Infrared Physics & Technology 67 (2014) 397–407

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

The infrared and visible image fusion algorithm based on target separation and sparse representation

Lu Xiaoqi [a,b], Zhang Baohua [a,b,*], Zhao Ying [b], Liu He [b], Pei Haiquan [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 firing times of PCNN.
- DENCLUE is used to accurately locate the infrared target region.
- The noise of the background region is suppressed based on sparse representation.
- Different fusion rules are built according to the target and background region.

ARTICLE INFO

Article history:
Received 11 June 2014
Available online 16 September 2014

Keywords:
Infrared and visible image
Sparse representation
Image fusion
Kernel Singular Value Decomposition
DENCLUE

ABSTRACT

Although the fused image of the infrared and visible image takes advantage of their complementary, the artifact of infrared targets and vague edges seriously interfere the fusion effect. To solve these problems, a fusion method based on infrared target extraction and sparse representation is proposed. Firstly, the infrared target is detected and separated from the background rely on the regional statistical properties. Secondly, DENCLUE (the kernel density estimation clustering method) is used to classify the source images into the target region and the background region, and the infrared target region is accurately located in the infrared image. Then the background regions of the source images are trained by Kernel Singular Value Decomposition (KSVD) dictionary to get their sparse representation, the details information is retained and the background noise is suppressed. Finally, fusion rules are built to select the fusion coefficients of two regions and coefficients are reconstructed to get the fused image. The fused image based on the proposed method not only contains a clear outline of the infrared target, but also has rich detail information.

1. Introduction

Image fusion is an important application of multi-sensor information fusion. Based on the fused images which contain more reliable, clear and accurate description of the scene, the researchers could perform further analysis, understanding image information, tracking target detection and identification. Infrared and visible image fusion is an important branch of image fusion. The infrared target scenes are obtained by thermal radiation imaging through the infrared sensors with only contour information. In contrast, the visible light sensors capture information by the reflection of target scenes with the edge of the scene and texture feature. The fusion of the infrared and visible image takes full advantage of their complementary which could be beneficial for improving detection capabilities and the ability to work around the clock of the detection system.

The main purpose of merging infrared and visible light images is to detect and locate infrared targets, so the infrared target extraction plays an important role in the fusion method. Due to the gray values of the target region are higher than surroundings' significantly, the target region could be distinguished by comparison of the pixel values. However, the distinguished boundaries may be vague, so the key to improve fusion results is the precise positioning of the infrared target.

DENCLUE is a generalized clustering algorithm based on the kernel density estimation [1–3], which can define clusters of arbitrary shapes formally. The effects on the surrounding points caused by the sample points is described by kernel function and superimposed to form a curved surface model on which the local maximum point is selected as a cluster attractor. If the density function value of the attractor is equal or greater than the preset threshold, the points matching to the condition will be clustered together.

* Corresponding author at: Aerding Street 7, Kundulun District, Baotou 014010, China.

E-mail address: zbh_wj2004@imust.cn (B. Zhang).

http://dx.doi.org/10.1016/j.infrared.2014.09.007
1350-4495/

Otherwise, if little data is attracted, attracted points are deemed as noise data [4]. The infrared target region can be accurately described by DENCLUE, and the distribution of gray values can be distinguished accurately.

Currently, many algorithms, such as adaptive thresholding method [5], non-parametric model [6], morphology method [7] and spatial analysis method [8,9], have got good results in infrared target detecting, but there are some limitations. For example, all of them rely on the priori knowledge of the target to construct a statistical model in order to simplify the problem, and the boundary of the target region is not precisely located; on the other hand, the corresponding spatial information of infrared and visible image express differently. Infrared images have little texture, poor contrast, low signal to noise and insensitive to lighting conditions. Multiscale analysis is widely used in this field; top-hat selection transform is used to extract the regions of the original infrared and visual images at each scale [10]. Shear wave is combined with Bayesian in the literature [11] to suppress the background noise of the infrared target, the separation of small infrared targets and background clutter is achieved. In the literature [12].

Factional order integral theory is used to suppress background noise of the infrared image, target and background regions are separated based on characteristics of the fractional order integral Statistical properties of the target region are considered in the methods mentioned above, but the sparsity characteristics of the human visual recognition mechanism that the image features can be accurately represented by few coefficients [13] is ignored. Yang proposed a multi-focus image fusion algorithm based on sparse representation. In his method, the image is decomposed by the redundant DCT dictionary, and the fusion coefficient is selected by comparing the 11 norm [14]. Jin proposed a coefficients optimizing method with teaching phase and learning phase to adjust the weighted coefficients automatically [15]. Aharon presented an image fusion method based on the KSVD algorithm in which sparse representation of the original image is obtained by the redundant dictionary which is trained by the KSVD algorithm, as a result, the method has strong anti-jamming ability, but it cannot extract infrared targets accurately, so the fusion result is not satisfied.

In order to accurately locate the edge of the infrared target and suppress the noise, this paper presents a new algorithm based on sparse representation and kernel density estimation clustering, firstly, the infrared target is detected by PCNN. Secondly, DENCLUE is used to find the clustering centers and accurately located the infrared target region, and then source images are classified into the target region and the background region. The background region is trained by K-SVD dictionary to get its sparse representation, so that more detail information can be retained and background noise can be suppressed. Finally, the fusion coefficients of the two regions are selected based on different rules and coefficients are reconstructed to get the fused image. The fused image based on the proposed method not only contains a clear outline of the infrared target, but also suppresses the noise.

The outline of this paper is as follows: in Section 2, we briefly review the Kernel density estimation clustering. Dictionary learning based block sparse representation is proposed in Section 3. The comparative experimental results are presented to show the performance of the proposed methods in Section 4. Finally, conclusions are drawn in Section 5.

2. Kernel density estimation clustering

Kernel density estimation (KDE) is a non-parametric density estimation method which does not depend on the priori knowledge or any hypotheses and a feature distribution estimation method driven by data. KDE could extract object effectively according to each pixel's kernel density component and different target component which can be distinguished by probability information of the gray value appeared in the region [16]. To reduce the volatility clustering results and improve the operational efficiency of DENCLUE, the image is divided into several regions based on PCNN in this paper, the number of sample points in each region is reduced and the content of the split sub-image is simplified.

In summary, the regional clustering method based on kernel density estimation includes three steps: (1) Pulse Coupled Neural Network (PCNN) is used to extract the infrared target region from the source image, and the image is divided into multiple regions according to the firing times of PCNN in this step; (2) DENCLUE is used to find the regional density attractor, namely, the local maxima of the density function; and (3) the sample points in different regions are clustered based on the density attractor, and in the region contains infrared target, precise positioning of the infrared target boundary is achieved.

2.1. Infrared target region extraction

PCNN has good pulse propagation characteristics which have been widely applied in image processing, pattern recognition and other fields. In PCNN, the gray values of the pixel are not necessarily related to neurons firing. During the course of the pulse generating, firing neurons motivate neighboring neurons firing by connecting function $L(n)$ during the course of the pulse generating, and the firing of neighboring neurons will motivate the surrounding neurons fire, thus a pulse wave is generated in the active region and propagation outside. The neighboring pixels with similar gradation tend to synchronize firing; the synchronous nature is used to achieve image segmentation. A lot of information contains in the activation process of neurons rather than in firing binary matrix [17].

PCNN is composed of the receiver domain, modulation domain and pulse generated domain:

The receiver domain:

$$F_{ij}(n) = S_{ij}[n]; \tag{1}$$

$$L_{ij}[n] = e^{-\alpha_L} L_{ij}[n-1] + V_L \sum_{kl} w_{ijkl}[n-1]; \tag{2}$$

The modulation domain:

$$U_{ij}[n] = F_{ij}[n](1 + \beta L_{ij}[n]); \tag{3}$$

The pulse generated domain:

$$Y_{ij}[n] = \begin{cases} 1, & U_{ij}[n] > T_{ij}[n] \\ 0, & \text{else} \end{cases}; \tag{4}$$

$$T_{ij}[n] = \exp(-\alpha_T) T_{ij}[n-1] + V_T Y_{ij}[n]; \tag{5}$$

where $F_{ij}(n)$ is the feedback input of the ij neuron; $S_{ij}[n]$ is the input for external incentives; T_{ij} is the neuron dynamic threshold; α_T is the time constant; V_T is the normalization constant; w_{ijkl}, are the synaptic connections right; U_{ij} is the internal activity term; β is the weight coefficients of $F_{ij}(n)$; σ is the average factor which regulates the internal activity level; Y_{ij} is the ij neuron output; n is the iteration number.

Considering the pulse propagation characteristics of PCNN, pixels with same firing times have similar regional characteristics which can be classified according to the number of firing times. An example of the regional classification is shown as Fig. 1, Fig. 1(a) is the infrared image, Fig. 1(b) is the firing map, Fig. 1(c)–(e) represent the maps that firing 1–3 times respectively. Pixels with the same firing times represent the contour information of different regions. In Fig. 1(e), the outline of the infrared objectives is clear, but in Fig. 1(c), it mainly contains background information distributed in pixels with lower firing times. The separation of the target region and the background