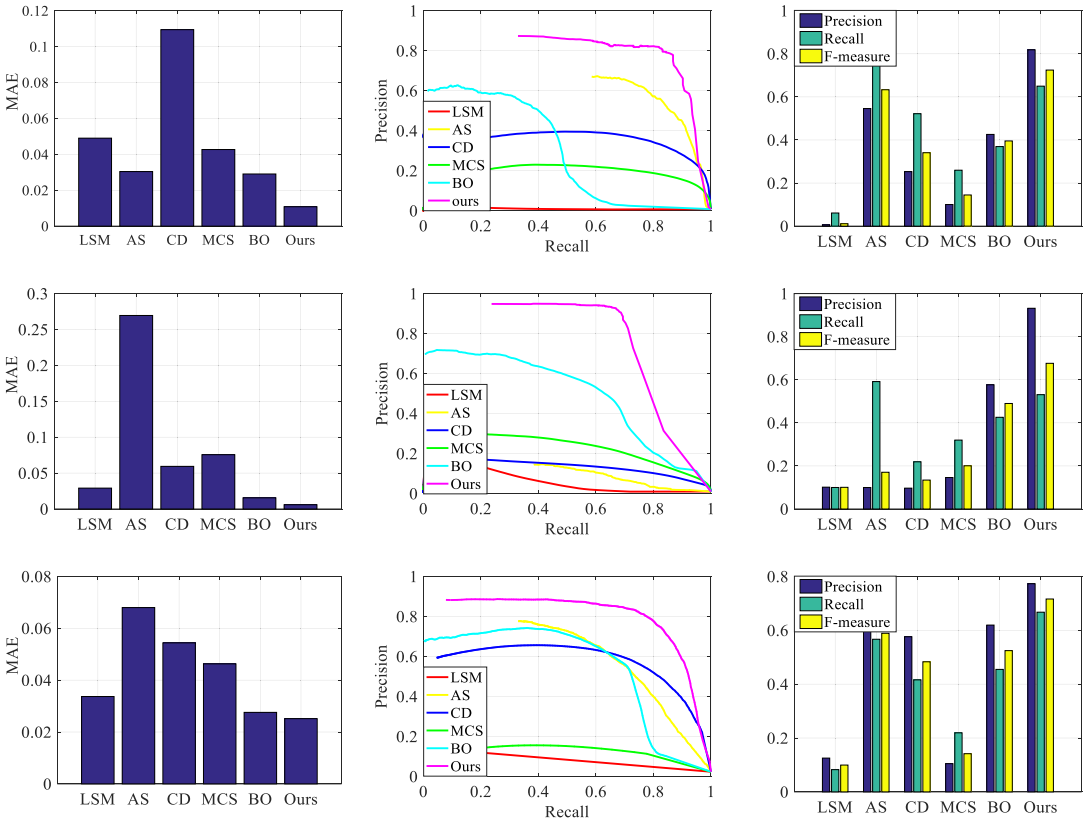


FIGURE 14. Objective comparison among the proposed method and 4 saliency models of infrared images with MAE, PR curves, and F-measure. From top to the bottom are on datasets OSU, IMS, and DIP.

BO perform much better than LSM and MCS. CD even has a much better ability than AS to suppress the background, while it cannot highlight the inner part of pedestrians. BO performs well in some images, but there still exists wrong saliency distribution and missed detection due to the inaccuracy of pre-detection. Comparing with the five methods above, the proposed method performs much better in suppressing background and highlighting pedestrians.

2) OBJECTIVE COMPARISONS

PR curves, F-measure, and MAE are also used to compare the objective performance of the proposed method and other models designed for infrared pedestrian images. For dataset OSU, it is obvious that the proposed method is superior to the other methods in all the datasets and evaluation metrics. AS and BO achieve comparative performance to the proposed method. This fact is consistent with their subjective performances, which also verifies the effectiveness of the evaluation metrics. The proposed method achieves the highest precision close to 0.9, while the precisions of the other methods apart from AS and BO are lower than 0.4. Meanwhile, the proposed method obtains the lowest MAE close to 0.01 and the highest F-measure up to 0.7.

For dataset IMS, we can see from Fig. 14 that the proposed method achieves the lowest MAE value 0.01. The F-measure value of the proposed method is close to 0.7, while the F-measure for all the other models are even lower than 0.2. It is noteworthy that, the highest precision in the PR curve of the proposed method is up to 0.95, while the PR curves for others are lower than 0.3.

For dataset DIP, the proposed method achieves the highest precision 0.9, the lowest MAE 0.027 and the highest F-measure 0.72. We can find that AS and CD perform much better than LSM and MCS in PR curve and F-measure, while they perform worse in MAE. This is because AS and CD not highlight only pedestrians but also the background. So, superior performance of the proposed method in both PR curves and MAE verifies its ability to highlight pedestrians and suppress background.

In summary, the proposed method performs much better than previous saliency models designed for infrared pedestrian images.

G. RUN TIME COMPARISONS

The run time experiment is performed via MATLAB 2015b on an Intel i5-3450 (3.10GHz) CPU with 8 GB RAM. The proposed method is compared with both state-of-the-art saliency models and saliency models for infrared pedestrian images in Table 1. Our method is slower than most of the others because of the calculation for the vertical edge weight, which consists of PB algorithm-based boundary map extraction and the edge weight calculation. The PB algorithm takes

TABLE 1. Run time comparisons.

Method	State-of-the-art saliency models									Saliency models for infrared pedestrians					
	FT	CA	GS	BD	BSCA	MAP	MB+	RS	HCA	LSM	AS	CD	MCS	BO	Ours
Time (s)	0.36	16.71	2.11	2.72	5.60	1.21	0.94	1.78	4.63	1.87	7.41	30.76	11.56	40.5	24.49
Code	EXE	M+C	M+C	M	M+C	M	BAT	M+C	M+C	M	M	M	M	M	M

more than half of the total time, and the calculation of edge weight needs to traverse all the superpixels, which also takes lots of time. However, better results could be obtained as shown in the above comparisons of performances, at the cost of more computational time.

IV. CONCLUSION

In this paper, by analyzing the thermal and appearance characteristics of infrared pedestrian images, a novel saliency detection method for infrared pedestrian images is proposed. Two features on thermal analysis-based saliency and appearance analysis-weighted saliency are first proposed. And then, a mutual guidance based saliency propagation method is introduced to facilitate the two features and improve the final saliency. We have also built two datasets DIP and IMS with 600 infrared pedestrian images, and have made them available to the public. All the experiments on three infrared pedestrian datasets demonstrate the effectiveness of the proposed method.

REFERENCES

[1] L. Li, F. Zhou, and X. Bai, "Infrared pedestrian segmentation through background likelihood and object-biased saliency," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 9, pp. 2826–2844, Sep. 2018.

[2] L. Marchesotti, C. Cifarelli, and G. Csurka, "A framework for visual saliency detection with applications to image thumbnailing," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 2232–2239.

[3] H. Kuang, K.-F. Yang, L. Chen, Y.-J. Li, L. L. H. Chan, and H. Yan, "Bayes saliency-based object proposal generator for nighttime traffic images," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 814–825, Mar. 2018.

[4] X. Yang, X. Qian, and Y. Xue, "Scalable mobile image retrieval by exploring contextual saliency," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1709–1721, Jun. 2015.

[5] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.

[6] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Salient region detection by modeling distributions of color and orientation," *IEEE Trans. Multimedia*, vol. 11, no. 5, pp. 892–905, Aug. 2009.

[7] C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, and D. A. Clausi, "Statistical textural distinctiveness for salient region detection in natural images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogit.*, Jun. 2013, pp. 979–986.

[8] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogit.*, Jun. 2009, pp. 1597–1604.

[9] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. 14th ACM Int. Conf. Multimedia*, Oct. 2006, pp. 815–824.

[10] K. Shi, K. Wang, J. Lu, and L. Lin, "PISA: Pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2115–2122.

[11] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.

[12] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 733–740.

[13] Z. Ren, S. Gao, L.-T. Chia, and I. W.-H. Tsang, "Region-based saliency detection and its application in object recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 5, pp. 769–779, May 2014.

[14] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1529–1536.

[15] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing Markov chain," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep. 2013, pp. 1665–1672.

[16] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs to model saliency in images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1698–1705.

[17] L. Zhang, C. Yang, H. Lu, R. Xiang, and M.-H. Yang, "Ranking saliency," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 9, pp. 1892–1904, Sep. 2017.

[18] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, Dec. 2006, pp. 545–552.

[19] C. Li, Y. Yuan, W. Cai, Y. Xia, and D. D. Feng, "Robust saliency detection via regularized random walks ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2710–2717.

[20] H. Li, H. Lu, Z. Lin, X. Shen, and B. Price, "Inner and inter label propagation: Salient object detection in the wild," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3176–3186, Oct. 2015.

[21] Y. Qin, H. Lu, Y. Xu, and H. Wang, "Saliency detection via cellular automata," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 110–119.

[22] Y. Qin, M. Feng, H. Lu, and G. W. Cottrell, "Hierarchical cellular automata for visual saliency," *Int. J. Comput. Vis.*, vol. 126, no. 7, pp. 751–770, 2018.

[23] X. Bai, P. Wang, and F. Zhou, "Pedestrian segmentation in infrared images based on circular shortest path," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2214–2222, Aug. 2016.

[24] J. Zhao, Y. Chen, H. Feng, Z. Xu, and Q. Li, "Fast image enhancement using multi-scale saliency extraction in infrared imagery," *Optik*, vol. 125, no. 15, pp. 4039–4042, Aug. 2014.

[25] B. Ko, D. Kim, and J. Nam, "Detecting humans using luminance saliency in thermal images," *Opt. Lett.*, vol. 37, no. 20, pp. 4350–4352, Oct. 2012.

[26] L. Zhang, Y. Zhang, W. Wei, and Q. Meng, "An associative saliency segmentation method for infrared targets," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 4264–4268.

[27] L. Li, Y. Zheng, and F. Zhou, "Contrast and distribution based saliency detection in infrared images," in *Proc. IEEE Int. Workshop Signal Process.*, Oct. 2015, pp. 1–6.

[28] X. Wang, C. Ning, and L. Xu, "Saliency detection using mutual consistency-guided spatial cues combination," *Infr. Phys. Technol.*, vol. 72, pp. 106–116, Sep. 2015.

[29] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

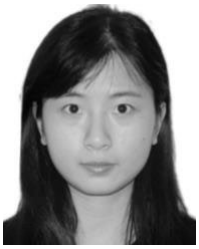
[30] M. Donoser and H. Bischof, "Efficient maximally stable extremal region (MSER) tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 553–560.

[31] J. R. Schott, R. V. Raqueno, and C. Salvaggio, "Incorporation of a time-dependent thermodynamic model and a radiation propagation model into IR 3D synthetic image generation," *Opt. Eng.*, vol. 31, no. 7, pp. 1505–1516, Jul. 1992.

[32] F. Xu, X. Liu, and K. Fujimura, "Pedestrian detection and tracking with night vision," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 1, pp. 63–71, Mar. 2005.

[33] *Physics of Emissivity*. Accessed: 2016. [Online]. Available: http://www.optotherm.com/emiss-physics.htm

[34] A. Fernández-Caballero, M. T. López, and J. Serrano-Cuerda, "Thermal-infrared pedestrian ROI extraction through thermal and motion information fusion," *Sensors*, vol. 14, pp. 6666–6676, Apr. 2014.
[35] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 530–549, May 2004.
[36] *OTCBVS Benchmark Dataset Collection*. Accessed: 2005. [Online]. Available: http://vcipl-okstate.org/pbvs/bench/
[37] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A benchmark," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5706–5722, Dec. 2015.
[38] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 29–42.
[39] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2814–2821.
[40] J. Sun, H. Lu, and X. Liu, "Saliency region detection based on Markov absorption probabilities," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1639–1649, May 2015.
[41] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech, "Minimum barrier salient object detection at 80 FPS," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1404–1412.
[42] Y. Zheng, F. Zhou, L. Li, and X. Bai, "Propagation based saliency detection for infrared pedestrian images," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2017, pp. 1527–1531.

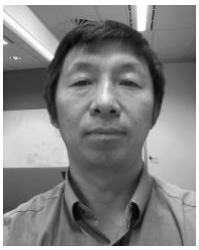


YU ZHENG received the B.S. degree from Beihang University, in 2015, where she is currently pursuing the Ph.D. degree. Her research interests include saliency detection, image segmentation, and analysis.

FUGEN ZHOU received the B.S. degree in electronic engineering and the M.S. and Ph.D. degrees in pattern recognition and intelligent systems from Beihang University, Beijing, China, in 1986, 1989, and 2006, respectively. His research interests include target detection and recognition, multimodality image processing, and biomedical image processing and recognition.

LU LI received the B.S. degree in automation from Wuhan University, in 2005, and the M.S. and Ph.D. degrees in pattern recognition and intelligent systems from Beihang University, in 2008 and 2018, respectively. Her research interests include saliency detection, and image segmentation and analysis.

XIANGZHI BAI received the B.S. and Ph.D. degrees from Beihang University, in 2003 and 2009, respectively, where he is currently a Full Professor with the Image Processing Center and the State Key Laboratory of Virtual Reality Technology and Systems. He holds 12 national invention patents and has published over 100 international journal and conference papers in the field of fuzzy theory, mathematical morphology, image analysis, pattern recognition, and bioinformatics. He also acts as an Active Reviewer for around 60 international journals and conferences.



CHANGMING SUN received the Ph.D. degree in computer vision from Imperial College London, London, U.K., in 1992. He joined CSIRO, Sydney, Australia, where he is currently a Principal Research Scientist, carrying out research and working on applied projects. His current research interests include computer vision, image analysis, and pattern recognition. He has served on the program organizing committees of various international conferences. He is an Associate Editor of the *EURASIP Journal on Image and Video Processing*.