Infrared Physics and Technology 92 (2018) 68–77

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

Regular article

General fusion method for infrared and visual images via latent low-rank representation and local non-subsampled shearlet transform

Boyang Cheng[a,b,*], Longxu Jin[a], Guoning Li[a]

[a] Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin 130033, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Keywords:
Latent low-rank representation
Local non-subsampled shearlet transform
Salient features
Image fusion

ABSTRACT

This study establishes the general fusion method for infrared and visual images via latent low-rank representation (LatLRR) and local non-sampled shearlet transform (LNSST) to effectively combine the salient information of both images and solve problems on low-contrasting heterogeneous image fusion. First, LNSST is used as a multi-scale analysis tool to decompose the source images into low-pass and high-pass sub-images. Second, the LatLRR, which is an effective method for exploring multiple subspace structural data, is used to extract the salient information of image sources. Thus, the LatLRR can be adopted to guide the adaptive weighted fusion of low-pass sub-images. Simultaneously, the average gradient, which can reflect image edge details, is regarded as the fusion rule for high-pass sub-images. A series of images from diverse scenes are used for the fusion experiments, and the results are evaluated subjectively and objectively. The subjective and objective evaluations show that our algorithm exhibited superior visual performance, and the values of the objective evaluation parameters increase by about 5–10% compared with other typical fusion methods.

1. Introduction

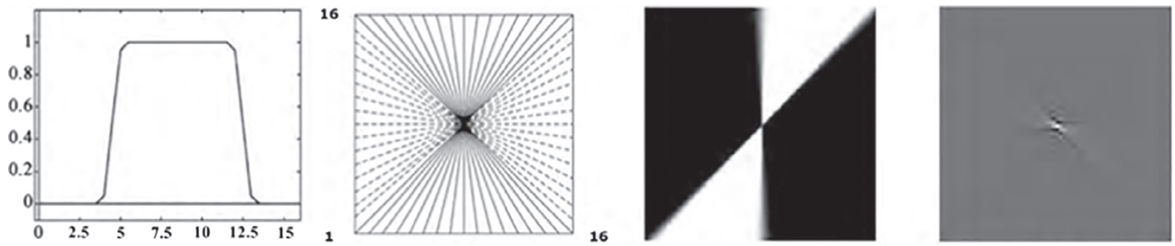
The development of infrared (IR) and visible image fusion technology is largely aimed at developing modern military detection technology. A visible image (VI) is a reflection image with several high-frequency components, and VI images can reflect scene details under certain illumination conditions. However, when illumination is not good, the resultant contrast of the VI image is relatively low. Meanwhile, an IR image is a radiation image. The gray level of IR images is determined by the temperature difference between the target and background, but resultant images cannot reflect real scenes [1]. Image fusion technology for IR and VI images can effectively synthesize and explore the combined characteristic information of two complementary images with the same resolution, enhance the understanding of a scene, and highlight image targets; thus, image fusion technology can find objects quickly and accurately despite confusing situations [2].

Several fusion approaches have been recently proposed, especially for pixel-level-based VI and IR image fusion [3]. A number of multi-scale analysis tools, such as contourlet transform (CT) [4], non-subsampled contourlet transform (NSCT) [5], and local non-subsampled shearlet transform (LNSST) [6], have been successfully used in the field of image fusion. LNSST is regarded the fastest MGA tool with the most disaggregation effect. Furthermore, LNSST can exhibit good local characteristics in space and frequency domains, avoid blocking effects, weaken the Gibbs-ringing phenomenon by using local small-sized shearlet filters, and improve the calculation efficiency of time domain convolutions. Thus, many researchers favor LNSST over other techniques. Lei et al. [7] proposed an adaptive fusion method based on the LNSST and non-negative matrix factorization to construct an algorithm that could guide the fusion of low-frequency coefficients, but the final fused images were dim and lost considerable textural details. Zhang et al. [8] presented a fusion algorithm based on saliency analysis and LNSST. This method utilized saliency detection to integrate IR target information into the VI image, but the fusion effect of background information required improvements. Wu et al. [9] combined LNSST and deep Boltzmann machine programming to solve fusion problems, but deep learning technology is not yet mature for fusion applications. Wang et al. [10] proposed a fusion algorithm for RDU-PCNN and ICA bases in the LNSST domain. Although the PCNN has a bionic mechanism, the final fused images introduce artifacts and have hazy image edges. Kong et al. [11] forwarded a technique for gray-scale VI and IR image fusion based on the LNSST. This method makes use of regional averaged energy and local directional contrast, but the fused image loses some important IR saliency information.

Latent low-rank representation (LatLRR) [12], an upgraded version

* Corresponding author at: Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin 130033, China.
E-mail address: boyangwudi@163.com (B. Cheng).

https://doi.org/10.1016/j.infrared.2018.05.006
Received 23 December 2017; Received in revised form 23 April 2018; Accepted 7 May 2018



(a) Meyer wavelet window **(b)** Pseudo-polarized coordinates **(c)** Frequency domain shearlet filter **(d)** Time domain shearlet filter

Fig. 1. Shearlet filter formation procedure when $L = 15$.

of the low-rank representation (LRR), is an effective method for exploring the multiple subspaces of data structures. LatLRR can robustly extract salient features from images because the method utilizes an unsupervised feature extraction algorithm. Moreover, LatLRR is robust to noise. When an image matrix is decomposed by LatLRR, the image can be represented as a superposition of principal features, salient features, and sparse noise. Salient features show the spatial distribution of salient information of images, and the weighted-mean is usually treated as the fusion rule for exploring the low contrast and unnatural reconstruction of heterogeneous fused images. The LatLRR algorithm can precisely identify salient objects and regions in images to form a saliency map. The saliency map contains the weight information of the spatial distribution of a grayscale image and thus integrates a weighting function. The fusion rule can be changed from the weighted-mean to the weighted-adaptive approach to effectively merge the salient information into the fused image and improve the fusion effect.

Based on the above review, this study proposes a general fusion method for IR and visual images via LatLRR and LNSST. To the best of our knowledge, this is the first time that the LatLRR has been used in the field of heterologous image fusion, in which the heterologous source images have the same resolution [13]. In this study, the LNSST is first used as a multi-scale analysis tool to decompose image sources into a low-pass sub-image and a series of high-pass sub-images. Second, the saliency information of the image is extracted by LatLRR to guide the adaptive weighted fusion of low-pass sub-image and high-pass sub-images. Finally, each sub-image is modeled and the corresponding fusion coefficients are produced. An algorithm is adopted to effectively express image characteristics and obtain a good fusion effect by using IR and visible light images in the fusion experiments.

The remainder of this paper is organized as follows: Section 2 introduces the theory relevant to LNSST and LatLRR. Section 3 elaborates the algorithm based on the new fusion rule. Section 4 presents five experimental results and intuitively compares the proposed method with other methods. Section 6 provides a summary of the findings.

2. Relevant theory

2.1. LNSST [14,15]

When the dimension is $n = 2$, the shearlet system function with discrete parameters is as follows:

$$S_{AB}(\varphi) = \{\varphi_{j,l,k} = |\det A|^{j/2}\varphi(B^l A^j x - k);\ j,\ l \in Z,\ k \in Z^2\}. \tag{1}$$

where $\varphi \in L^2(R^2)$, A, and B are 2×2 reversible matrices; $|\det B| = 1$; j is the scale parameter; l is the direction parameter; and k is the spatial position.

For $j \geq 0$, $-2^j \leq l \leq 2^j - 1$, $k \in Z^2$, and $d = 0,1$, the Fourier transform of the shearlet can be expressed on the basis of the tight support frame.

$$\hat{\varphi}^{(d)}_{j,l,k} = 2^{3j/2} V(2^{-2j}\xi) W^{(d)}_{j,l}(\xi) e^{-2\pi i \xi A_d^{-j} B_d^{-l} k}. \tag{2}$$

where $V(2^{-2j}\xi)$ is the scale function; $W_{j,l}{}^{(d)}$ is the window function localized on the trapezoidal pair; A_d is the heterosexual expansion matrix; and B_d is the shear matrix. The shearlet transform of the $f \in L^2(R^2)$ function can be calculated by Eq. (3).

$$\langle f, \hat{\varphi}^{(d)}_{j,l,k} \rangle = 2^{3j/2} \int_{R^2} \hat{f}(\xi) \overline{V(2^{-2j}\xi) W^{(d)}_{j,l}(\xi)} e^{-2\pi i \xi A_d^{-j} B_d^{-l} k} d\xi. \tag{3}$$

As shown in Eq. (3), the shearlet transform is divided into two steps. The first step is a multi-scale decomposition [i.e., $\hat{f}(\xi)\overline{V(2^{-2j}\xi)}$] and the second step is the direction of the localization, i.e., $\hat{f}(\xi)\overline{V(2^{-2j}\xi)W^{(d)}_{j,l}(\xi)}$.

Multi-scale decomposition: The image is subjected to non-subsampled pyramid decomposition using a non-subsampled $2D$ filter bank of dual channels to generate a low-pass sub-image and multiple high-pass sub-images with perfect reconstruction.

Directional localization: Directional localization is achieved by small-scale shearlet filters and high-pass sub-images convolution calculations. The local window is $L \times L$, where $L = n \cdot (2^{j-1} + 1)$ with j as the scale parameter and n as any positive integer. The local small-size shearlet filter can avoid the blocking effect, weaken the Gibbs-ringing phenomenon, and improve the calculation efficiency of time domain convolution. Thus, j is usually 2 or 3 and the local window is usually 15×15.

The above shearlet transformation is called the LNSST, a technique that removes the sampling operation in the decomposition stage. LNSST involves translation invariance because the local small-size shearlet filter can avoid spectrum aliasing to improve image decomposition and reconstruction. The shearlet filter formation process for $L = 15$ is shown in Fig. 1.

The image f is decomposed by the m-layer LNSST to obtain $\Sigma_m 2^{dm}$ high-pass directional sub-images and a low-pass sub-image. Each sub-image is the same size as the original image, and d_m is the number of m-layers of directional localization. A two-layer LNSST decomposition of Linda is shown in Fig. 2. The number of high-pass sub-images in the first layer is 4 (the number of stages is 2), the number of high-pass sub-images in the second layer is 4 (the number of stages is 2), and the size of the shearlet filter is 15×15.

2.2. LatLRR

The image sources of the fusion usually contain a certain amount of noise, but LatLRR can automatically extract salient features from noisy images. LatLRR is robust to noise, and the saliency map obtained is more accurate than other saliency detection-based methods [16–18].

The core idea behind LatLRR is that an image matrix can be represented as a superposition of principal features, salient features, and sparse noise given the low rank and sparse optimization criteria. For an image matrix $X \in R^{M \times N}$, the idea may be interpreted as

$$X = XL + SX + E. \tag{4}$$

where L represents the low-rank matrix, $L \in R^{N \times N}$; S represents the sparse matrix, $S \in R^{M \times M}$; E represents the sparse noise, $E \in R^{M \times N}$; XL