Research Article

Detecting Saliency in Infrared Images via Multiscale Local Sparse Representation and Local Contrast Measure

Xin Wang,[1,2] Chunyan Zhang,[1] Chen Ning,[3] Yuzhen Zhang,[2] and Guofang Lv[1]

[1]*College of Computer and Information, Hohai University, Nanjing, Jiangsu 211100, China*
[2]*Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China*
[3]*School of Physics and Technology, Nanjing Normal University, Nanjing, Jiangsu 210023, China*

Correspondence should be addressed to Xin Wang; wang_xin@hhu.edu.cn

For infrared images, it is a formidable challenge to highlight salient regions completely and suppress the background noise effectively at the same time. To handle this problem, a novel saliency detection method based on multiscale local sparse representation and local contrast measure is proposed in this paper. The saliency detection problem is implemented in three stages. First, a multiscale local sparse representation based approach is designed for detecting saliency in infrared images. Using it, multiple saliency maps with various scales are obtained for an infrared image. These maps are then fused to generate a combined saliency map, which can highlight the salient region fully. Second, we adopt a local contrast measure based technique to process the infrared image. It divides the image into a number of image blocks. Then these blocks are utilized to calculate the local contrast to generate a local contrast measure based saliency map. In this map, the background noise can be suppressed effectually. Last, to make full use of the advantages of the above two saliency maps, we propose combining them together using an adaptive fusion scheme. Experimental results show that our method achieves better performance than several state-of-the-art algorithms for saliency detection in infrared images.

1. Introduction

In the field of image processing and computer vision, a number of methods for saliency detection in visible images have been proposed [1–5]. These techniques have been proved to be very mature. However, when applied to infrared (IR) images, they tend to fail. The reason is that infrared images, which are generated by infrared imaging sensors, always have low resolutions, low signal-to-noise ratios (SNR), low contrasts, and few useful image features. If the commonly used saliency detection methods that have been originally designed for visible images are directly applied to infrared images without considering the special characteristics of infrared images, their effect cannot be fully played, which directly leads to reduced saliency detection performance [6].

To this end, some techniques have been later proposed especially for saliency detection in infrared images [7–10]. The idea of many of these methods is to predict the size and shape of the background and salient region as accurately as possible and then adopt the method of background suppression to eliminate background and simultaneously preserve the salient region. To some extent, they have achieved good detection effect, but the premise is that the salient region should have a certain signal-to-noise ratio; that is to say, the gray intensity of the salient region must be much higher than the surrounding background region in infrared images. Thus, when the salient region has a low signal-to-noise ratio, most of the above methods cannot effectively restrain background noise as well as highlight the salient region.

In recent years, the sparse representation (SR) theory has been developed in the field of signal processing and analysis [11]. Image sparse representation is the linear representation of image signal by a few atoms in the learned overcomplete dictionary. The key of sparse representation is to construct a good dictionary to make the image signal sparser. Sparse representation has been used in many fields, such as image

denoising [12], image restoration [13], and image target detection [14]. Nowadays, researchers begin to apply sparse representation to image saliency detection. For example, in [15], Xia et al. presented a saliency model based on a nonlocal reconstruction. In the method, the saliency was measured by the sparse reconstruction residual of representing the central patch with a linear combination of its surrounding patches sampled in a nonlocal manner. In [16], Fareed et al. proposed detecting salient region through sparse reconstruction and graph-based ranking. In the first step, the original image was segmented into superpixels. In the second step, the sparse representation measure and uniqueness of the features were computed. Then both were ranked on the basis of the background and foreground seeds, respectively. Thirdly, a location prior map was used to enhance the foci of attention. Rigas et al. [17] presented an algorithm for efficient modeling of visual saliency based on local sparse representation and the use of Hamming distance. This method was based on an efficient comparison scheme for the local sparse representations deriving from nonoverlapping image patches. The sparse coding stage was implemented via an overcomplete dictionary trained with a soft-competitive bioinspired algorithm and the use of natural images. The resulting local sparse codes were pairwise compared using the Hamming distance as a gauge of their coactivation. The calculated distances were used to quantify the saliency strength for each individual patch, and then, the saliency values were nonlinearly filtered to form the final map. Fan and Qi [18] introduced a saliency detection method based on global and local short-term sparse representation. They employed the ICA algorithm to learn a set of basis functions at first and then represented the input image by this set of basis functions. Next, a global and local saliency framework was employed to measure the saliency, where the global saliency was obtained through Low-Rank Representation (LRR), and the local saliency was obtained through a sparse coding scheme. Based on the above introduction, we can see that sparse representation is indeed an effective tool for image saliency detection, so we tried applying the existing sparse representation based saliency detection methods to infrared images. Unfortunately, they cannot obtain satisfying results due to the special characteristics of infrared images. Because of the special imaging mechanism, infrared images usually have low contrast and lots of background noise. The saliency maps obtained by the existing sparse representation based methods always contain much noise and hardly present complete saliency regions.

Hence, to utilize the advantages of sparse representation for saliency detection in infrared images and at the same time overcome the above-mentioned deficiencies, we present a novel saliency detection algorithm based on multiscale local sparse representation and local contrast measure in this paper. First, a multiscale local sparse representation (MLSR) based approach is proposed to detect saliency in infrared images. Based on this approach, multiple saliency maps with various scales are obtained for an input infrared image. Then, to enhance the salient region completely, the maps are combined together to derive a fused saliency map. Subsequently, in order to suppress the disturbances of background and further improve the saliency detection

performance, a local contrast measure (LCM) based technique is adopted to process the input infrared image. Based on this technique, the original infrared image is divided into many image blocks. Then the blocks are used to calculate the local contrast so as to compute a local contrast measure based saliency map. Since this technique takes the characteristics of infrared images into consideration, it can effectively suppress the background noise. Finally, by combining the advantages of multiscale local sparse representation which can fully highlight salient region and local contrast measure which can restrain background noise, the proposed algorithm can guarantee that the detected salient regions of infrared images are accurate and integral, and the background clutters are restrained. Experimental results show that it can provide more accurate and reliable results of saliency detection in infrared images than *tested* conventional algorithms.

The rest of this paper is organized as follows. Section 2 describes the proposed multiscale local sparse representation and local contrast measure based saliency detection method in detail. Section 3 presents and discusses the evaluation results on real-life infrared images. Section 4 concludes the paper.

2. The Proposed Algorithm

The proposed saliency detection method for infrared images involves three main stages, as shown in Figure 1, which correspond to the multiscale local sparse representation based saliency computation, local contrast measure based saliency computation, and adaptive saliency map combination, respectively. We describe these three main stages of the method in detail as follows.

2.1. Multiscale Local Sparse Representation Based Saliency Computation. It has been verified that sparse representation is a useful tool for saliency detection, but it is difficult to obtain complete salient regions when applied to infrared images [15]. To solve this problem, here we propose a multiscale local sparse representation based approach to compute saliency for infrared images. This approach combines the advantages of local sparse representation that can highlight the edge profile of salient object and the advantages of multiscale idea that can enhance the internal area of salient object, which guarantees the integrality of the salient region.

2.1.1. Scale Selection. Multiscale analysis is very useful for estimating visual saliency, since it can help capturing salient objects in the visual field no matter what sizes the objects are. Hence we apply multiscale idea to infrared image salient detection, that is, extracting visual saliency over different scales. However, finding an appropriate number of image scales is not easy for infrared images. While we wish to preserve the fine image structures by selecting a large number of image scales, we also want to avoid decomposing connected regions into smaller noisy image patches. After deep consideration, in this work, a scheme is implemented based upon our previous work [19] that aims to decompose the
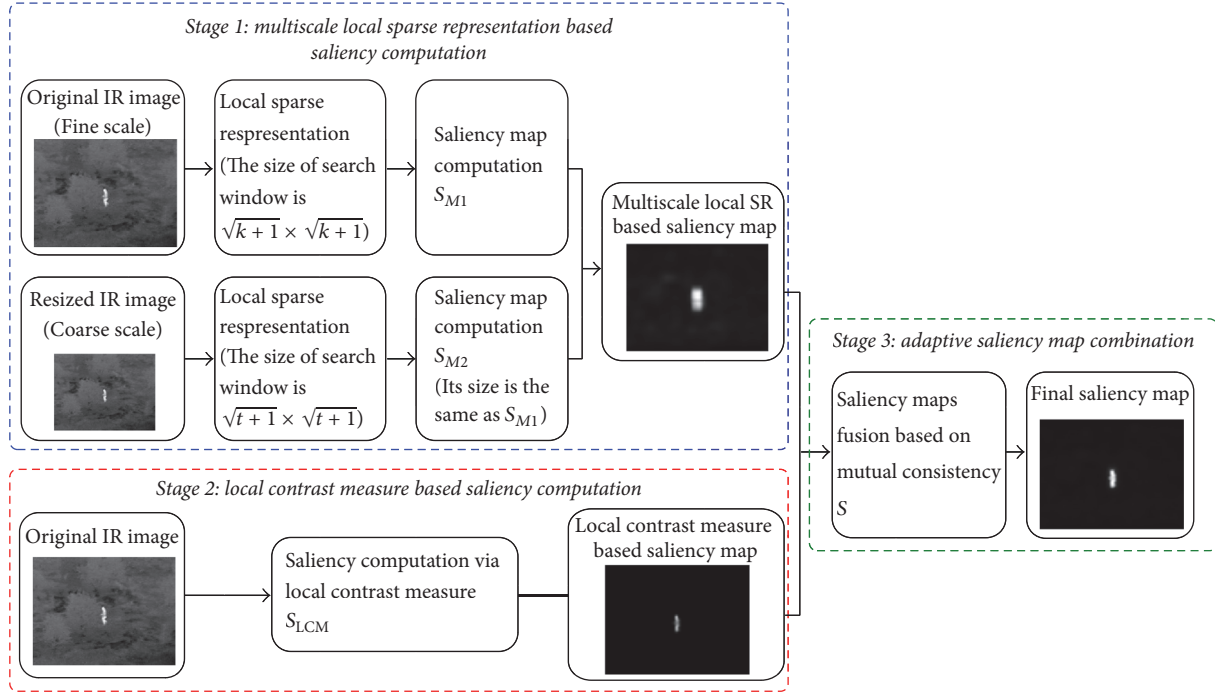
FIGURE 1: The framework of the proposed saliency detection method.

infrared image with different numbers of image scales. The scales are given as follows:

$$l_d = l_{\min} + (d - 1) \cdot \frac{l_{\max} - l_{\min}}{M - 1}, \tag{1}$$

where $l_{\min}$ is the smallest scale which is used to represent the coarse structure of infrared images. $l_{\max}$ is the largest scale which is used to represent the fine structure of infrared images. M is the total number of scales. $d = 1, \ldots, M$ corresponds to the scale d.

For simplicity and efficiency, here we consider two different scales; that is, $M = 2$. For fine scale, $l_{\max}$ is 1. For coarse scale, experiments showed that $l_{\min} = 0.7$ is sufficient to represent the coarse structure of infrared images. Finally, two saliency maps at both the large and the small scales will be calculated, respectively, based on local sparse representation.

2.1.2. Multiscale Local Sparse Representation Based Saliency Map. Given an original infrared image, we get two images with different scales based on the aforementioned multiscale idea and compute the saliency map using local sparse representation [15]. The detailed steps are given as follows.

Step 1. Given an original infrared image F, calculate its saliency map S_{M1} at scale $l_{\max} = 1$. Specifically, in image F, given a pixel $x \in R^2$, we use $\varphi(x) \in R^n$ to represent the vector of the square local patch with n pixels at x, which is formed by sequentially stacking each column of the patch. Let $R(x)$ be the $\sqrt{k+1} \times \sqrt{k+1}$ search window centered at x. The central patch $\varphi(x)$ can be represented by using a linear

combination of patches whose central pixels are in the search window $R(x) - \{x\}$:

$$\varphi(x) = \sum_{i=1}^{k} \alpha_i \varphi(y_i) = \Phi A, \tag{2}$$

where $\Phi = [\varphi(y_1), \ldots, \varphi(y_k)] \in R^{n \times k}$ is the data matrix and $A = [\alpha_1, \ldots, \alpha_k]^T$ is the coefficient vector of the linear combination. In order to make full use of the sparse representation for the original infrared image, we should ensure that the coefficient vector contains only few nonzero elements. So we transform (2) to the following form:

$$\widehat{A} = \arg\min_{A} \quad \|A\|_{l_0}, \\ \text{s.t.} \quad \Phi A = \varphi(x), \tag{3}$$

where $\|\cdot\|_{l_0}$ denotes the pseudo l_0-norm. Orthogonal Matching Pursuit (OMP) can be used to compute the coefficient $\widehat{A}$, and then reconstructed patch is calculated by

$$\widehat{\varphi}(x) = \Phi\widehat{A}. \tag{4}$$

Subsequently, the reconstruction residual is calculated by

$$r(x) = \|\varphi(x) - \widehat{\varphi}(x)\|_2, \tag{5}$$

where $\|\cdot\|_2$ denotes the l_2-norm.

By applying the above local sparse representation to the whole infrared image, that is, changing the central patch by scanning the original image from left to right and top to bottom, we can finally get the saliency map S_{M1} at scale $l_{\max} = 1$ of the original infrared image.

Step 2. For the coarse scale $l_{\min} = 0.7$, resize the original infrared image to get a cutdown image and, at the same time, reduce the size of the search window of local sparse representation to $\sqrt{t+1} \times \sqrt{t+1}$, where $t = \lfloor k \times l_{\min} \rfloor$. Similar to Step 1, we can obtain the saliency map S_{M2} at scale $l_{\min} = 0.7$ of the original infrared image. Note that S_{M2} is ultimately rescaled to the original size of the infrared image.

Step 3. Combine the saliency map S_{M1} and the saliency map S_{M2} by weighted additive fusion:

$$S_M = \lambda \times S_{M1} + (1 - \lambda) \times S_{M2}, \tag{6}$$

where $\lambda = 0.5$ gives the same importance of S_{M1} and S_{M2}. The final multiscale local sparse representation based saliency map S_M is obtained for the original infrared image. Since this result combines the advantages of multiscale analysis and local sparse representation, the salient region can be well highlighted. Nevertheless, when the infrared image has complex background clutters, this saliency map will include much background noise. Therefore, in the following section, we will adopt the local contrast measure [9] as an auxiliary tool to solve this problem.

2.2. Local Contrast Measure Based Saliency Computation. As an important factor in human visual perception, contrast has been extensively studied by researchers in psychology and computer vision. Numerous methods have been proposed computing contrast saliency maps using various visual properties, such as color, intensity, texture, and structure. For infrared images, intensity is usually regarded as a very important and distinctive visual characteristic [9]. Motivated by these, we focus on calculating intensity contrast for saliency detection for infrared images, which is unaffected by temperature variations in infrared image applications.

Considering that infrared images usually exhibit much noise due to the infrared sensors, electronic circuit noise, background noise, and so forth, to reduce the influence of noise and generate a more robust saliency map, we calculate a saliency map by using local contrast measure as follows.

First, given an infrared image F, slide a window on it from up to down and left to right at a certain step to get a set of subblocks.

Second, let o be a subblock. Calculate the average intensity value u_0 for o:

$$u_0 = \frac{1}{N_o} \sum_{(p,q) \in o} F(p,q), \tag{7}$$

where N_o represents the number of pixels in o. Also, search the maximum intensity value $I_{\max}$ in o:

$$I_{\max} = \max(F(p,q)), \quad (p,q) \in o. \tag{8}$$

Third, apply an image area whose side length is three times the subblock's side length on the image and take o as the central subblock. Then eight adjacent subblocks of o can be obtained in this area. Calculate their average intensity values $u_1, \ldots, u_8$. The local contrast measure of o can be defined by

$$\text{LCM} = \min_i \frac{I_{\max} u_0}{u_i}, \quad i = 1, \ldots, 8. \tag{9}$$

Obviously, if the central block o is a target block, $\max(u_i)$ is usually lower than $I_{\max}$. Thus, LCM $> u_0$, then target is enhanced. If o is a background block, there may be $\max(u_i) \geq I_{\max}$. Thus, LCM $\leq u_0$; then the background is suppressed.

Finally, apply the above local contrast measure to the whole infrared image, we can obtain a local contrast measure based saliency map S_{LCM}. In this map, the background noise can be effectively restrained.

2.3. Adaptive Saliency Map Combination. After obtaining the multiscale local sparse representation based saliency map S_M and the local contrast measure based saliency map S_{LCM}, we design an adaptive fusion scheme to combine them together.

In the human visual attention, the fusion of different characteristics is very important. Currently, there exist various fusion methods, such as additive fusion [20], multiplicative fusion [21], and maximum fusion [22]. Considering that multiplicative fusion is likely to lose much information, while maximum fusion is easily influenced by background noise, we select additive fusion for saliency maps combination.

For additive fusion, computing the weight of each saliency map is an interesting and important task. There are two ways for the weight calculation: one is the fixed weighting methods [20], and another is the adaptive weighting method [19]. Comparing these two kinds of weighting methods, the former is simple and fast to implement but lacks flexibility and easily results in a bad result; on the contrary, the latter has strong flexibility and adaptability and usually achieves better fusion effect. Therefore, here we adopt a mutual consistency guided fusion method, the effectiveness of which has been verified in our recent related work [19], to adaptively combine S_M and S_{LCM}.

Given the multiscale local sparse representation based saliency map S_M and the local contrast measure based saliency map S_{LCM}, their mutual consistencies are first calculated:

$$P_M = \frac{\sum_{i,j} [S_M(i,j) \times S_{\text{LCM}}(i,j)]}{\sum_{i,j} [S_M(i,j)]}$$

$$P_{\text{LCM}} = \frac{\sum_{i,j} [S_M(i,j) \times S_{\text{LCM}}(i,j)]}{\sum_{i,j} [S_{\text{LCM}}(i,j)]}, \tag{10}$$

where P_M denotes consistency of $S_M(i,j)$ relative to $S_{\text{LCM}}(i,j)$. P_{LCM} denotes the consistency of $S_{\text{LCM}}(i,j)$ relative to $S_M(i,j)$. Both P_M and P_{LCM} take values in $[0, 1]$.

Then, the relative adaptive weights for $S_M(i,j)$ and $S_{\text{LCM}}(i,j)$ are calculated by

$$w_M = \frac{P_M}{P_M + P_{\text{LCM}}}$$

$$w_{\text{LCM}} = \frac{P_{\text{LCM}}}{P_M + P_{\text{LCM}}}. \tag{11}$$

Finally, the fused saliency map S is obtained by

$$S(i,j) = w_M \times S_M(i,j) + w_{\text{LCM}} \times S_{\text{LCM}}(i,j). \tag{12}$$

Figures 2 and 3 show two saliency detection examples. For each example, the original infrared image, the multiscale
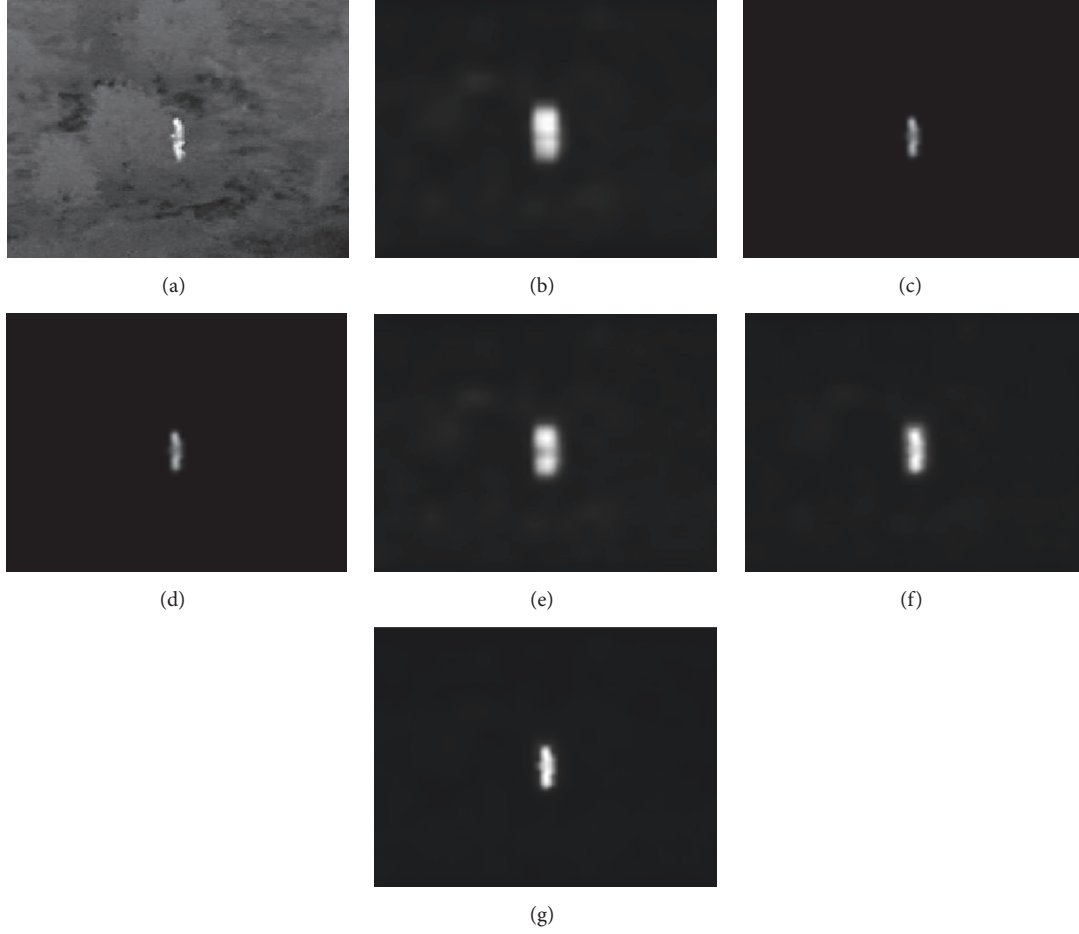
FIGURE 2: Saliency maps generated by different fusion schemes (the salient object is a pedestrian). (a) Original infrared image. (b) Multiscale local sparse representation based saliency map. (c) Local contrast measure based saliency map. (d)–(g) Saliency maps generated by multiplicative fusion, maximum fusion, additive fusion with fixed weights, and our adaptive fusion approaches, respectively.

local sparse representation based saliency map, the local contrast measure based saliency map, and the combined saliency maps using different fusion strategies, including multiplicative fusion, maximum fusion, additive fusion of fixed weights (i.e., $w_M = w_{LCM} = 0.5$), and additive fusion of our adaptive weights, are illustrated. From these two examples, we can see that the multiscale local sparse representation based saliency map can well highlight the whole salient object, but it includes too much noise (see Figures 2(b) and 3(b)). While the saliency map based on local contrast measure can effectively suppress the background noise, but the detected salient region is incomplete (see Figures 2(c) and 3(c)). In order to achieve better results, these two kinds of saliency maps are combined together. By comparing various fusion strategies, it can be seen that, since multiplicative fusion loses much information, the salient region is incomplete in the fused saliency map (see Figures 2(d) and 3(d)); maximum fusion easily tends to the response of one predominant channel and cannot effectively suppress background noise (see Figures 2(e) and 3(e)); fixed weights-based additive fusion method lacks flexibility, so it also leads to bad results (see Figures 2(f) and 3(f)). On the contrary, our

adaptive fusion method produces superior results. As shown in Figures 2(g) and 3(g), it can not only enhance the salient regions, but also simultaneously suppress the background noise.

3. Experimental Results

In this section, we evaluate the performance of the proposed method. We compare it with ground truth data as well as eight state-of-the-art saliency detection algorithms. The first selected model, called a structured matrix decomposition model, was recently proposed by Peng et al. [23] for salient object detection. It utilized a tree-structured sparsity-inducing regularization and a Laplacian regularization to complete the saliency detection (we refer to this method as SMD). Second, Chen et al. [24] presented a contrast measuring method for infrared salient object detection. It firstly measured the dissimilarity between the current location and its neighborhoods and then used an adaptive threshold to obtain the salient object (we refer to this method as LC). Third, Hou and Zhang [25] proposed a dynamic visual attention model for saliency detection, which introduced the Incremental

(a)   (b)   (c)

(d)   (e)   (f)
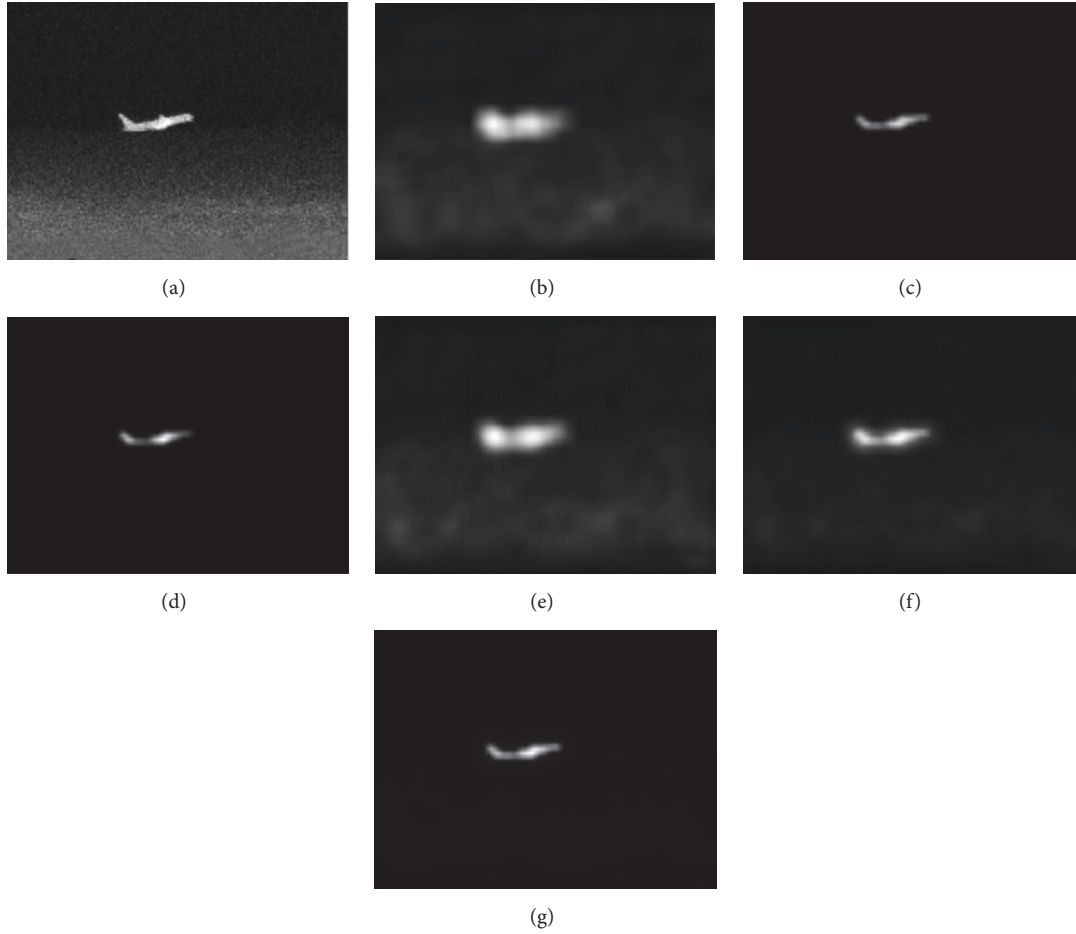
(g)

FIGURE 3: Saliency maps generated by different fusion schemes (the salient object is an airplane). (a) Original infrared image. (b) Multiscale local sparse representation based saliency map. (c) Local contrast measure based saliency map. (d)–(g) Saliency maps generated by multiplicative fusion, maximum fusion, additive fusion with fixed weights, and our adaptive fusion approaches, respectively.

Coding Length to measure the perspective entropy gain of each feature (we refer to this method as ICL). Fourth, the selected model was proposed by Itti et al. [5]. In this model, color, intensity, and orientation features were used to calculate saliency in an image: given an image, it was first decomposed into color, intensity, and orientation channels to generate three conspicuity maps by using the multiscale and center-surround operations; then these three conspicuity maps were linearly combined together to produce the final saliency map for the input image (we refer to this method as ITTI). Fifth, in [26], Zhai and Shah presented a visual attention detection algorithm using spatiotemporal cues. In the spatial attention model, they developed a fast method for computing pixel-level saliency maps using color histograms of images (We refer to this method as SC). Sixth, in [27], Hou and Zhang proposed a spectral residual approach for visual saliency detection. This model was independent of features, categories, or other forms of prior knowledge of the objects. By analyzing the log-spectrum of an input image, the spectral residual of an image was extracted in spectral domain, and then the corresponding saliency map was constructed in spatial domain (we refer to this method as SR). Seventh,

in [28], Achanta et al. provided a frequency-tuned saliency model relying on difference of Gaussian bandpass filters. This method exploited features of color and luminance and was computationally efficient (we refer to this method as FT). The last selected saliency detection model for comparison was based on local sparse representation [15], which divided a test image into a number of patches at first; then, each patch was sparsely coded with its surrounding patches. Based on the learned dictionary, the sparse reconstruction error of each patch was computed as the saliency values for the corresponding patch (we refer to this method as LSR).

In this paper, we mainly compare the above methods by using the publicly available dataset [29]. Our test dataset contains a number of various infrared images with different salient objects, such as pedestrian, airplane, and vessel. The salient regions in each image are manually labeled to generate the ground truths. Both qualitative evaluation and quantitative evaluation are made.

3.1. Qualitative Evaluation. In this section, we present some qualitative comparison results of our algorithm and other eight methods, as shown in Figures 4 and 5. Figure 4 displays
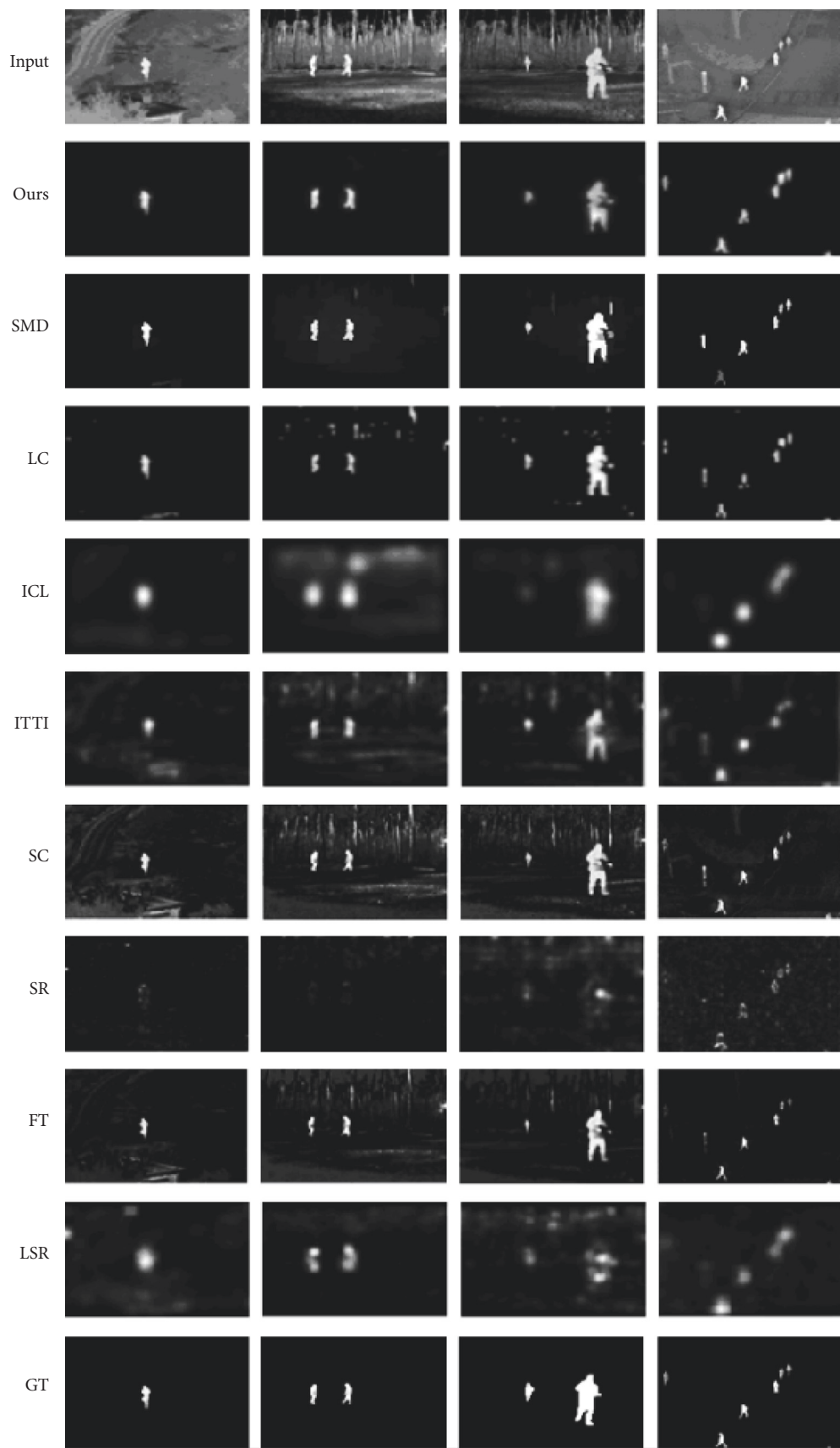
FIGURE 4: Saliency detection results generated from different approaches (the salient objects are pedestrians).
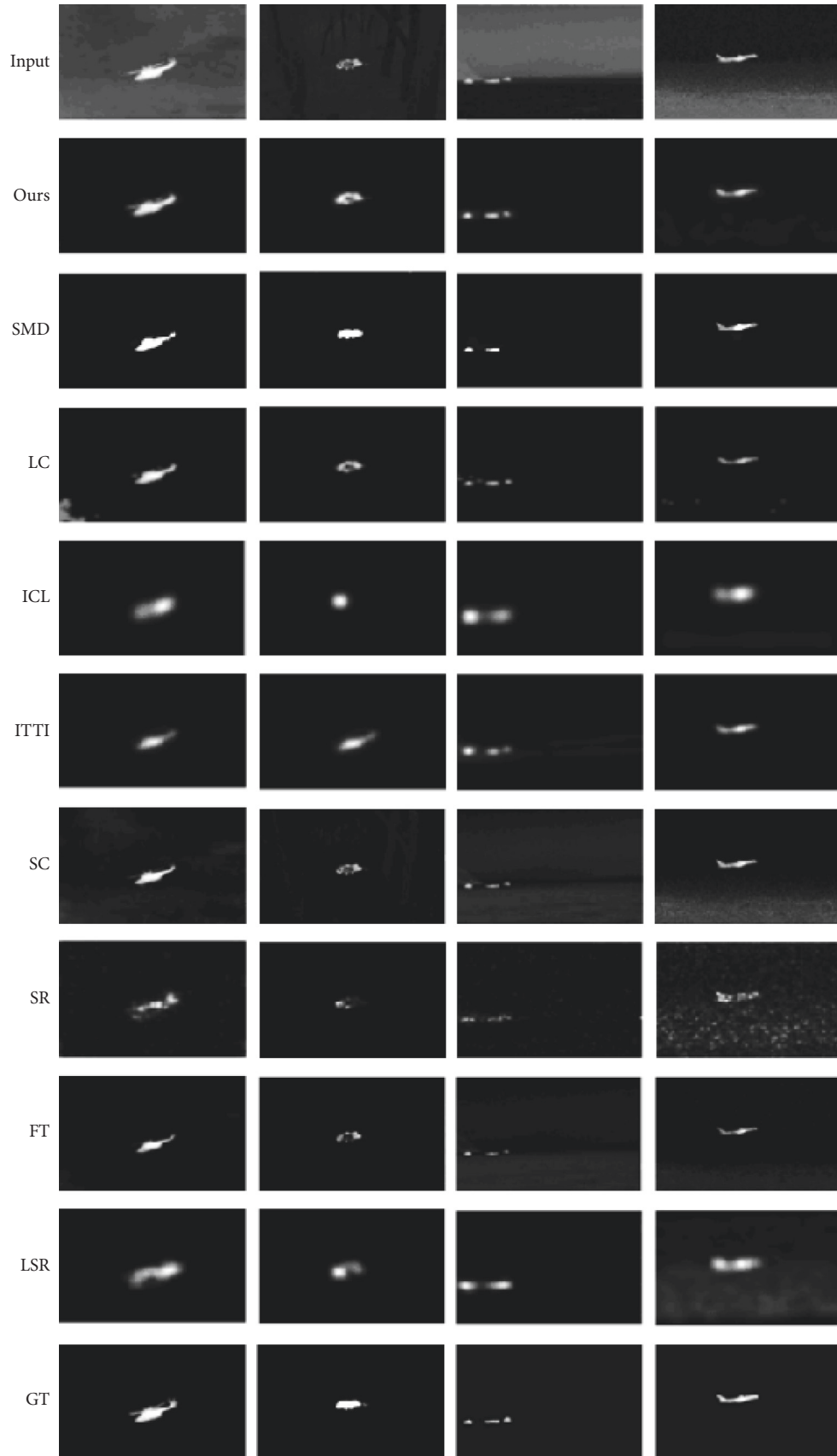
FIGURE 5: Saliency detection results generated from different approaches (the salient objects are planes, vessels, etc.).

TABLE 1: Computational cost comparison among different saliency detection methods for a single image.

Evaluation indicators	Ours	SMD	LC	ICL	ITTI	SC	SR	FT	LSR
Time (sec)	2.701	1.050	0.989	0.408	0.366	0.452	0.187	0.467	1.553
Code	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab	Matlab

the computed saliency maps of infrared images which contain various salient human objects, while Figure 5 shows the computed saliency maps of infrared images which include different salient nonhuman objects.

From Figures 4 and 5, we can see that our method can highlight all the salient regions completely no matter how many human objects the images have, and at the same time, it can suppress the background noise effectively. On the contrary, other eight approaches achieve inferior performance for infrared images. For example, SMD, LC, SC, and FT can highlight most salient regions, but the background noise cannot be suppressed. ICL, LSR, and ITTI tend to generate diffuse saliency maps. SR produces the worst performance, since it not only fails to detect the salient regions, but also contains too much noise. Thus, the proposed approach outperforms the other methods.

3.2. Quantitative Evaluation. To further evaluate our algorithm, we also make quantitative comparison. We select three metrics, namely, recall, precision, and F-measure [30, 31], which are defined as follows:

$$\text{recall} = \frac{\text{num}\,(\text{sm} \cap \text{gt})}{\text{num}\,(\text{gt})}$$

$$\text{precision} = \frac{\text{num}\,(\text{sm} \cap \text{gt})}{\text{num}\,(\text{sm})} \qquad (13)$$

$$F\text{-measure} = \frac{\left(1 + \beta^2\right) * \text{precision} * \text{recall}}{\beta^2 * \text{precision} + \text{recall}},$$

where sm denotes the binary salient region mask and gt is the corresponding binary ground truth. num(sm) and num(gt) are the numbers of pixels in sm and gt, respectively. In general, precision and recall are two conflicting goals. To consider precision and recall simultaneously, F-measure metric is always calculated. Similar to [32], β^2 is set to 0.3. The binary salient region mask sm is found by applying a simple threshold operation on a saliency map [21].

The quantitatively results are shown in Figure 6. As can be seen, our method achieves the best performance compared with the state-of-the-art saliency detection algorithms for infrared images. It should be emphasized again that such favorable results can be obtained since our proposed method is especially designed for infrared images based on the analysis of characteristics of infrared images.

Besides, all experiments are implemented on an Intel Dual Core 2.3 GHz laptop with 4 GB RAM. The programming platform is Matlab R2013b. We compute the saliency detection time for all test images with the above-mentioned
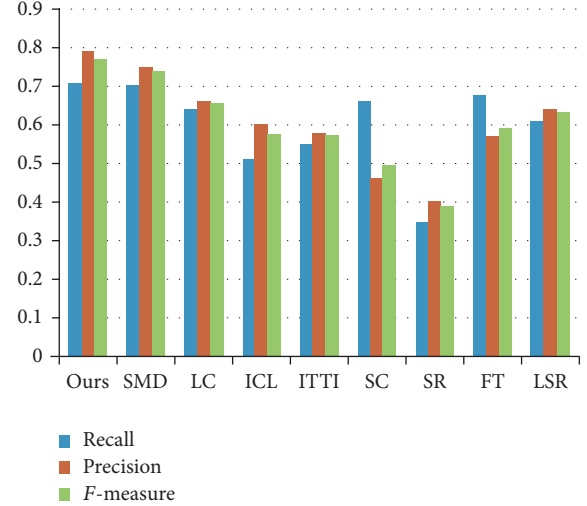
FIGURE 6: Recall, precision, and F-measure achieved using different saliency detection methods.

algorithms. The results are reported in Table 1, where the time is computed as an average over all the test images.

When compared with other eight methods, that is, SMD, LC, ICL, ITTI, SC, SR, FT, and LSR, which only take about 1-2 seconds to handle one image, the efficiency of our multiscale local sparse representation and local contrast measure based approach is much lower, for it involves online learning processes. As shown in Table 1, for an image with the size of 320×400, our average computational time is about 3 s in the above experimental setup. An efficient C/C++ implement or even a parallel architecture would reduce the overall execution time dramatically and would make the proposed method feasible for real-world application.

4. Conclusion

This paper proposes a novel saliency detection method for infrared images based on multiscale local sparse representation and local contrast measure. We first develop a multiscale local sparse representation based approach to compute saliency for infrared images, which can highlight the salient regions fully. We then adopt a local contrast measure based scheme to compute saliency for infrared images, which can suppress the background noise effectively. By incorporating these two categories of saliency maps into a unified one, we can generate reliable saliency maps. Compared with several up-to-date approaches, we have shown that the proposed method is able to achieve more accurate and robust results of saliency detection for infrared images. In the future, we

will apply it to infrared video analysis, such as video retrieval, abstraction, and event recognition.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013.

[2] N. Liu and J. Han, "DHSNet: deep hierarchical saliency network for salient object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*, pp. 678–686, June 2016.

[3] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 185–198, 2010.

[4] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS '07)*, vol. 19, pp. 545–552, 2007.

[5] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[6] S. Qin, L. Wang, H. Cheng, Q. Feng, M. Zhang, and C. Gao, "Infrared image saliency detection based on human vision and information theory," in *Proceedings of the International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, pp. 484–488, October 2016.

[7] B. C. Ko, D. Y. Kim, and J. Y. Nam, "Detecting humans using luminance saliency in thermal images," *Optics Letters*, vol. 37, no. 20, pp. 4350–4352, 2012.

[8] X. Wang, G. Lv, and L. Xu, "Infrared dim target detection based on visual attention," *Infrared Physics and Technology*, vol. 55, no. 6, pp. 513–521, 2012.

[9] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, and Y. Fang, "A robust infrared small target detection algorithm based on human visual system," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 12, pp. 2168–2172, 2014.

[10] S. Qi, J. Ma, C. Tao, C. Yang, and J. Tian, "A robust directional saliency-based method for infrared small-target detection under various complex backgrounds," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 495–499, 2013.

[11] X. Wang, S. Shen, C. Ning, F. Huang, and H. Gao, "Multiclass remote sensing object recognition based on discriminative sparse representation," *Applied Optics*, vol. 55, no. 6, pp. 1381–1394, 2016.

[12] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[13] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3336–3351, 2014.

[14] X. Wang, S. Shen, C. Ning, M. Xu, and X. Yan, "A sparse representation-based method for infrared dim target detection under sea-sky background," *Infrared Physics and Technology*, vol. 71, pp. 347–355, 2015.

[15] C. Xia, P. Wang, F. Qi, and G. Shi, "Nonlocal center-surround reconstruction-based bottom-up saliency estimation," *Pattern Recogn*, vol. 48, no. 4, pp. 206–210, 2015.

[16] M. M. S. Fareed, G. Ahmed, and Q. Chun, "Salient region detection through sparse reconstruction and graph-based ranking," *Journal of Visual Communication and Image Representation*, vol. 32, pp. 144–155, 2015.

[17] I. Rigas, G. Economou, and S. Fotopoulos, "Efficient modeling of visual saliency based on local sparse representation and the use of hamming distance," *Computer Vision and Image Understanding*, vol. 134, pp. 33–45, 2015.

[18] Q. Fan and C. Qi, "Saliency detection based on global and local short-term sparse representation," *Neurocomputing175*, pp. 81–89, 2016.

[19] X. Wang, C. Ning, and L. Xu, "Saliency detection using mutual consistency-guided spatial cues combination," *Infrared Physics and Technology*, vol. 72, pp. 106–116, 2015.

[20] Z. Ren, S. Gao, L.-T. Chia, and D. Rajan, "Regularized feature reconstruction for spatio-temporal saliency detection," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3120–3132, 2013.

[21] W. Kim and C. Kim, "Spatiotemporal saliency detection using textural contrast and its applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 4, pp. 646–659, 2014.

[22] Y. Li, Y. Tan, J.-G. Yu, S. Qi, and J. Tian, "Kernel regression in mixed feature spaces for spatio-temporal saliency detection," *Computer Vision and Image Understanding*, vol. 135, pp. 126–140, 2015.

[23] H. Peng, B. Li, H. Ling, W. Hu, W. Xiong, and S. J. Maybank, "Salient object detection via structured matrix decomposition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 818–832, 2017.

[24] C. L. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A local contrast method for small infrared target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 574–581, 2014.

[25] X. Hou and L. Zhang, "Dynamic visual attention: searching for coding length increments," in *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS '08)*, pp. 681–688, December 2008.

[26] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th Annual ACM International Conference on Multimedia (MULTIMEDIA '06)*, pp. 815–824, October 2006.

[27] X. D. Hou and L. Q. Zhang, "Saliency detection: a spectral residual approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, June 2007.

[28] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proceedings*

of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 1597–1604, June 2009.

[29] J. W. Davis, OTCBVS Benchmark Dataset Collection, http://vcipl-okstate.org/pbvs/bench/.

[30] J.-G. Yu and J. Tian, "Saliency detection using midlevel visual cues," *Optics Letters*, vol. 37, no. 23, pp. 4994–4996, 2012.

[31] J. Han, S. He, X. Qian, D. Wang, L. Guo, and T. Liu, "An object-oriented visual saliency detection framework based on sparse coding representations," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 12, pp. 2009–2021, 2013.

[32] L. Zhang, Y. Zhang, W. Wei, and Q. Meng, "An associative saliency segmentation method for infrared targets," in *Proceedings of 2013 20th IEEE International Conference on Image Processing, ICIP 2013*, pp. 4264–4268, aus, September 2013.