Reconstructed Saliency for Infrared Pedestrian Images

LU LI, (Member, IEEE), FUGEN ZHOU, (Member, IEEE), YU ZHENG, (Student Member, IEEE), AND XIANGZHI BAI, (Member, IEEE)
Image Processing Center, Beihang University, Beijing 100191, China

Corresponding authors: Lu Li (lilu@buaa.edu.cn) and Xiangzhi Bai (jackybxz@buaa.edu.cn)

ABSTRACT Accurately and completely detecting infrared pedestrian is a challenging problem in an intelligent transportation system due to the low SNR and the inhomogeneous luminance distribution in the infrared images, especially for the complex background environment. In this paper, we introduce a reconstruction optimization-based saliency detection method for infrared pedestrian images to solve the problem. First, appearance-based infrared saliency was introduced to enhance the salient areas of the infrared images from locally and globally contrast features. Then, considering the essential characteristic of the infrared pedestrian images, thermal radiation prior, and pedestrian shape prior was combined to construct infrared object prior information. Finally, the infrared pedestrian saliency map can be calculated through a random walk-based saliency reconstruction optimization method with the appearance saliency and infrared object prior. The extensive experiments on real infrared images captured by intelligent transportation systems demonstrate that our saliency algorithm consistently outperforms the state-of-the-art saliency detection methods, in terms of higher precision, F-measure, and lower mean absolute error.

INDEX TERMS Visual attention, reconstruction optimization, infrared image, infrared pedestrian, saliency detection.

I. INTRODUCTION

Accurately and completely detect the pedestrians in infrared images is an important problems in computer vision and image processing, which can be widely applied in road security, autonomous vehicle and advanced driver assistance systems. However, due to the most essential imaging mechanism of infrared image, it is still a challenging problem right now.

Unlike visible light images, infrared images have no color information. Their luminance distribution is mainly affected by the thermal radiation characteristics of the target and background in the image. The intrinsic resolution of the infrared spectrum, the atmospheric absorption, and scattering effects during transmission make the infrared image lack high contrast, high resolution and rich texture. Traditional object detection algorithms based on visible features cannot describe the difference between the target and the background in the infrared image. Thus, most of the traditional object detection methods, such as threshold-based methods 1,

edge-based methods [1], [2] cluster-based methods [3], [4] and methods based on some optimization theories [5]–[7] are difficult to achieve good results in the infrared pedestrian images. Besides, in the infrared pedestrian image, the background is always complicated by the influence of trees, electric poles, street lamps, which may have similar luminance distribution and shape structure with pedestrians in infrared images. What's more, the occlusion effect, the pose variation and the luminance variations greatly increase the difficulty of infrared pedestrian detection.

Different with traditional object detection, saliency detection can simulate human vision systems to distinguish salient regions from a cluttered scene [8]. It has been well studied by cognitive scientists and recently attracted a lot of interest and applications in the computer vision community including object detection [9]. Early saliency detection methods focuses on exploring low-level cues, such as color [10], orientation [11], and texture [12], to detect the difference between the saliency regions and the background. Generally, local- or global-contrast are the most widely utilized assumption in almost all saliency methods. Local contrast [13]–[15]

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenliang Zhang.

assumes that the more different the object is comparing with its neighborhoods, the more salient the object is. While global contrast [16] assumes that objects with more unique features in the whole image are usually more salient. To obtain saliency map with better performance, it is a common strategy to combine local and global contrast together [17]. Later, a psychophysical study [18] showed that humans favor the center of images. Accordingly humans usually frame objects of interest near the image center when taking photographs. Therefore, salient objects are more likely to be placed near the center of images, and the background generally spreads over the whole image. Thus high level visual cues such as center-bias [19], [20], and background prior [21], [22] are introduced with the location heuristic to get a better performance of saliency. Recently, optimization based methods attract more attention in saliency detection and have achieved state-of-the-art performance. Harel *et al.* [11] introduce a Graph-based visual saliency method by MRF (Markov Random Fields) to simulate the human eye movement, and get better results. Liu *et al.* [23] provide an elegant CRF (Conditional Random Fields) framework to combine multiple feature maps. Yang *et al.* [24] introduce smoothness constraint which has been well documented for graph-based object segmentation for saliency map refining. Besides smoothness constraint, Zhu *et al.* [25] produce a principle framework that integrates low level cues and directly aim for saliency optimization. Qin [26] proposed a cellular automata based saliency propagation to exploit the intrinsic relevance between neighboring cells, which is also an optimization method to improve the saliency result.

With the development of saliency detection, some researchers have introduced saliency into infrared pedestrian detection and proposed various solutions. KO [27] calculated the luminance saliency map which estimates the luminance contrast using a center-surrounded scheme. Zhang [28] proposed an associative saliency, generated from both the region and edge contrasts. Li *et al.* [29] applied the gradient information of pedestrians to enhance the uniqueness of intensity, and combined it with multi-scale contrasts to obtain the final saliency. Wang *et al.* [30] proposed a mutual consistency-guided fusion strategy to adaptively combine the luminance contrast saliency map and contour saliency map for infrared images.

However, most of the existing methods mentioned above just introduce the low level human vision features into infrared saliency detection, without analysis the essence imaging mechanism of infrared imaging. Therefore, they are not effective for complicate infrared scenes. Besides, pedestrian information, such as pedestrian always appears specific shape structures in images, are rarely taking into account for the infrared saliency detection. Based on the above considerations, a reconstruction saliency detection model for infrared pedestrian images is proposed in this paper. The main contributions of this paper are as follows:

1. By fully exploring the mechanism of infrared imaging, the appearance based infrared saliency was calculated locally and globally. This effective and robust appearance saliency features can derive an abstraction for infrared pedestrians.

2. By considering the shape information and the infrared character of the pedestrians, the thermal radiation prior and pedestrian shape prior is introduced to obtain the infrared object prior information.

3. By comprehensively considering both the human visual appearance of the infrared object and the essential mechanism of infrared pedestrian, we introduce reconstruction optimization algorithm to combine the saliency and the object prior together. The reconstructed saliency map can drives the visual attention of the target to achieve effective target detection.

The remainder of the paper is organized as follows. In Section II, we describe the problem of infrared saliency detection and our solution. In section III, we elaborate the details of our proposed reconstructed Saliency for infrared pedestrian images. In section IV, performance evaluation of proposed scheme with comparisons to the state-of-the-art algorithms is presented. Finally Section V concludes the paper.

II. PROBLEM DESCRIPTION

Saliency detection has got a great useful application in various research fields, by combining with different human vision mechanism. When it comes to infrared images, it can be also helpful to enhance the performance of infrared segmentation or infrared object detection. However, the traditional saliency detection algorithm is modeled by lower contrast character, it does not work well for the complex scenes with background interference and low contrast. That's because there always exist similar contrast areas between the object and the background, it's hard to distinguish them by the contrast only. What's more, to some specific object with complex structure, color or texture, it's even harder to obtain the complete saliency areas by the traditional vision system.

Thus, one of the feasible solution is to effectively express the infrared imaging mechanism together with the prior knowledge of the object to reconstruct the saliency model. It can combine the visual task into the human brain memory with the object knowledge storage to perform visual selection. Thus, the reconstructed model can improve the accuracy and integrity of target detection in complex background saliency maps.

The salient map of image represent probability of the foreground object may appear. When applied with different vision system, there may be different probability density value for the same pixel. Obviously, the visual information contained in these probability densities is not the same, and the context information around the region is different also. By comprehensively considering the probability density map obtained with different visual mechanisms, it can effectively compensate for the problem of information loss caused by insufficient single visual information. Based on that, the reconstructed saliency can be described below:

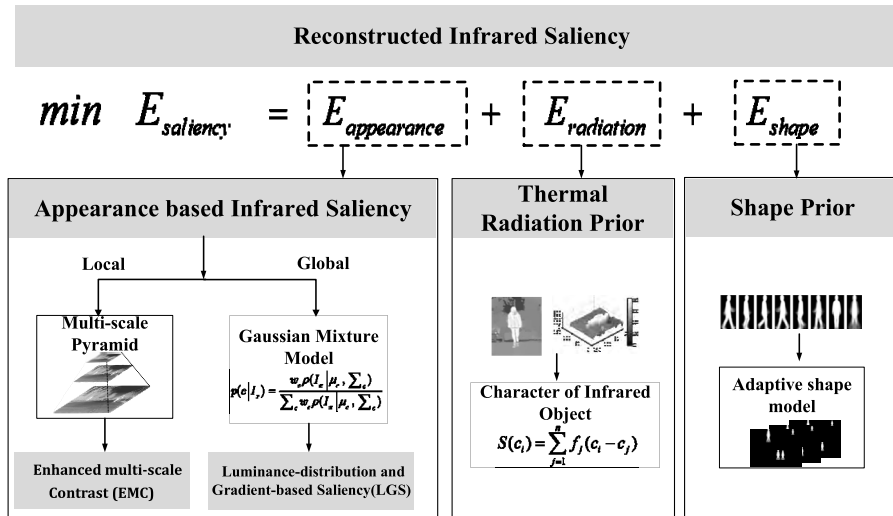$$S = Opt(S_{V\text{is}}, S_{Pri}) \tag{1}$$

FIGURE 1. Framework of the proposed algorithm.

In which, S_{Vis} is the saliency result by visual mechanism, S_{Pri} is the prior knowledge of the object or the imaging mechanism. $S = Opt(\cdot)$ is the reconstruction operator to combine S_{Vis} and S_{Pri}. Here we introduce random walk as an optimization operator to combine visual mechanism and prior knowledge.

Random walk is a mathematical formalization of a random sequence path, which leads an element to a seed location with the highest likelihood [31]. Random walks first appeared in computer vision in the early work of Wechsler and Kidode for texture discrimination [37]. More recently, the average hitting time of a random walk from an object boundary has been studied as a measure to characterize object shape [38]. This approach was recently applied to automatic image segmentation [30] by choosing the designated node randomly and recursively partitioning until a measure of partition quality is violated. The goal of the random walk segmentation algorithm is to obtain the walk probability of the unmarked points in the graph, which can be minimized by the energy function gives in the Eq.(2)

$$E = D[x] = \frac{1}{2}\sum_{e_{ij} \in E} w_{ij}(x_i - x_j)^2 = \frac{1}{2}x^T L x \qquad (2)$$

In which L is Laplace matrix obtained by the conversion of the undirected graph, which can be elaborated in [31].

In computer vision literature, random walk was often introduced as an optimization method to combine different information together. In [32], random walk was used to get segmentation result naturally by adding energy terms and performing minimization on the total energy. In [21], random walk was also used to improve the saliency estimation through combining pixel-wise graph term and a newly formulated fitting constraint, to take local image data and prior estimation into account. Here, we introduce random walk as an optimization method to reconstruct saliency map by comprehensive considering the appearance of image $E_{appearance}$,

the imaging mechanism prior of infrared image $E_{radiation}$, and the shape prior of object itself E_{shape},

$$
\begin{aligned}
E_{saliency} &= E_{Vis} + E_{Pri} \\
&= E_{appearance} + E_{radiation} + E_{shape} \qquad (3)
\end{aligned}
$$

By using the least square optimization method to solve the value under the minimum energy, the optimal saliency map can be obtained as below,

$$S^* = \arg\min_s \{E\} \qquad (4)$$

III. PROPOSED METHOD

Fig. 1 shows the framework of our proposed saliency detection algorithm for infrared pedestrian images. First, appearance based infrared saliency is proposed locally and globally respectively and will be elaborated in Section III.A. Second, infrared object prior can be described from thermal radiation and shape characteristic, as illustrated in Section III.B. and III.C. Finally, by introducing the above appearance based infrared saliency and object prior, we construct random walk based saliency optimization method to obtain the finally saliency map, which will be presented in Section III.D.

A. APPEARANCE BASED INITIAL SALIENCY

From the perspective of human visual attention, infrared images have the following two main characteristics [9], [33]:

1. Local features: such as grayscale contrast. For infrared pedestrian images, the pedestrian target tends to have a higher luminance distribution compared to the background, especially in the local areas.

2. Global characteristics: the proportion of the target in the image is relatively low, while the background appears with higher frequency. In general, the target has a higher brightness, a compact spatial distribution, and a distinguishable contour edge.

Combing the local and global features of visual saliency can enhance the effect of saliency detection result in infrared images especially in complex backgrounds, which has already been verified in our previous work [29]. So here, we obtain the appearance based infrared saliency through combining the local and global saliency together, just like what we did before in [29].

$$I_{appearance} = f_c(x, I) * f_d(x, I) \quad (5)$$

In which, $f_c(x, I)$ corresponds to the local multi-scale Contrast. $f_d(x, I)$ corresponds to the global features. Multi-scale contrast $f_c(x, I)$ is used to extract regions with large differences, in which the target is better positioned to suppress the influence of background noise,

$$f_c(x, I) = \sum_{l=1}^{L} \sum_{x' \in N(x)} \left\| (I_{whc}^{(l)}(x) - I_\mu^l)^2 - (I_{whc}^{(l)}(x') - I_\mu^l)^2 \right\| \quad (6)$$

where I_μ is the arithmetic mean pixel value of the image and I_{whc} is the Gaussian blurred version of the original image to eliminate fine texture details as well as noise. $I^{(l)}(x)$ denotes the pixel x in the lth-level image in the pyramid and the number of pyramid levels L is set to 6. $N(x)$ is a 9×9 window.

The object region tends to have relatively rich gradient information, the gradient-weighted luminance spatial distribution is used to enhance the compact object and extract a complete and significant object region. Here, we use Luminance-distribution and Gradient-based Saliency as $f_d(x, I)$

$$f_d(x, I) = p(c | I_x) = \frac{w_c \rho(I_x | \mu_c, \sum_c)}{\sum_c w_c \rho(I_x | \mu_c, \sum_c)} \\ * gradient(I - I_\mu)^2 \quad (7)$$

Since appearance based saliency have already taken local contrast and global distribution into account, we introduce appearance based saliency as initial information in Eq. (3).

$$p_l(v_i) = I_{appearance}(v_i) \quad (8)$$

$$E_{appearance} = \frac{1}{2} x^T L x \quad (9)$$

In which $p_l(v_i)$ represent the probability of the foreground, which is a one-dimensional column vector. L is the Laplace Matrix converted from the undirected graph.

Compared with intensity profiling, appearance based saliency can describe the density of each pixels belonging to background and foreground more accurately. Fig.2 presents the results of the proposed appearance based initial saliency. As can be observed, in the saliency map obtained by the original multi-scale contrast approach in Fig.2 (b), the pedestrian on the left is hardly visible while strong background still exists, especially in the up-left corner of the image. On the other hand, in the saliency map obtained by the enhanced multi-scale contras approach, as shown in Fig.2 (c), the two problems in original multi-scale contrast approach are all
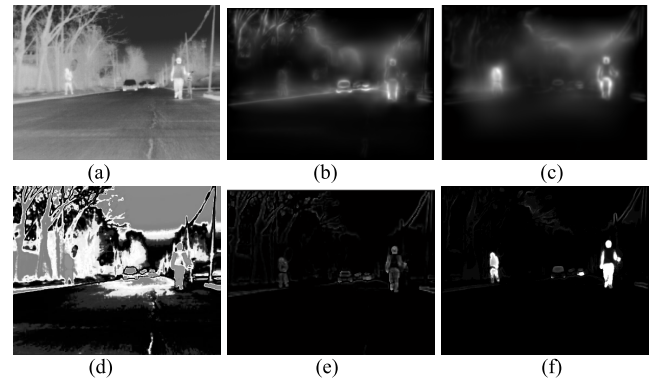

FIGURE 2. (a) Original Infrared Image; (b) saliency map of multi-scale contrast [9]; (c) saliency map of enhanced multi-scale contrast, Eq.(6); (d) luminance-distribution saliency map [33]; (e) luminance-distribution and gradient-based saliency, Eq.(7); (f)appearance based initial saliency.

well-handled. Besides, compared to the saliency map of luminance distribution based GMM in Fig.2 (d), the pedestrian objects in Fig.2 (e) are much improved and the background is well suppressed. So by combining the local enhanced multi-scale contrast with the global luminance-distribution and gradient based saliency map, the appearance based initial saliency yields significantly better results, as can be seen in Fig.2 (f). The enhanced multi-scale contrast make the blurred objects prominent in local regions while the luminance-distribution and gradient based saliency removed most of the homogenous regions outside the salient object.

B. SALIENCY WITH THERMAL RADIATION PRIOR

In general, the higher the temperature of the target, the stronger its heat radiation is, which always show brighter areas in the image. In natural environment, the intensity of infrared heat radiation generated by pedestrians is often higher than the intensity of radiation generated by surrounding scenes. Thus, the brightness of pedestrian area is generally higher than the surrounding background in infrared images.
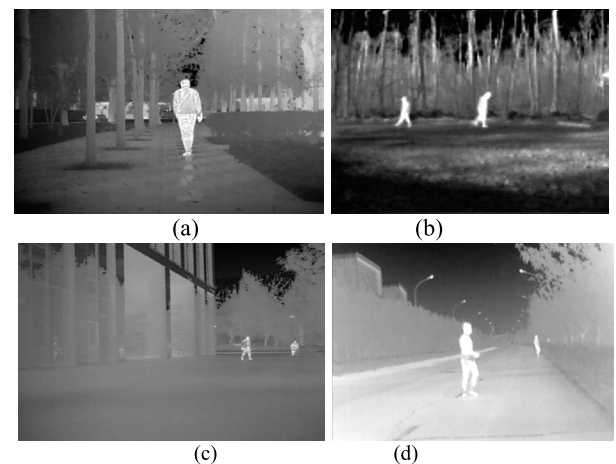

FIGURE 3. Typical infrared pedestrian images.

Fig.3 shows typical infrared pedestrian images in different scenarios. Fig.3 (a) and Fig.3 (b) are infrared images in

natural scenes. There are trees, shrubs, ground, and sky in the background. Pedestrian thermal radiation in the image is higher than the background, so pedestrian in the infrared image is the brightest area. Fig.3 (c) is the infrared image of the building scenes. Similarly, the temperature of the pedestrian is higher than that of the trees on the surface of the building and the background. Although the target is relatively small, the luminance radiation characteristic from the infrared image can still identify the pedestrian's posture characteristics. Fig.3 (d) is an infrared image of a pedestrian crossing road. The thermal radiation of the pedestrian is the largest in the infrared image, although there are similar thermal radiation of street lights and automotive lighting distractions on the background.

Thus, for each image I, we introduce thermal radiation as prior

$$E_{radiation} = \frac{1}{2}(x - p_l)^T(x - p_l) \tag{10}$$

By considering appearance and radiation, the energy of the image can be rewrite as

$$E_{app+rad} = E_{appearance} + E_{radiation}$$
$$= \frac{1}{2}x^T L x + \frac{1}{2}\mu_l(x - p_l)^T(x - p_l) \tag{11}$$

In which, μ_l is the coefficient of the infrared saliency.

As we discussed above, thermal radiation is one of the most typical characteristic in infrared images. For infrared images, there are only luminance information for each pixels, which reflects the thermal radiation of different object and the background. In addition, combining the different heat radiation characteristics of the target and background in infrared images mentioned above, the infrared visual attention can be attracted not only by the high contrast areas, but also by the brighter areas of the infrared image. Considering the four boundaries of the image, most of the pixel values in the boundaries is lower than pixels in the object itself. The dark areas of the boundaries might be mistaken detected as target by directly using the absolute value calculation method. Thus, we modify the saliency calculation method by retaining the sum of the difference between the pixel and other pixels when it is above 0, while setting the saliency value to 0 when the sum of difference is less than zero.

Here, we calculate the thermal radiation by rewriting the formula of global contrast as below:

$$radiation(c_i) = \begin{cases} S(c_i) = \sum\limits_{\forall I_i \in I} D(I_k, I_i), & S(c_i) \geq 0 \\ 0, & S(c_i) < 0 \end{cases} \tag{12}$$

where $D(I_k, I_i)$ is the color Euclidean distances for pixels I_k and pixels I_i in RGB space, $D(I_k, I_i) = ||I_k - I_i||$.

Realign the elements for Laplacian matrix L, x and prior probability p_l.

$$E_{app+rad} = \frac{1}{2}\begin{bmatrix} x_M^T & x_U^T \end{bmatrix}\begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix}\begin{bmatrix} x_M \\ x_U \end{bmatrix}$$
$$+ \frac{\mu_l}{2}\left(\begin{bmatrix} x_M \\ x_U \end{bmatrix} - \begin{bmatrix} p_{l_M} \\ p_{l_U} \end{bmatrix}\right)^T\left(\begin{bmatrix} x_M \\ x_U \end{bmatrix} - \begin{bmatrix} p_{l_M} \\ p_{l_U} \end{bmatrix}\right)$$

$$= \frac{1}{2}\left(x_M^T L_M x_M + 2x_U^T B^T x_M + x_U^T L_U x_U\right)$$
$$+ \frac{\mu_l}{2}\begin{pmatrix} x_M^T x_M - 2p_{l_M}^T x_M + p_{l_M}^T p_{l_M} + x_U^T x_U \\ -2p_{l_U}^T x_U + p_{l_U}^T p_{l_U} \end{pmatrix} \tag{13}$$

Make $\frac{dE_{app+rad}}{dx_U} = 0$, then

$$B^T x_M + L_U x_U + \mu_l\left(x_U - p_{l_U}\right) = 0 \tag{14}$$

Finally, we can get the probability of an unmarked pixel

$$x_U = (L_U + \mu_l I)^{-1}\left(\mu_l p_{l_U} - B^T x_M\right) \tag{15}$$

FIGURE 4. Infrared image and thermal radiation prior saliency map.

Fig.4 shows a typical infrared image and the thermal radiation prior saliency map. It can be seen from the results on the right that the global contrast method can enhance the target areas with higher thermal radiation in the infrared image, while suppress the background areas with lower thermal radiation. In the saliency map, the brightness of the pedestrian target is much higher than the background areas, which provides an effective prior to accurately and completely detecting the pedestrian target in the infrared image.

C. SALIENCY WITH PEDESTRIAN SHAPE PRIOR

Although the random walk algorithm, combining saliency appearance and radiation prior, can obtain the global optimum result, it might lead to false result especially when there are similar luminance distribution among background and object. Considering the special shape structure of pedestrian in infrared pedestrian images, we introduce the shape prior of pedestrian into random walk algorithm. Thus, through reconstructing the probability of infrared pedestrian with prior information, the robustness of algorithm and the accuracy of object extraction can be efficiently improved.

In [6], we use statistical methods to obtain the common characteristics of pedestrian contours in different walking directions, to estimate the pedestrian shape prior model. For difference pedestrian's walking direction, the spatial position and shape of the arms and legs will be obvious different, while
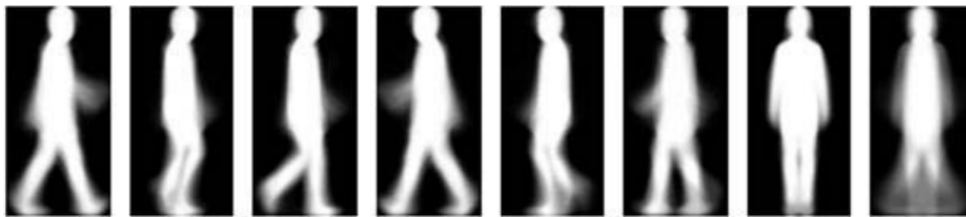
FIGURE 5. Statistical model for different pedestrians.

the posture of the trunk part is basically unchanged. The shape and area of the pedestrian's head are basically not affected by the direction of action.

In order to express the shape similarity characteristics of pedestrians more clearly, all of the templates are clustered into 7 different classes, which means that the pedestrian binary template images in the same category are superimposed to form a prior shape probability of the pedestrian shape template. As shown in Fig.5, the 1st to 7th images from left to right are the results of superimposed pedestrian contours in the same direction, and the eighth image $T(x)$ is the result of statistical superposition of the first to seven images, as shown in the following equation:

$$T(x) = c \sum_{s=0}^{7} G(x - \mu^s, D^s) \qquad (16)$$

Here, $G(x - \mu^s, D^s)$ is the 1st to 7th image,Fig.5, c is the weighing coefficient. Thus we can use $T(x)$ as statistic prior model for the head, the trunk, arms and legs of different pedestrians with different walking direction. Since it is similar spatial character for the pedestrian appeared under natural images and infrared images, so we can use $T(x)$ to describe the statistic character of infrared pedestrian. In this paper, we use the spatial distribution difference of pixels to define the shape probability for the different areas of infrared images, as shown in equation (17).

$$F = \sum_{i,j} I(i,j)d(i,j)^2 \qquad (17)$$

In which, $I(x, y)$ is the pixel value for pixel (x, y), $d(x, y)^2$ is the geometrical distance between (x, y) to the center of the image. Thus, the similarity of the image with the prior shape template can be calculated as

$$Sim = \frac{F}{F_T} = \frac{\sum_{i,j} I(i,j)d(i,j)^2}{\sum_{i,j} T(i,j)d(i,j)^2} \qquad (18)$$

The closer the value of Sim is to 1, the more similar the region is to the pedestrian prior template.

For any input infrared images, there might be multi pedestrians objects. Considering we have already got the saliency map $Sal(c_i)$ before, it can also be combined with the shape prior to get a more adaptive template for any input images. First, the salient map is binarized to extract the connected domain. The regions which does not conform to the pedestrian aspect ratio is eliminated. The obtained binary mask map

is a potential pedestrian target region with potential position and size information. Then, the adaptive pedestrian shape prior model can be obtained by the combination with the shape template

$$Shape_{ij} = \begin{cases} Ibw_{ij} \cdot \exp(-(Sim - 1)^2), & Ibw_{ij} \neq 0 \\ 0, & Ibw_{ij} = 0 \end{cases} \qquad (19)$$

In which Ibw_{ij} is the binary image after thresholding the saliency map.

Thus, an adaptive shape prior template *Shape* can be produced, according to the characteristics of different images. Then, refer to the generation method of matrix L, the template *Shape* can also be mapped as graph $G_S = (V, E)$, the weight of edge connection is defined as

$$w_S(e_{ij}) = \exp(-\beta_S(g_{S_i} - g_{S_j})^2) \qquad (20)$$

Define the Laplace matrix S as,

$$S_{ij} = \begin{cases} d_{S_i}, & i = j \\ -w_S(e_{ij}), & v_i \text{ and } v_j \text{ is adjacent} \\ 0, & \text{other} \end{cases} \qquad (21)$$

The shape prior model E_{shape} can be defined as

$$E_{shape} = \frac{1}{2}(x - p_s)^T S (x - p_s) \qquad (22)$$

$$E_{app+sha} = E_{appearance} + \mu E_{shape}$$
$$= \frac{1}{2}x^T Lx + \frac{1}{2}\mu_s(x - p_s)^T S(x - p_s) \qquad (23)$$

where

$$p_s = \begin{bmatrix} p_{s_M} \\ p_{s_U} \end{bmatrix}, \quad S = \begin{bmatrix} S_M & R \\ R^T & S_U \end{bmatrix} \qquad (24)$$

Thus

$$
\begin{aligned}
E_{app+sha} &= \frac{1}{2}\begin{bmatrix} x_M^T & x_U^T \end{bmatrix}\begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix}\begin{bmatrix} x_M \\ x_U \end{bmatrix} \\
&+ \frac{\mu_s}{2}\left(\begin{bmatrix} x_M \\ x_U \end{bmatrix} - \begin{bmatrix} p_{s_M} \\ p_{s_U} \end{bmatrix}\right)^T \begin{bmatrix} S_M & R \\ R^T & S_U \end{bmatrix} \\
&\times \left(\begin{bmatrix} x_M \\ x_U \end{bmatrix} - \begin{bmatrix} p_{s_M} \\ p_{s_U} \end{bmatrix}\right) \\
&= \frac{1}{2}\left(x_M^T L_M x_M + 2x_U^T B^T x_M + x_U^T L_U x_U\right) \\
&+ \frac{\mu_s}{2}(x_M^T S_M x_M - 2x_M^T S_M p_{s_M} + p_{s_M}^T S_M p_{s_M} \\
&+ 2x_M^T R x_U - 2x_M^T R p_{s_U} - 2x_U^T R^T p_{s_U} + p_{s_M}^T R p_{s_M} \\
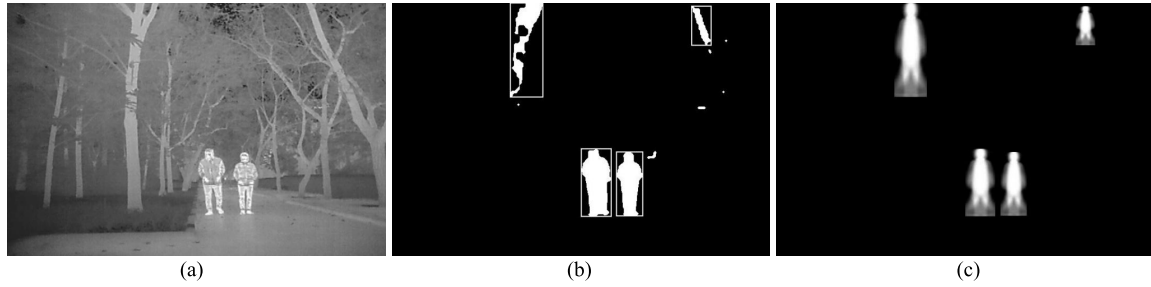&+ x_U^T S_U x_U - 2x_U^T S_U p_{s_U} + p_{s_U}^T S_U p_{s_U}) \qquad (25)
\end{aligned}
$$

FIGURE 6. Pedestrain shape prior. (a) Infrared image (b) region of interest (ROI) (c) adaptive *Shape* model.

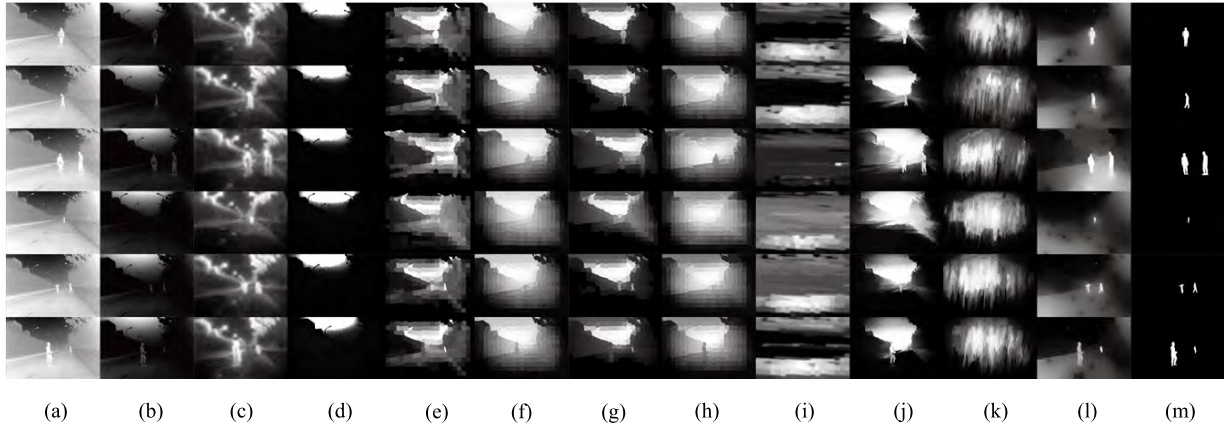(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m)

FIGURE 7. Visual comparison on the IMS datasets between the proposed method and state-of-the-art saliency models.(a) OR. (b) FT. (c) CA. (d) SF. (e) GS.(f) MR. (g) BD. (h) BSCA. (i) MAP. (j) MB+. (k) DSR. (l) our. (m) GT.

Make $\frac{dE_{app+sha}}{dx_U} = 0$, the combined energy can be minimized by solving the following system:

$$B^T x_M + L_U x_U + \mu_s \left(R^T \left(x_M - p_{s_M} \right) + S_U \left(x_U - p_{s_U} \right) \right) = 0 \tag{26}$$

Here we can get the probability of unmarked point,

$$x_U = (L_U + \mu_s S_U)^{-1} \left(-(B^T + \mu_s R^T)x_M \right.$$
$$\left. + \mu_s (R^T p_{s_M} + S_U p_{s_U}) \right) \tag{27}$$

As shown in the above figure, it is an adaptive template automatically generated by using the shape prior template and the ROI area of the image. This template can guide the shape prior of the target in the image when constructing the shape prior optimization function.

D. RECONSTRUCTED SALIENCY MODEL

To obtain foreground region, our foreground mask is decided by the reconstructed energy function. We propose a new energy function combined with constraints from appearance initial saliency, radiation prior, and shape prior model. The total energy function is defined as follows:

$$E_{saliency} = E_{appearance} + E_{radiation} + E_{shape}$$
$$= \frac{1}{2} x^T L x + \frac{\mu_l}{2} (x - p_l)^T (x - p_l)$$
$$+ \frac{\mu_s}{2} (x - p_s)^T S (x - p_s) \tag{28}$$

Realign the elements of seed point and others, Eq. (28) can be rewritten as

$$E_{saliency} = D[x_U] = \frac{1}{2} \begin{bmatrix} x_M^T & x_U^T \end{bmatrix} \begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix} \begin{bmatrix} x_M \\ x_U \end{bmatrix}$$
$$+ \frac{\mu_l}{2} \left(\begin{bmatrix} x_M \\ x_U \end{bmatrix} - \begin{bmatrix} p_{l_M} \\ p_{l_U} \end{bmatrix} \right)^T \left(\begin{bmatrix} x_M \\ x_U \end{bmatrix} - \begin{bmatrix} p_{l_M} \\ p_{l_U} \end{bmatrix} \right)$$
$$+ \frac{\mu_s}{2} \left(\begin{bmatrix} x_M \\ x_U \end{bmatrix} - \begin{bmatrix} p_{s_M} \\ p_{s_U} \end{bmatrix} \right)^T \begin{bmatrix} S_M & R \\ R^T & S_U \end{bmatrix}$$
$$\left(\begin{bmatrix} x_M \\ x_U \end{bmatrix} - \begin{bmatrix} p_{s_M} \\ p_{s_U} \end{bmatrix} \right) \tag{29}$$

Make $\frac{dE_{saliency}}{dx_U} = 0$, then the solution is given by

$$(L_U + \mu_l I + \mu_s S_U) x_U$$
$$- \left(B^T + \mu_s R^T \right) x_m + \mu_l p_{l_U} + \mu_s \left(R^T p_{s_U} + S_U p_{s_U} \right) \tag{30}$$

IV. PERFORMANCE EVALUATION AND COMPARISONS

This section first introduce the datasets and evaluation metrics in our experiments. Then, the performance of the proposed saliency detection scheme with comparisons to the state-of-the-art saliency detection approaches is evaluated. Finally, to further verify the effectiveness of the proposed method, we also compare the saliency performance with well-known infrared saliency methods.

A. DATASETS AND EVALUATION METRICS

To evaluate the effectiveness of the proposed method, the experiments are tested on two datasets.
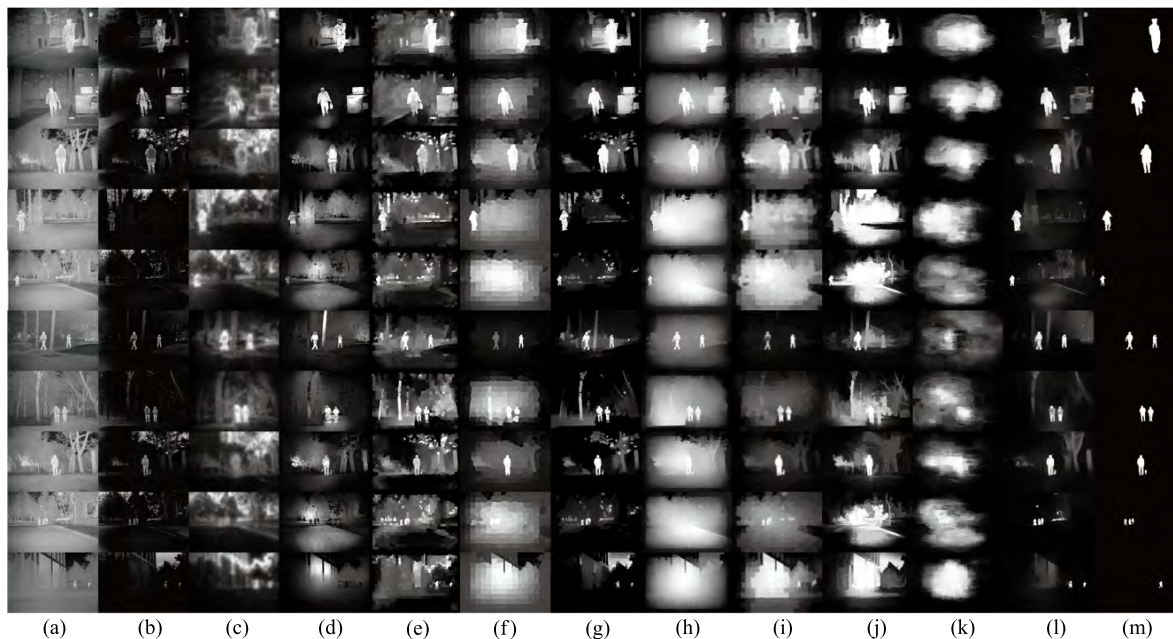
FIGURE 8. Visual comparison on the DIP datasets between the proposed method and state-of-the-art saliency models.(a) OR. (b) FT. (c) CA. (d) SF. (e) GS. (f) MR. (g) BD. (h) BSCA. (i) MAP. (j) MB+. (k) DSR. (l) our.(m) GT.
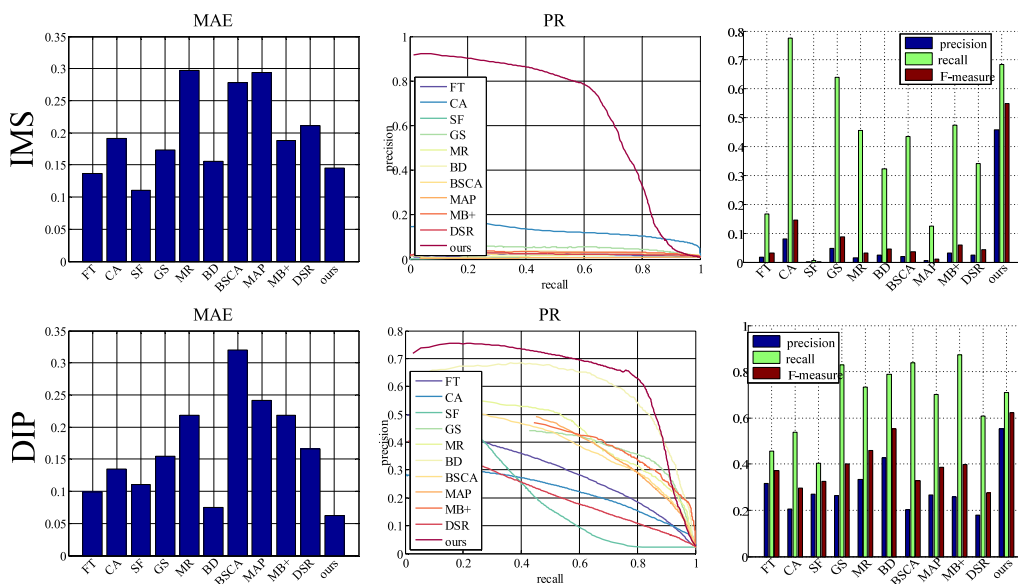

FIGURE 9. Objective comparison on IMS and DIP dataset, from left to right: MAE, PR, PRF.

1) IMS: This dataset is provided by our cooperation partner, which consists of 190 images with one or two pedestrians in each image. There are 39 images containing one pedestrian and other images contain two pedestrians. In this dataset, pedestrians either walk away or walk towards the camera. Thus the size of pedestrians changes a lot.

2) DIP: As the dataset IMS is relatively simple and do not comprehensively cover all the situations, we built a more complex dataset to comprehensively testify the effectiveness of the proposed method. 400 infrared images are contained in this dataset, which were scanned by a Tau 2 LWIR camera.

There are totally 634 pedestrians in the dataset, which contains 220 images with single pedestrian and 180 images with multiple pedestrians. This dataset includes 31 different scenes in all, which results in more complicated backgrounds. Also, different postures and various sizes are all taken into account. Because of its comprehensiveness, this dataset is important in this paper to verify the effectiveness of the proposed method. Moreover, the human-segmented ground truths (GT) are also provided.

Both subjective and objective performance evaluation and comparisons are conducted to verify the effectiveness
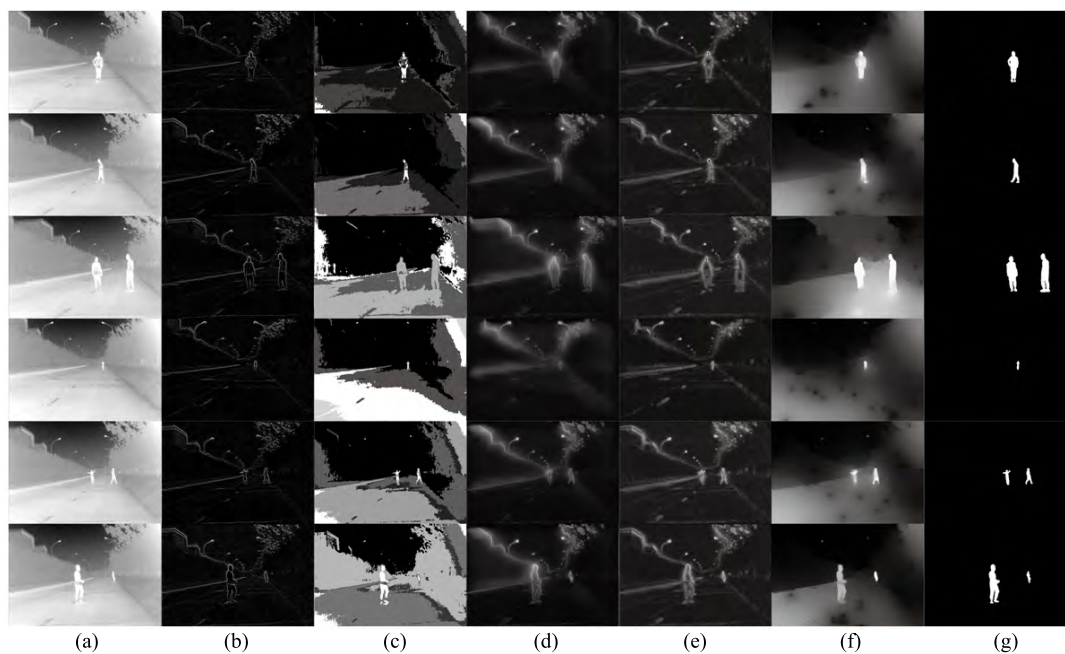
FIGURE 10. Visual comparison on the IMS with saliency models of infrared pedestrian images. (a) OR. (b) LSM. (c) AS. (d) CD.(e) MCS. (f) our. (g) GT.

of the proposed scheme, in which objective performance of saliency detection are evaluated in terms of the PR (Precision-Recall) curves [21], [25], MAE (Mean Absolute Error) [12], [18], [24] and F-measure [12], [14], [24]. It should be noted that a satisfactory saliency algorithm should yields a low value of MAE and high values of Precision, Recall and F-measure.

B. PERFORMANCE COMPARISON WITH THE-STATE-OF-ART SALIENCY DETECTION METHOD

In this section, the comparisons of the proposed saliency detection method against ten state-of-the-art saliency methods are presented, including FT [13], CA [16], SF [9], GS [34], MR [35], BD [25], BSCA [26], MAP [36], MB+ [37], DSR [38].

Fig.7 shows the subject result on IMS datasets. As can be seen, besides CA [16] can highlight the outline area of pedestrians to some extent, the other algorithms can hardly distinguish pedestrians from the low-contrast infrared images. From the second-to-last column of the result in Fig.11, we can see that the proposed algorithm can accurately locate the infrared pedestrian's area and can separate the object and the background completely. The proposed algorithm is effective for images with single and small pedestrian, big pedestrians with uneven gray distribution, and multiple pedestrian images with different scales.

Fig.8 shows the subjective visual results on DIP datasets. Compared with IMS, this datasets is more challenging with complex background and pedestrians with different posture and inhomogeneous distribution. It can be seen that, BSCA [26], MAP [36], MB+ [37] and DSR [38], which

use the probabilistic propagation method for significant target detection, cannot distinguish the object form the background in the infrared image. In the saliency map obtained by MR [35], the saliency value of the four borders in the image is lower, while the center region of the image is higher. This method has some significant effects on large-scale pedestrians, but it cannot distinguish pedestrians with smaller scale or weaker contrast with the background. The target and the background in the saliency map obtained by CA [16] are relatively vague. In the saliency map of SF [9], GS [34] and BD [25], the objects are relatively complete, and the significant value is high and uniform. But here are still background noise which cannot be removed. The result of FT [13] has a certain degree of background suppression, but at the same time it also suppresses the saliency value of the object. However, the proposed method can not only enhance saliency of the pedestrian uniformly, but also suppress the background to the greatest extent. What's more, we can also obtain the most satisfactory results for images of different scales in different scenes.

Besides subjective comparisons, we further compare the objective performance with different saliency methods on MAE, PR, and F-measure, as shown in Fig.9. For dataset IMS, the proposed method achieves the second lowest MAE value, and the highest precision in almost all the recall range [0, 1] up to 0.8, while the precisions of all the other methods are lower than 0.2. This demonstrates the effectiveness of the proposed method for suppressing background. For dataset DIP, the proposed method is superior to the other state-of-the-art methods in all the evaluation metrics. It get the lowest MAE value around 0.05, the highest PR curve
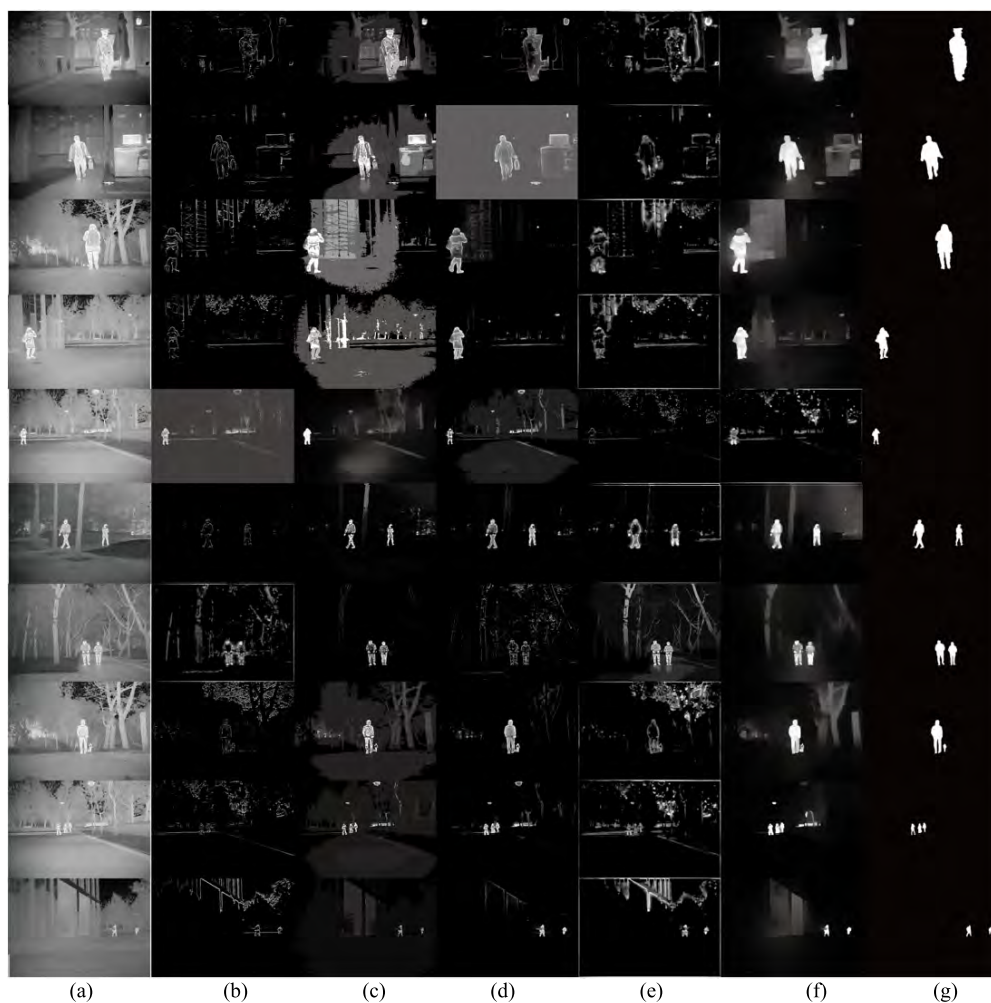
FIGURE 11. Visual comparison on the DIP with saliency models of infrared pedestrian images.(a) OR.(b) LSM. (c) AS. (d) CD. (e) MCS. (f) our. (g) GT.

above 0.7 and the highest PRF value This is in keeping with the visual performance that the dark sky region tends to be taken as salient region, while the pedestrian regions are totally suppressed with these compared methods. All the experiments indicate the superiority of the proposed method and demonstrate the good performance for infrared pedestrian saliency detection regardless of the complexity of images.

C. PERFORMANCE COMPARISON WITH INFRARED SALIENCY DETECTION METHOD

Actually, the above state-of-the-art saliency models are designed for visible images. To demonstrate the superiority of the proposed saliency model more persuasively, we also compare our saliency performance with state-of-the-art infrared saliency models, includes LSM [27], AS [28], CD [29] and MCS [30]. The experiments are also operated on the three datasets IMS, and DIP.

Fig.10 shows the subjective results on dataset IMS. It's difficult because of the inhomogenous luminance distribution and the low contrast of the images. All the four infrared saliency methods fail to detect the most saliency pedestrians

from the images, while the proposed method can obtain the pedestrians acutely and completely. Besides, the proposed method is suitable for both single object and multi-objects, even when there are both small and big objects in one image.

Fig.11 show the subjective results on dataset DIP. Since the complexity of the background and diversity of the objects in this dataset, it's hard to find a perfect saliency model suitable for all the different scenes. As can be seen in Fig.11, LSM [27] use central contrast and CD [29] use gradient to calculate the saliency map which can only detect the significant varied contour of the objects. AS [28] and MCS [30] combine region contrast and edge contrast, but the problem of inhomogeneous luminance distribution of the object itself is not taken into consideration. So the saliency object in the result is comparable vague, and pedestrians can not be distinguished from the background. While the proposed method can obtain the saliency map with homogeneous luminance distribution and distinct edges of the salient objects.

We also compare the objective performance with different saliency methods on MAE, PR, and F-measure, as can be seen in Fig.12. For dataset IMS, the proposed method outperform
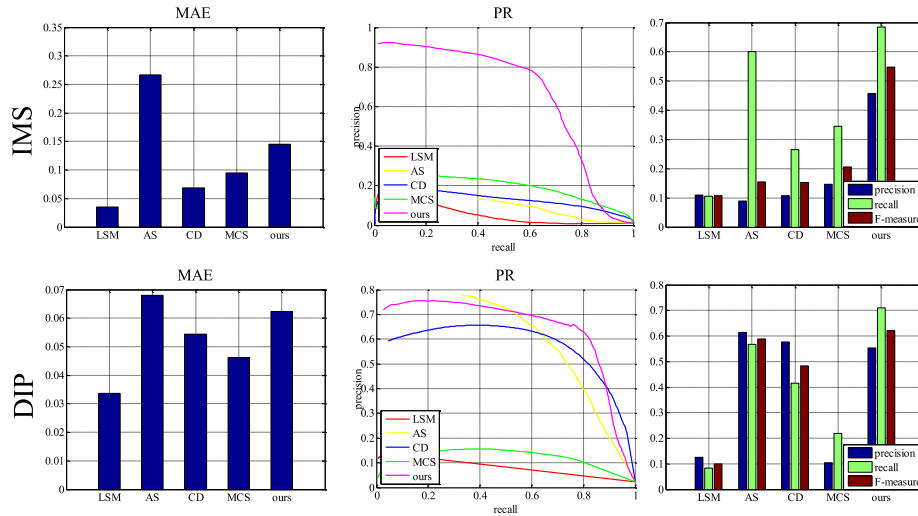
FIGURE 12. Objective comparison on IMS and DIP dataset, from left to right: MAE, PR, PRF.

all the other comparable methods: we get the highest PR up to 0.9 while the other methods all lower than 0.4, and we get the highest F-measure up to 0.55 while the other all lower than 0.2. Also for dataset DIP, the proposed method get the highest PR up to 0.75, the lower MAE 0.06 and the highest F-measure 0.6. All the above results demonstrate that the superiority performance of the proposed method.

V. CONCLUSIONS

In this paper, we have studied the challenging problem of saliency detection in infrared pedestrian images and developed an effective reconstruction optimization based infrared saliency method. First, by fully exploring the mechanism of infrared imaging, the appearance based infrared saliency was calculated to derive an abstraction for infrared pedestrians. Second, by considering the shape information and the infrared character of the pedestrians, the thermal radiation prior and pedestrian shape prior is introduced to obtain the infrared object prior information. Finally, the infrared pedestrian saliency map can be calculated through random walk based saliency reconstruction optimization method with the appearance saliency and infrared object prior. Extensive experimental results on real infrared images captured by intelligent transportation system have shown the superior performance of the proposed approach against the state-of-the-art algorithms.

REFERENCES

[1] R. Medina-Carnicer, A. Carmona-Poyato, R. Muñoz-Salinas, and F. J. Madrid-Cuevas, "Determining hysteresis thresholds for edge detection by combining the advantages and disadvantages of thresholding methods," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 165–173, Jan. 2010.

[2] C. G. Harris and M. A. Stephens, "combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, vol. 15, no. 50, 1988, pp. 147–152.

[3] S. P. Chatzis and T. A. Varvarigou, "A fuzzy clustering approach toward hidden Markov random field models for enhanced spatially constrained image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 5, pp. 1351–1361, Oct. 2008.

[4] S. Shirakawa and T. Nagao, "Evolutionary image segmentation based on multiobjective clustering," in *Proc. IEEE Congr. Evol. Comput.*, May 2009, pp. 2466–2473.

[5] M. P. Kumar, P. H. S. Torr, and A. Zisserman, "OBJCUT: Efficient segmentation using top-down and bottom-up cues," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 530–545, Mar. 2010.

[6] Y. Ma, H. Derksen, W. Hong, and J. Wright, "Segmentation of multivariate mixed data via lossy data coding and compression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1546–1562, Sep. 2007.

[7] J. Han, K. N. Ngan, M. Li, and H.-J. Zhang, "Unsupervised extraction of visual attention objects in color images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 141–145, Jan. 2006.

[8] Ali Borji, M.-M. Cheng, Q. Hou, H. Jiang, and J. Li. (2014). "Salient object detection: A survey." [Online]. Available: https://arxiv.org/abs/1411.5878

[9] 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.

[10] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," in *Proc. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 409–416.

[11] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Conf. Adv. in Neural Inf. Process. Syst.* Cambridge, MA, USA: MIT Press, 2007, pp. 545–552.

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

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

[14] 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.

[15] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. 11th ACM Int. Conf. Multimedia*, 2003, pp. 374–381.

[16] 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.

[17] L. Zhou, Z. Yang, Q. Yuan, Z. Zhou, and D. Hu, "Salient region detection via integrating diffusion-based compactness and local contrast," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3308–3320, Nov. 2015.

[18] B. W. Tatler, "The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions," *J. Vis.*, vol. 7, no. 14, pp. 1–17, 2007.

[19] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 2106–2113.

[20] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2976–2983.

[21] 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. (CVPR)*, Jun. 2015, pp. 2710–2717.

[22] H. Jing, X. He, Q. Han, and X. Niu, "Background contrast based salient region detection," *Neurocomputing*, vol. 124, pp. 57–62, Jan. 2014.

[23] T. Liu, J. Sun, N.-N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.

[24] C. Yang, L. Zhang, and H. Lu, "Graph-regularized saliency detection with convex-hull-based center prior," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 637–640, Jul. 2013.

[25] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. Comput. Vis. Pattern Recognit.*, 2014, pp. 2814–2821.

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

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

[28] 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. 2014, pp. 4264–4268.

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

[30] 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.

[31] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.

[32] J.-M. Morel and S. Solimini, *Variational Methods in Image Segmentation*. Basel, Switzerland: Birkhäuser, 1995, pp. 83–91.

[33] 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.

[34] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 29–42.

[35] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 3166–3173.

[36] 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.

[37] 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.

[38] H. Lu, X. Li, L. Zhang, X. Ruan, and M. H. Yang, "Dense and sparse reconstruction error based saliency descriptor," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1592–1603, Apr. 2016.

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.

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 the Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 1986, 1989, and 2006, respectively. His research interests include target detection and recognition, multi-modality image processing, and biomedical image processing and recognition.

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, and image segmentation and analysis.

XIANGZHI BAI received the B.S. and Ph.D. degrees from Beijing University (BUAA), in 2003 and 2009, respectively, where he is currently a Full Professor with the Image Processing Center. He holds 12 national invention patents and has published more than 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.