Infrared Physics & Technology 73 (2015) 286–297

Contents lists available at ScienceDirect

Infrared Physics & Technology

journal homepage: www.elsevier.com/locate/infrared

A fusion algorithm for infrared and visible images based on saliency analysis and non-subsampled Shearlet transform

Zhang Baohua [a,b], Lu Xiaoqi [a,b,*], Pei Haiquan [b], Zhao Ying [b]

[a] School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China
[b] School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China

HIGHLIGHTS

- The infrared target is detected based on saliency analysis.
- Super-pixels-based method is used to coarsely locate the infrared target region.
- The multi-directional detect operators are used to refine the contour of the target.
- Non-subsampled Shearlet transform is used to select fusion coefficients of the background.

ARTICLE INFO

Article history:
Received 26 September 2015
Available online 26 October 2015

Keywords:
Image fusion
Saliency analysis
Non-subsampled Shearlet transform
Super-pixels
Adaptive threshold segmentation

ABSTRACT

This paper proposed a novel fusion method for the infrared and visible image based on the accurate extraction of the target region. Firstly, the super-pixels-based saliency analysis method is used to extract the salient regions of the infrared image and obtain the coarse contour of the infrared target. Then the multi-directional detection operators and the adaptive threshold algorithm are used to refine the boundary of the target region and obtain the fusion decision map. In order to capture the details of the visible image, non-subsampled Shearlet transform (NSST) is used to select the fusion coefficients of the background. Experimental results indicate that the proposed method is superior to other state of the-art methods in subjective visual and objective performance.

1. Introduction

The infrared images reflect the physical information of the thermal object by recording the intensity of thermal radiation. Because of good characteristics of unaffected by the external influences, such as the sunlight, the smog and other condition factors, the infrared image is used to search for the hidden targets and identify the camouflaged. However, it is difficult to identify the surroundings using the infrared image with insufficient background information and lack of details [1]. The visible images record the reflective properties of the spectrum information of the scene, which are consistent with the human visual characteristics. But it is greatly influenced by the lighting which makes the target region in the visible image is always inconspicuous. To get a more accurate, reliable and comprehensive description of the scene information, the complementary of the infrared image and the visible image could be used to get a fused image [2,3], which includes the infrared target and the details of the scenes.

The fusion methods are usually classified as the pixel-based and the region-based methods according to different processing objects. The former selects the fusion coefficients based on single pixel [4], which is unable to describe local information of the infrared target. On the other hand, the pixel-based fusion methods change the gray values of the source images, and make it easy to introduce the side effects, such as the false contour and the glitches [5]. Relatively, the region-based methods depend on the similarity information (such as the edge intensity, the texture and the spatial frequency of the source images) to instruct the fusion [6], which is useful to protect the local features and reduce the side effects. Considering the characteristics of the infrared image, the region-based fusion method is more suitable for the infrared and visible images than the pixel-based method.

The block segmentation based fusion method [7] is a typical region-based fusion method, in which the target regions are selected from the blocks according to certain criteria. Due to the fixed size of the block, it is easy to produce 'blocking effects'. So

* Corresponding author at: School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China.

http://dx.doi.org/10.1016/j.infrared.2015.10.004
1350-4495/

the improved methods [8,9] were proposed to improve the fusion effects by dynamically computing the optimal block size according to the contents of the image. However, the blocking effect was still inevitable and high computational complexity was unsatisfactory. To avoid the 'blocking effect', the visual saliency detection was applied to substitute the segmentation algorithm in multi-focus fusion [10–12], which prevented to produce the side effect and reduced the complexity of the algorithm. However, since the saliency analysis cannot capture the weak edges of the target region, the saliency analysis was limit to extract the target with clear contour [13]. The multi-scale top-hat transform was used to extract the interest regions of the source images [14] and combine the extracted regions of the infrared and visible images to get the fused image. Zhang et al. [12] extracted the infrared target in the infrared image by using the region growing method and replacing the corresponding position in the fused image. Peng et al. [15] used the saliency analysis to mining important information of the source image, and fused multiband images based on the description of the important region of the scene. In [16], the morphological–spectral unsupervised segmentation was employed to divide the source images, and then the different fusion rules were built on sub-images. The methods mentioned above had better infrared target representation, but the detail information of the visible image had been ignored, which lost some useful detail information.

To overcome the above problems, this paper proposed a novel image fusion method based on the saliency analysis. The coarse target region of the infrared image was extracted based the super-pixels-based saliency analysis firstly. Then, the multi-directional edge detect operator was used to refine the boundaries of the target region. Subsequently, the adaptive threshold algorithm was used to get the fusion decision map and the Shearlet transform was used to select the fusion coefficients to protect the details of the background region.

The remainder of the paper is organized as follows. In Section 2, the detection method for the target region and NSST are presented. Section 3 describes the novel fusion algorithm in details. In Section 4, the fusion scheme is tested on several groups of images and the experimental results and discussions are presented. Finally, Section 5 concludes the paper.

2. Related work

2.1. Super-pixels-based saliency analysis

Super-pixels segmentation is an over-segmentation representation based on the image features [17]. Compare with the block-based segmentation method, it simplifies the image characterization and protects the edges and contours of the target region well. On the other hand, the super-pixels based saliency analysis is in line with the human visual mechanism [18], which is more effective in highlighting salient objects uniformly and robust to the background noise. So, the super-pixels segmentation is suitable for the extraction of the infrared target because it is able to capture the local features of the image without destroying the edge information of the target region.

The saliency analysis methods can be classified as the local features-based and the global features-based methods. Itti's algorithm [19] and GBVS algorithm [20] are the representative methods of local feature-based methods, the potentially important information of the image is extracted based on the low-level visual features, such as color, brightness, and direction information. Relatively, SR algorithm [21] and IG algorithm [22] consider the integrity of the image and extract the salient region in frequency domain. Although the methods mentioned above are able to extract the salient region by simulating the visual attention mechanism, they are unable to preserve the edge information of the image effectively and identify the boundary of the target precisely. To solve this problem, Li [18] proposed an improved super-pixels-based saliency analytical method. The simple linear iterative clustering (SLIC) algorithm [23] was used to coarsely segment the image by using the reconstruction error information to detect the boundary and details information. In Li's method, the edge information was well preserved and used to extract the infrared target. Super-pixels-based saliency analysis can be divided into two parts:

(1) Coarse segmentation based on the SLIC method

Super-pixels segmentation can be attributed to the clustering of the pixels with local similarity. To better capture the structural information, the SLIC is used to segment the source image into multiple, uniform and compact regions. The SLIC can be described as follows:

a. Initialization

Divide the source image into k square grids, the length of the square is $S = \sqrt{N/k}$;

b. Calculate the distance from each pixel to the cluster center

Denote $C_k = [l_k, a_k, b_k, x_k, y_k]^T$ is the cluster center, which has the minimum gradient value in the 3×3 neighborhoods. At the beginning all pixels are denoted by the distance $dis(P) = \infty$, in which P denote any pixel. The distance from the pixel P to the cluster center C_k in $2S \times 2S$ range is calculated and denoted as D_P.

$$D_P = \sqrt{\left(d_c^2 + \left(\frac{d_s}{S}\right)^2\right)m^2} \tag{1}$$

where d_c represents the color distance between two pixels in the CIELAB color model; d_s represents their distance in space; m is the weight, the smaller the value of m, the smaller impact of spatial distance on D_P, and the segmentation results are closer to the actual boundary.

$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \tag{2}$$

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \tag{3}$$

l_* is the category sign, a_* and b_* are color information respectively, x_* and y_* are location information. If $D_P < dis(i)$, let $dis(i) = D_P$, $l_i = k$, otherwise, retain their original values.

(2) Reconstruction error

The entire image is represented as $Q = [q_1, q_2, \ldots q_N] \in \mathbb{R}^{D \times N}$, where N is the number of segments and D is the feature dimension. The dense reconstruction error of segment i $den(\varepsilon_i)$ is defined as follows:

$$den(\varepsilon_i) = \|q_i - (U_{BT}\beta_i + \overline{q})\|_2^2 \tag{4}$$

where BT is the background templates, $BT = [bt_1, bt_2, \ldots bt_N]$, $BT \in \mathbb{R}^{D \times N}$, $\overline{q}$ is the mean value of Q, U_{BT} is the normalized covariance matrix $U_{BT} = [u_1, u_2, \ldots u_D]$, β_i is the reconstruction coefficient of segment i, $i \in [1, N]$, U'_{BT} is the transpose of U_{BT}:

$$\beta_i = U'_{BT}(q_i - \overline{q}) \tag{5}$$

The sparse reconstruction error of segment i is defined as follows:

$$spa(\varepsilon_i) = \|q_i - BT\alpha_i\|_2^2 \tag{6}$$

$$\alpha_i = \arg\min \|q_i - BT\alpha_i\|_2^2 + \omega\|\alpha_i\|_1 \tag{7}$$