Infrared Image Saliency Detection based on Human Vision and Information Theory

Shiyu Qin*, Lan Wang*, Hua Cheng†, Qi Feng*, Minwen Zhang*, Chenqiang Gao*

*Key Laboratory of Signal and Information Processing of Chongqing
Chongqing University of Posts and Telecommunications
Chongqing, China

†China Ship Development and Design Center
Wuhan, China

Abstract—**Robust image saliency detection can process the image correctly without any prior knowledge and additional assumptions. Therefore, the saliency detection is still one of the important steps in the field of computer vision including object recognition and tracking, image and video encoding and image segmentation. Although infrared imaging has extensive applications, there is few saliency extraction algorithms based on infrared spectroscopy. We propose an infrared image-based saliency extraction algorithm based on human vision and information theory. The proposed algorithm uses both human visual attention mechanism and theory of information, and it can also produce a saliency image with full resolution. The detection results of the proposed algorithm get a higher accuracy and better recall rate, when tested on one of the largest infrared data sets which is publicly now and a data set created by ourselves.**

Keywords- Infrared Image; Saliency Detection; Human Vision; Information Theory

I. INTRODUCTION

Human beings are receiving an amount of image information all the time, and can focus on the important regions of the images easily. Such region is called a saliency region. It will dramatically improve resource utilization, if the computer is able to extract the saliency image and allocate resources preferentially when processing an image. So it is necessary to calculate and detect saliency regions of the image in the following research fields: object recognition [1] and tracking [2], image and video encoding [3], image segmentation [4], image scaling [5], and image searching [6]. Saliency detection is closely related to the processing of human perceptual system which is involved in multiple disciplines including computer vision [7], behavioral science [8], cognitive science [9], neurology [10], and information science [11].

Currently, most saliency research works are based on visible light instead of infrared images, because of the expensiveness of infrared imaging device. With the decreasing cost and increasing quality of infrared sensor, infrared image studies gradually come into researchers' sight. At the same time, the infrared thermal imaging instrument has the advantages of strong anti-interference ability, good concealment, high sensitivity, all-weather work, mobility and so on. In civil fields, infrared thermal imaging system can be used in the detection and evaluation of material defects, equipment state thermal diagnosis, production process monitoring, disaster reduction and disaster prevention. Nevertheless, there is a strong noise and background clutter interference in the receiver of the infrared imaging instrument, leading to the low contrast, SNR (signal to noise ratio) and resolution of infrared images [12]. Effective algorithm should be explored to handle the difficulties of infrared saliency detection.

As the information theory presented, information can be divided into redundancy and changing components. Human's vision is more sensitive to changing components. One of the basic principles in the visual system is to suppress the response to the frequently occurring features, and remain sensitive to the nonconventional features at the same time. On the other hand, the theory of human visual attention suggests that the theory of human visual attention: The visual system of human is very sensitive to the contrast.

We proposed an infrared image saliency detection algorithm based on human vision and information theory. Using one of the characteristics of human visual system and the theory that information can be divided into redundant and changing parts in information theory, we produce a saliency image with full resolution. Our algorithm is dedicated to the processing of infrared images.

We have carried out an objective test of our algorithm on the infrared data set, and compared with some saliency detection algorithms [7,13,14,15]. The results show an obvious improvement on precision and recall. Generally speaking, at the cost of a slight increase in the computation time, we get a higher correct rate and recall rate.

II. RELATED WORK

There are two kinds of saliency detection algorithm that one of them is task independent and bottom-up data driven, the other is task dependent and top-down target driven. The proposed algorithm in this paper is a bottom-up saliency one.

Itti and Koch [16] proposed the It model which simulate biological visual attention mechanisms. The model based on research results of feature integration theory and Behavioral science, image is decomposed into three categories: the features of visual brightness, color and direction. The idea of Hou and Zhang [17] based on similarity representation redundancy is proposed, which is based on spectral redundancy. This method is proposed for the first time Saliency

This research is funded by the National Natural Science Foundation of China (No.61571071, 61102131), the Natural Science Foundation of Chongqing Science and Technology Commission (No. cstc2014jcyjA40048), the Wenfeng innovation and start-up project of Chongqing University of Posts and Telecommunications (No. WF201404).

were calculated in frequency domain. Achanta et al. [7] from the frequency domain perspective, proposed a new approach to detect the salient regions based on global correlation for the first time. Hu et al. [18] proposed a robust subspace estimation and saliency detection algorithm based on analysis technology. In this method, firstly, the input image is transformed into polar coordinate system, and then a new linear subspace estimation method based on generalized principal component analysis is used to represent the transformed image. At last, saliency of the image sub region is calculated by using the statistical value of the projection distance of the pixels on the estimated subspace. Davis [19] and Jiang transform the problem of significant region detection into the problem of the objective function with the characteristic of the sub model. Mai et al. [20] proposed an approach which is based on data driven significant aggregation. This approach can gain a better saliency region detection result by integrating the saliency map generated by the multiple saliency detection method.

These algorithms have been able to achieve very good results in the visible light image, but because of the particularity of infrared image, the algorithms are often not achieving good results when they are applied directly to the infrared images. So the research of saliency detection algorithm for infrared image is of great practical importance.

III. PROPOSED ALGORITHM

In our approach, we use the theory of human visual attention mechanism and information theory simultaneously. First, the input image is segmented into regions. Then we calculate the contrast of each region to other regions and the spectral residual images of whole image. Finally, the spatial correlation and spectral residual factors of spatial distance information are combined to get the final saliency image.

A. Human vision-based attention mechanism

The study of visual saliency in biological vision shows that the human visual system is very sensitive to the contrast information of the visual signal, especially those image regions which are extremely distinct from surrounding. So, the region based contrast method is applied to obtain the course saliency map [13]. The approach of graph-based segmentation is used to segment the image into regions [21], and then the color histogram for each region is built. The saliency value of region r_k is computed by

$$S_V\left(r_k\right)=\sum_{r_k \neq r_i} w\left(r_i\right) D_b\left(r_k, r_i\right) \exp\left(-D_s(\mathrm{r}_k, \mathrm{r}_i) / \delta^2\right), \quad (1)$$

where $D_s(\mathrm{r}_k, \mathrm{r}_i)$ is the Euclidean distance between centroids of regions r_k and r_i. δ controls the intensity of spatial weighting. The larger the values of δ is, the less the effect of spatial weighting is. Thus, contrast with farther regions would contribute more to the saliency of the current region. $w\left(r_i\right)$ is the weight of region r_i and define $D_b\left(r_k, r_i\right)$ as the brightness distance between two regions,

$$D_b\left(r_1, r_2\right)=\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} f\left(k_1, i\right) f\left(k_2, j\right) D\left(k_{1,i}, k_{2,j}\right). \quad (2)$$

$f\left(k_1, i\right)$ is the probability that the i-th luminance $k_{1,i}$ occurs in all n_1 kinds of luminance of region r_1 and $f\left(k_2, i\right)$ is as the same.

However, this kind of method cannot extract the salient regions accurately in infrared domain because there are few fine grained features in infrared images. On the other hand, the differences between regions are indistinct, generating low region contrast.

B. Mechanism based on information theory

Inspired by [14], we utilize the information theory to separate the prominent part of the infrared images correctly. According to the theory, the image can be divided into two parts as follows:

$$H=H_P+H_R. \quad (3)$$

H_P represents the prominent information part in the image, and H_R represents the redundant information part in the image. Removing the redundant part of the image information can generate the prominent part of the image information, the significant region.

To solve the low region contrast of the infrared images, we add a weight component to the regions according to the prominent information. For a region r_k, we first calculate the prominent information $P\left(I_i\right)$ for the entire infrared image I_i. Add the prominent information values within the range of r_k together as the weight $w_I\left(r_k\right)$.

$$P\left(I_i\right)=g\left(r_k\right) * \mathscr{F}^{-1}\left[\exp\left(R\left(I_i\right)+P\left(I_i\right)\right)\right]^2, \quad (4)$$

where $R\left(I_i\right)$ is the spectral residual obtained by Log amplitude Spectra $L_A\left(I_i\right)$ minus $L_A\left(I_i\right)$ after mean filtering

$$R\left(I_i\right)=L_A\left(I_i\right)-h_n\left(I_i\right) * L_A\left(I_i\right). \quad (5)$$

h_n is the convolution kernel of a n*n mean filtering, and we set n=3 in our experiment. The log amplitude spectra were obtained after the logarithm of the amplitude spectrum,

$$L_A(I_i) = \log(A(I_i)), \tag{6}$$

$$A(I_i) = \Re(\mathscr{F}[I_i]). \tag{7}$$

$P(I_i)$ is the phase spectrum computed by

$$P(I_i) = \Im(\mathscr{F}[I_i]). \tag{8}$$

Then we normalize the weights to [0, 1],

$$w_I^*(r_k) = \frac{w_I(r_k)}{\sum_k w_I(r_k)}. \tag{9}$$

Ultimately, we add the prominent weight to the region contrast computing by

$$S(r_k) = \sum_{r_k \neq r_i} w(r_i) D_b(r_k, r_i) \exp\left(-D_s(r_k, r_i)/\sigma^2\right) w_I^*(r_k) \tag{10}$$

In our implementation, we use $\delta^2 = 0.4$.

IV. EXPERIMENTS

We use the GUIDIR IR300 refrigeration infrared thermal imager to shoot and produce a new infrared data set, including 1062 images in size of 293 x 256 pixels totally. We manually mark the saliency region as the ground truth (GT) for the algorithm testing. Our algorithm is compared with frequency-tuned salient region detection (FT) algorithm [7], histogram-based contrast method (HC) algorithm [13], region-based contrast (RC) algorithm [13], a spectral residual approach (SR) algorithm [14], and Visual attention detection in video sequences using spatiotemporal cues (LC) algorithm [15].

In order to test the detection effect of our proposed algorithm objectively, as in [7], we set the threshold to change from 0 to 255. We use all threshold values to get binarization saliency image, and then get the accuracy and recall rate. In addition, we also compared the F-measurement (Figure 7) on the test data set, where the measurement is defined as:

$$F_\beta = \frac{(1+\beta^2) * \mathrm{P} * \mathrm{R}}{\beta^2 * \mathrm{P} + \mathrm{R}}. \tag{11}$$

We assume that the weight of the precision is greater than the recall. P represents Precision and R represents Recall. Our method get a better result than other 5 methods after compared in figure 1.

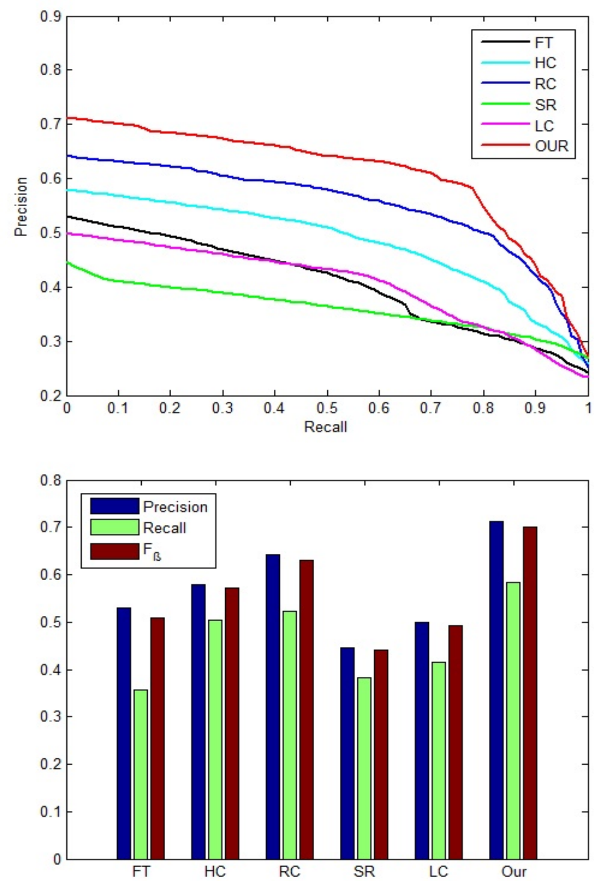
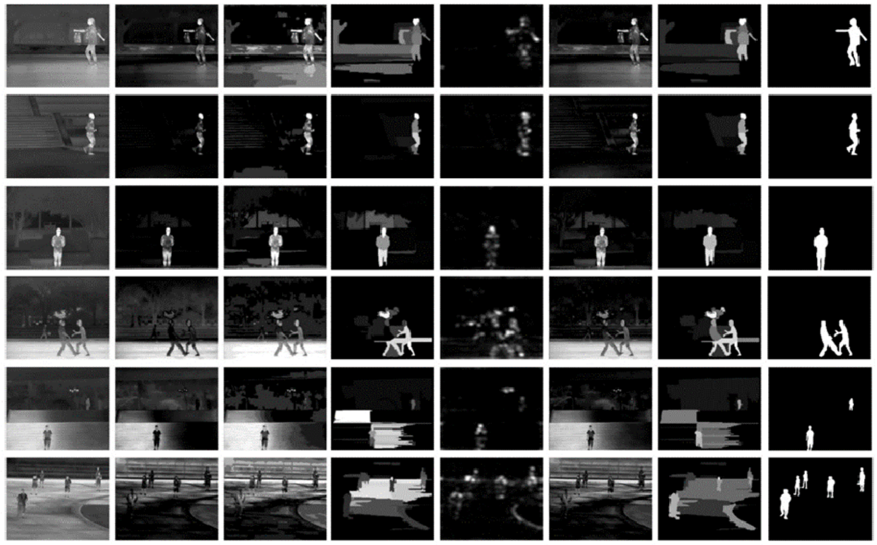


Figure 1. PR curves (left) and F0.3 metric (right) comparison

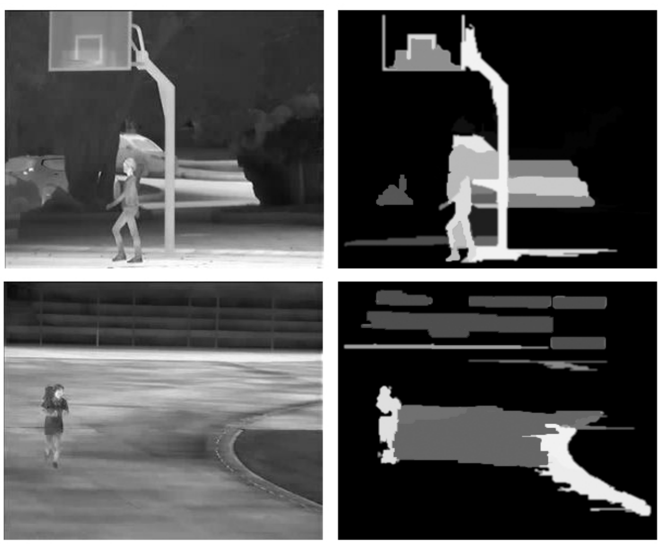
Saliency maps obtained by different algorithms are shown in Figure 2, (a) as the original input image, (b) and (c) are the result of the FT saliency detection algorithm and the HC saliency detection algorithm respectively. These two algorithm are both based on the pixels, for some rare pixels, the detection effect is obvious, but when the pixel distribution is more complex, the consistency of test results is usually poor. (d) is the result of the RC saliency detection algorithm. The algorithm is on good consistency in the region, but for low contrast infrared image, due to the histogram is relatively close of the region, but the content is different, it will affect the detection results. (e) is the result of the SR saliency detection algorithm, it is sensitive to the frequency information of the infrared image, which is the texture of the edge or the change, but the region detection effect is not good; (f) is the result of the LC saliency detection algorithm. This method is very simple, but the strategy for the global contrast resulted in rare brightness value is dominant, with high contrast, which in many cases is not reasonable. (g) is the image of saliency



(a)original (b)FT (c)HC (d)RC (e)SR (f)LC (g)Ours (h)GT

Figure 2. Saliency image comparison

algorithm proposed in this paper. (h) is our manual precise marking ground truth. From figure 2, we can directly see the algorithm to get the saliency effect is more excellent, better suppression of the background information, better highlight the significant information.



(a)original image (b)Ours

Figure 3. An example case challenging for our algorithm

Figure 3 shows two difficult cases of our algorithm. In the saliency map of first example, we cannot clearly observe the entire pedestrian, as the pedestrian's head is divided into the area with other elements. In the second example, the edge of the runway is also detected as a saliency area, which is because Residual Spectral is sensitive to some high-frequency regions.

Our algorithm and comparison algorithms are tested using uniform experimental platform with Intel Celeron E3400 2.6GHz CPU, and 4GB RAM. We implemented the algorithm in C++, and then calculate the average time of each algorithm (Table 1). Although the algorithm of this paper takes more time, it has a better effect.

TABLE I. COMPARISON OF THE AVERAGE TIME OF EACH ALGORITHM

Algorithm	**FT**	**HC**	**RC**	**SR**	**LC**	**Our**
Time(s)	0.491	0.904	3.557	0.381	0.586	4.362

V. CONCLUSION AND FUTURE WORK

We propose a method for the saliency computation of infrared images based on the global and local information. Additionally, we collect and annotate a new infrared data set for saliency detection, and then we have carried out an objective test of our algorithm on the data set and compared with the previous 5 other methods. The experimental results show that our proposed method is superior to other methods in accuracy rate and recall rate.

In the future, we will use a better segmentation algorithm to output a more accurate result. At the same time, the problem of high sensitivity of Residual Spectral for some non-salient regions is further studied. Finally, we plan some applications based on infrared saliency image detection, such as infrared small target detection and tracking, pedestrian detection at night time, etc.

ACKNOWLEDGMENT

This research is funded by the National Natural Science Foundation of China (No.61571071, 61102131), the Natural Science Foundation of Chongqing Science and Technology Commission (No. cstc2014jcyjA40048), the Wenfeng

innovation and start-up project of Chongqing University of Posts and Telecommunications (No. WF201404).

REFERENCES

[1] U. Rutishauser, D. Walther, C. Koch, and P. Perona. “Is bottom-up attention useful for object recognition? ” Computer Vision and Pattern Recognition, 2004, pp. 37-44.

[2] Fan J, Wu Y, Dai S. “Discriminative spatial attention for robust tracking,”Europe Conference on Computer Vision Springer, 2010, pp. 480–493.

[3] Li Z, Qin S, Itti L. “Visual attention guided bit allocation in video compression,”Image and Vision Computing, 2011, 29(1), pp.1–14.

[4] J. Han, K. Ngan, M. Li, and H. Zhang. “Unsupervised extrac¬tion of visual attention objects in color images,” IEEE Transactions on Circuits and Systems for Video Technology, 2006, 16(1), pp.141-145.

[5] C. Christopoulos, A. Skodras, and T. Ebrahimi. “The JPEG2000 still image coding system: an overview,” IEEE transactions on consumer electronics, 2000, 46(4), pp.1103-1127.

[13] Cheng, M. M., Mitra, N. J., Huang, X., Torr, P. H., and Hu, S. M.. “Global contrast based salient region detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3), pp.569-582.

[14] X. Hou and L. Zhang. “Saliency detection: A spectral residual approach,” IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp.1-8.

[15] Y. Zhai and M. Shah. “Visual attention detection in video sequences using spatiotemporal cues,” Proceedings of the 14th ACM international conference on Multimedia, 2006, pp.815-824.

[16] 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, 1998, 20(11), pp.1254-1259.

[17] Hou X, Zhang L. “Saliency detection: A spectral residual approach,” IEEE Conference on Computer Vision and Pattern Recognition. 2007, pp.1–8.

[6] Yang L, Geng B, Cai Y, et al. “Object retrieval using visual query context,” IEEE Transactions on Multimedia, 2011, 13(6), pp.1295–1307.

[7] R. Achanta, S. Hemami, F. Estrada, and S. SUsstrunk. Frequency-tuned salient region detection. Computer vision and pattern recognition, 2009, pp. 1597-1604.

[8] Beck D M, Kastner S. “Top-down and bottom-up mechanisms in biasing competition in the human brain,” Vision research, 2009, 49(10), pp.1154–1165.

[9] J. M. Wolfe and T. S. Horowitz. “What attributes guide the deployment of visual attention and how do they do it?” Nature reviews neuroscience, 2004, 5(6), pp.495-501.

[10] S. K. Mannan, C. Kennard, and M. Husain. “The role of visual salience in directing eye movements in visual object agnosia,” Current biology, 2009, 19(6), pp.R247-R248.

[11] X. Hou and L. Zhang. Saliency detection: “A spectral residual approach,” IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp.1-8.

[12] Zhang, L., Zhang, Y., Wei, W., and Meng, Q.. “An associative saliency segmentation method for infrared targets,” 2013 IEEE International Conference on Image Processing, 2013, pp.4264-4268.

[18] Hu Y, Rajan D, Chia L T. “Robust subspace analysis for detecting visual attention regions in images,” Proceedings of the 13th annual ACM international conference on Multimedia. 2005, pp.716–724.

[19] Jiang Z, Davis L S. “Submodular Salient Region Detection,” IEEE Conference on Computer Vision and Pattern Recognition. 2013, pp.2043–2050.

[20] Mai L, Niu Y, Liu F. “Saliency Aggregation: A Data-Driven Approach,” IEEE Conference on Computer Vision and Pattern Recognition. 2013, pp.1131–1138.

[21] P. Felzenszwalb and D. Huttenlocher. “Efficient graph-based image segmentation,” International Journal of Computer Vision, 2004, 59(2), pp. 167-181.