ISSN 1674-1056
http://cpb.iphy.ac.cn
http://iopscience.iop.org/cpb
中国物理 B
Chinese Physics
中国物理 Chinese Physics B

An infrared and visible image fusion method based uponmulti-scale and top-hat transforms

Gui-Qing He(何贵青), Qi-Qi Zhang(张琪琦), Jia-Qi Ji(纪佳琪), Dan-Dan Dong(董丹丹), Hai-Xi Zhang(张海曦), Jun Wang(王珺)

Citation: Chin. Phys. B . 2018, 27(11): 118706. **doi:** 10.1088/1674-1056/27/11/118706

Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Integration of a field-effect-transistor terahertz detector with a diagonal horn antenna

Xiang Li(李想), Jian-dong Sun(孙建东), Zhi-peng Zhang(张志鹏), V V Popov, Hua Qin(秦华)

Chin. Phys. B . 2018, 27(6): 068506. **doi:** 10.1088/1674-1056/27/6/068506

Performance of dual-band short- or mid-wavelength infrared photodetectors based on InGaAsSb bulk materials and InAs/GaSb superlattices

Yao-yao Sun(孙姚耀), Yue-xi Lv(吕粤希), Xi Han(韩玺), Chun-yan Guo(郭春妍), Zhi Jiang(蒋志), Hong-yue Hao(郝宏玥), Dong-wei Jiang(蒋洞微), Guo-wei Wang(王国伟), Ying-qiang Xu(徐应强), Zhi-chuan Niu(牛智川)

Chin. Phys. B . 2017, 26(9): 098506. **doi:** 10.1088/1674-1056/26/9/098506

Very long wavelength infrared focal plane arrays with 50% cutoff wavelength based on type-II InAs/GaSb superlattice

Xi Han(韩玺), Wei Xiang(向伟), Hong-Yue Hao(郝宏玥), Dong-Wei Jiang(蒋洞微), Yao-Yao Sun(孙姚耀), Guo-Wei Wang(王国伟), Ying-Qiang Xu(徐应强), Zhi-Chuan Niu(牛智川)

Chin. Phys. B . 2017, 26(1): 018505. **doi:** 10.1088/1674-1056/26/1/018505

High-speed waveguide-integrated Ge/Si avalanche photodetector

Hui Cong(丛慧), Chunlai Xue(薛春来), Zhi Liu(刘智), Chuanbo Li(李传波), Buwen Cheng(步成文), Qiming Wang(王启明)

Chin. Phys. B . 2016, 25(5): 058503. **doi:** 10.1088/1674-1056/25/5/058503

Open critical area model and extraction algorithm based on the net flow-axis

Wang Le, Wang Jun-Ping, Gao Yan-Hong, Xu Dan, Li Bo-Bo, Liu Shi-Gang

Chin. Phys. B . 2013, 22(12): 128504. **doi:** 10.1088/1674-1056/22/12/128504

SPECIAL TOPIC — 80th Anniversary of Northwestern Polytechnical University (NPU)

An infrared and visible image fusion method based upon multi-scale and top-hat transforms*

Gui-Qing He(何贵青)[1], Qi-Qi Zhang(张琪琦)[1], Jia-Qi Ji(纪佳琪)[1], Dan-Dan Dong(董丹丹)[1], Hai-Xi Zhang(张海曦)[1,†], and Jun Wang(王珺)[2,‡]

[1] *School of Electronics and Information, Northwestern Polytechnical University, Xi'an* 710072, *China*
[2] *School of Information Technology, Northwestern University, Xi'an* 710072, *China*

(Received 3 July 2018; revised manuscript received 2 October 2018; published online 8 October 2018)

The high-frequency components in the traditional multi-scale transform method are approximately sparse, which can represent different information of the details. But in the low-frequency component, the coefficients around the zero value are very few, so we cannot sparsely represent low-frequency image information. The low-frequency component contains the main energy of the image and depicts the profile of the image. Direct fusion of the low-frequency component will not be conducive to obtain highly accurate fusion result. Therefore, this paper presents an infrared and visible image fusion method combining the multi-scale and top-hat transforms. On one hand, the new top-hat-transform can effectively extract the salient features of the low-frequency component. On the other hand, the multi-scale transform can extract high-frequency detailed information in multiple scales and from diverse directions. The combination of the two methods is conducive to the acquisition of more characteristics and more accurate fusion results. Among them, for the low-frequency component, a new type of top-hat transform is used to extract low-frequency features, and then different fusion rules are applied to fuse the low-frequency features and low-frequency background; for high-frequency components, the product of characteristics method is used to integrate the detailed information in high-frequency. Experimental results show that the proposed algorithm can obtain more detailed information and clearer infrared target fusion results than the traditional multi-scale transform methods. Compared with the state-of-the-art fusion methods based on sparse representation, the proposed algorithm is simple and efficacious, and the time consumption is significantly reduced.

Keywords: infrared and visible image fusion, multi-scale transform, mathematical morphology, top-hat transform

PACS: 87.50.W–, 85.60.Gz, 82.50.Bc, 61.80.Ba **DOI:** 10.1088/1674-1056/27/11/118706

1. Introduction

The visible image is obtained according to the reflection characteristics of ground features, and can provide rich information of scenes and of spatial details but cannot reveal hidden people or items. The infrared image is captured via the thermal radiation characteristics of an object, and has a high capability of target detection but low spatial resolution. The fusion of infrared and visible images can synthesize better target recognition of infrared images and higher spatial resolution and rich scene information of visible images, thus improving the ability of correctly detecting and identifying targets in complex environments. Such a fusion method has broad applications in military, transportation, security, and many others.[1–4]

Since multi-scale transforms (MSTs in short) have features of multiple resolutions and time-frequency localization, they have always been the hot topics in the field of image fusion for infrared and visible images with strong complementarity.[5,6] The most widely used multi-scale transforms include Laplace pyramid,[7] discrete wavelet transform (DWT),[8–11] dual-tree complex wavelet transform (DTCWT),[12–14] contourlet transform (CT),[15] and non-subsampled contourlet transform (NSCT).[16–18] The fusion process can usually be divided into three steps. Firstly, the source image is multi-scale transformed so as to acquire its low- and high-frequency components. Secondly, fusion coefficients are procured for low- and high-frequency components respectively via specific fusion rules. Finally, fusion coefficients are inversely multi-scale transformed in order to attain the fusion result. Such a multi-scale-transform based fusion method can efficaciously extract source image features in multiple scales and from diverse directions. However, for such a method, fusion rules are mostly designed for high-frequency components. Due to the incompleteness of multi-scale decomposition, certain details are still retained in the low-frequency component, which is even more pronounced when the number of decomposition layers is small. Hence, current fusion rules have not fully taken into account such features as edge con-

*Project supported by the National Natural Science Foundation of China (Grant No. 61402368), Aerospace Support Fund, China (Grant No. 2017-HT-XGD), and Aerospace Science and Technology Innovation Foundation, China (Grant No. 2017 ZD 53047).

†Corresponding author. E-mail: zh.haixi@gmail.com
‡Corresponding author. E-mail: jwang@nwu.edu.cn

© 2018 Chinese Physical Society and IOP Publishing Ltd

tours in the low-frequency component, incurring information loss: the energy and contours contained in the low-frequency component cannot be accurately transmitted to the fusion result, affecting the quality. Therefore, the study herein has focus on the design of fusion rules for the low-frequency component, making up for the limitation of features extraction from multi-scale transforms, and obtaining better fusion results.

In recent years, the top-hat transform in mathematical morphology has been successfully applied to infrared and visible image fusion due to the following reasons. As a toolbox in mathematical morphology, the top-hat transform can simultaneously extract from source images (the infrared and visible images to be fused) the bright features (local maximum values of grayscale) and darkness features (local minimum values of grayscale), and separate bright-&-dark features from background information, which is conducive to obtaining complementary information of source images during the fusion process. With top-hat transform, Bai and co-workers[19] have changed the scale of structural elements, extracted the bright-&-dark features of multiple levels of source images, and obtained rich fusion results. Chen and co-workers[20] have further improved the top-hat transform in Ref. [19]: the bright-&-dark features are fused separately from the background information, thus enhancing the fusion result for infrared and visible images. Such a top-hat transform-based fusion method does the fusion work in the image domain at a single scale. Considering the fact that images usually contain different features at diverse scales and in disparate directions and that such features are often prominent information which the image fusion needs to distinguish and to retain, it is proposed in the study that by combining multi-scale and top-hat transforms and thus taking advantage of both, a better fusion result may be achieved. In particular this means: on one hand for high-frequency components, the multi-scale transform is used to extract detailed information in multiple scales and from diverse directions, and on the other hand for the low-frequency component, the toolbox of top-hat transform in mathematical morphology is utilized to separate the bright-&-dark features from background information, whereby reaping a fusion image with richness and accuracy.

In the following sections, we first introduce the theoretical basis of multi-scale and top-hat transforms in Section 2. The proposed method is described in Section 3. In Section 4, the experimental results, analysis, and discussion are given, and followed by a conclusion in Section 5.

2. Theoretical basis of multi-scale and top-hat transforms

2.1. Multi-scale transforms

Currently in the field of infrared and visible image fusion, traditional multiscale transforms are capable of extracting source image features in multiple scales and from diverse directions, which is beneficial to obtain rich fusion results, thus becoming an active research direction. The most typical and widely used multi-scale transforms are DWT, DTCWT, and NSCT.

DWT is the most representative one in wavelet transforms and enjoys many edges such as directionality and localization. For the source image, a DWT can be performed to obtain the low-frequency component denoted as LL and three high-frequency components denoted as LH, HL, and HH, respectively for details of horizontal, vertical, and diagonal directions. DTCWT overcomes drawbacks of traditional wavelet transform: less directional selectivity and without translation invariance, by adopting the two-way DWT structure in a binary tree: one tree generating the real part of the DTCWT while another tree begeting its imaginary part, thus obtaining the approximate translation invariance, directional selectivity, and computational efficiency. NSCT is a true two-dimensional image transform and can acquire the essential geometric structure of the source image, and has no need for the downsampling and upsampling processes in the decomposition and reconstruction of the image while being translation invariance. NSCT consists of non-subsampled pyramid filter banks (NSPFB in short) and non-subsampled directional filter banks (NSDFB in short). The size of submap in each direction is equal to that of the source image.

From the above analysis, it is known that multi-scale transforms with better performance have been proposed, and the capability of feature extraction has been continuously enhanced, thus effectively advancing the image fusion technology. However, most existing multi-scale-based fusion methods focus on the research of high-frequency fusion rules, which is in contrast with the fact that for the low-frequency component the fusion rule is rather standard, single, and unique. Since the low-frequency component still contains much edge and contour information, and such important information in low-frequency component cannot be retained, and the fusion result is affected. Due to such consideration, the study herein aims at addressing this issue and focuses on fusion rules for multi-scale transform of low-frequency component in order to improve its accuracy and to enhance fusion effect.

2.2. Top-hat transform

Top-hat transform is an important tool in mathematical morphology, and in recent years has been successfully applied to the fusion of infrared and visible images, and achieved good fusion effects. The classical top-hat transform can only extract the bright features of the image, which correspond to the local pixel peak values of the source image, and the darkness features correspond to the local pixel valley values. In order to extract at the same time bright-&-dark features of the source image, multiple subtraction operations are entailed, and due to calculation errors are prone to the problem of inaccuracy and incompleteness of image feature extraction, and are inconvenient for practical applications. Hence, one new type of top-hat transform is utilized: when the open and close operations are performed, two structural elements $B_1(u,v)$ and $B_2(u,v)$ of unequal scales are applied with the corresponding scale being $S_{B_1} > S_{B_2}$. Assume the source image being $f(x,y)$, the novel open operation can be expressed as

$$f\circ = (f\Theta B_1)\oplus B_2. \tag{1}$$

Such an operation can eliminate both bright-&-dark features at the same time. The new top-hat transform is composed of the source image and subtraction of the new open operation, therefore, the bright-&-dark features of the image can be obtained simultaneously. Furthermore, the performance of the bright-&-dark features extracted by the new top-hat transform is consistent with the source image, viz. the bright features correspond to the local pixel peak values and the dark ones to valley numbers. Thus, the study herein adopts the new type of top-hat transform to extract image feature, which can simultaneously obtain not only bright-&-dark features but also avoid multiple subtraction operations, whereby improving the accuracy of feature extraction. Although one of the structural elements has a large scale, its computational complexity and load are comparatively and significantly reduced relative to the traditional top-hat transform. Definition of the new top-hat transform is as follows:

$$TH = f - f\circ. \tag{2}$$

In Fig. 1, characteristic images are shown after the classical (b) and the new top-hat transforms are applied, where for the former the structural element B is a square structure with scale of 3 and for the latter B_1 of 6 and B_2 being one with scale of 3. It is evident from the pictures that with equal conditions the new top-hat transform is relatively more efficacious to extract image features and more conducive to obtain richer bright-&-dark features. With such a consideration the study herein combines the new top-hat transform with traditional multi-scale transform, viz. for the source image, (i) a multi-scale transform is applied to secure the low-frequency component and several high-frequency components; (ii) consequently the new top-hat transform is utilized to extract the bright-&-dark features in the low-frequency component; and (iii) features acquired in (ii) and the background information are processed via different fusion rules, rendering for the low-frequency component in the fusion image having richer features and evenly distributed energy, thus attaining better fusion effect.



Fig. 1. Characteristic images from the classical and the new top-hat transforms. Corresponding images are shown. (a) Source image, (b) characteristic image from classical top-hat transform, (c) characteristic image from the new top-hat transform.

3. Fusion rules

For the infrared and visible image fusion, the source image is first multi-scale transformed to extract components of low-high-frequencies, and such information represents disparate image features: the components of high-frequency correspond to features with sharp variations of grayscale in the source image such as edges, textures, and corners. Coefficients of high-frequency components are thus far approximately sparse (with relatively fewer non-zero elements) and can comparatively better depict characteristic features. Therefore, taking into account the characteristic features of single pixels and local regions, the product of characteristics method is adopted as the fusion rule. The low-frequency component corresponds to features with slow variations of grayscale in the source image such as contours and energy. Subject to the scale and to the selective direction in the multi-scale transform, the low-frequency component usually contains many important features of the source image, and if dealt without any discretion, the fusion result will be affected adversely. Therefore, the study herein has utilized the new top-hat transform to first isolate the bright-&-dark features in the low-frequency component from its background, and afterwards nonidentical fusion strategies are employed to fuse separately with anticipation to rich features and to accuracy in the low-frequency component of the resultant fusion image. As stated above, the current study combines the classical multi-scale transform with the new top-hat one, and takes advantages of both in order to obtain superior fusion results. Figure 2 shows the framework of image fusion in the study, where I_1 and I_2 represent the source images: the infrared and visible ones, respectively, and both have been registered.

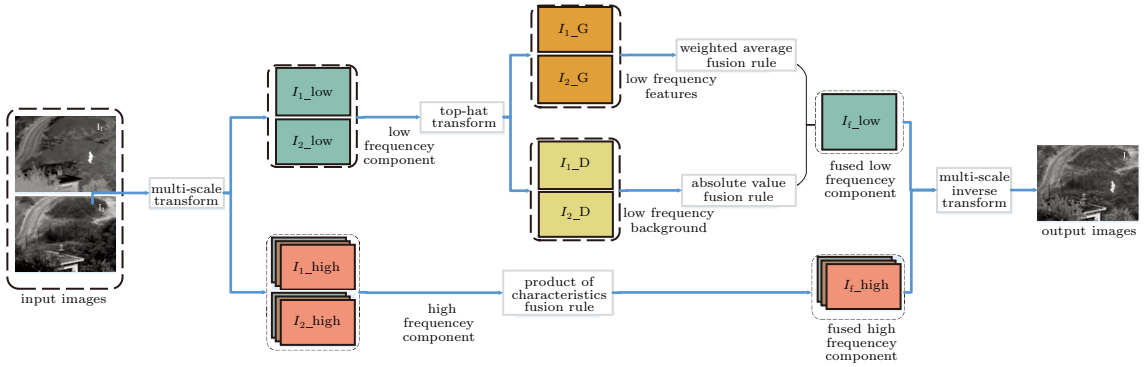


Fig. 2. (color online) Framework of image fusion in the study.

3.1. Fusion rules for the low-frequency component

Due to the incompleteness of multi-scale decomposition, certain information on details still retains in the low-frequency component, which is more pronounced for the case with relatively fewer decomposition levels. Traditional multi-scale transforms usually use weighted average to fuse the low-frequency component, neglecting its detailed features, thus losing the information on details of its source image and decreasing the contrast in the resultant fusion image. Hence, for the low-frequency component, this study adopts a fusion rule which is based upon the new top-hat transform: processing the bright-&-dark features separately from its background by utilizing the complementary and redundant information in the source image, thus enriching the fusion effect with more details and energy being more evenly distributed.

Firstly, perform the new open operation on the low-frequency component Ii_{low} $(i = 1,2)$, where B_1 and B_2 are square flat structural elements, and the scale of B_1 is greater than that of B_2. Obtain the background Ii_{G} $(i = 1,2)$ for the low-frequency:

$$Ii_{\text{G}} = (Ii_{\text{low}} \Theta B_1) \oplus B_2. \tag{3}$$

Next, apply the new top-hat transform, viz. the low-frequency component Ii_{low} $(i = 1,2)$ being subtracted by its background Ii_{G} $(i = 1,2)$ to obtain the bright-&-dark features Ii_{D} $(i = 1,2)$:

$$Ii_{\text{D}} = Ii_{\text{low}} - (Ii_{\text{low}} \Theta B_1) \oplus B_2 = Ii_{\text{low}} - Ii_{\text{G}}. \tag{4}$$

Figure 3 is the result image of the new top-hat transform for UNcamp image, where (b) and (c) are the image background and the bright-&-dark features, respectively.

The bright-&-dark features and background of the low-frequency component are first divided into small blocks by a sliding window strategy and then fused. Because region-based fusion rules are better than single-pixel fusion rules, the block size is $n \times n$, the sliding step size is step, obtain the background Ii_{Gblocks} $(i = 1,2)$ and the bright-&-dark features Ii_{Dblocks} $(i = 1,2)$: with both being to be fused.



(a) (b) (c)

Fig. 3. New top-hat transformed UNcamp image: bright-&-dark features and background. Corresponding images are shown. (a) UNcamp image, (b) background, (c) bright-&-dark features.

The bright-&-dark features Ii_{Dblocks} $(i = 1,2)$ in the low-frequency component are relatively sparse coefficients. The rule of taking the absolute value with the largest magnitude can retain more details in the low-frequency component, rendering richer detailed information in the fusion result. If_{Dblocks} $(i = 1,2)$ are the bright-&-dark features in the fused image, and the fusion rule is given as

$$If_{\text{Dblocks}} = \max(I1_{\text{Dblocks}}, I2_{\text{Dblocks}}). \tag{5}$$

The background of the low-frequency component Ii_{Gblocks} $(i = 1,2)$ needs to provide reasonable background information for the fused result. Therefore, guided by the bright-&-dark features in the low-frequency component, a weighted-average fusion rule is applied, which can maximally retain such features as contours in the low frequency and is conducive to actualize relatively sharper contrast in the fusion result. If_{Gblocks} represents the fused background and the weighted average is shown as

$$If_{\text{Gblocks}} = \sum_{i=1}^{2} wi \times Ii_{\text{Gblocks}} \Big/ \sum_{i=1}^{2} wi,$$
$$wi = \|Ii_{\text{Dblocks}}\|_2. \tag{6}$$

Finally, add the fused bright-&-dark features If_{Gblocks} with the fused background If_{Dblocks}, perform the reverse sliding window operation and put back sequentially, and take the average value per the number of superimposition of each pixel to land the fused low-frequency component If_{low}.

3.2. Fusion rules for high-frequency components

For high-frequency components, traditional fusion rules are all based upon one single characteristic, which is accomplished by calculating the local variance, local gradient, energy of the wavelet coefficients, etc. The biggest disadvantage of such rule is the incompleteness of feature analysis and the insufficiency of extracted image details considering only the individual information of a single pixel while neglecting the overall information of the local region, and vice versa. The study in Ref. [24] has applied the idea of multi-resolution in wavelet analysis to coefficient fusion, adopted the product of multiple characteristics as the fusion rule, and obtained the fusion effect which is more natural and more comprehensive. The study herein has used the same two characteristics in Ref. [24] as the decision basis for coefficient fusion. In the window l with size $N \times N$, define the product of characteristics $p_{l,k}^{d}$,

$$p_{l,k}^{d} = |\sigma_l \times \delta_{l,k}|, \tag{7}$$

where N takes the value of 3, d denotes the serial number of different high-frequency subband, and σ_l is the standard deviation of all pixels in window l, and $\delta_{l,k}$ represents the average pixel gradient of the k-th pixel in window l. Based upon $p_{l,k}^{d}$, the fusion rule for coefficients is given as

$$\text{coef}_{l,k}^{d} = \begin{cases} \text{coef}1_{l,k}^{d}, & p1_{l,k}^{d} \geq p2_{l,k}^{d}, \\ \text{coef}2_{l,k}^{d}, & \text{otherwise}, \end{cases} \tag{8}$$

where $\text{coef}1_{l,k}^{d}$ and $\text{coef}2_{l,k}^{d}$ represent coefficients of high-frequency components $I1_{\text{high}}$ and $I2_{\text{high}}$ in the d-th high-frequency subband, respectively. $\text{coef}1_{l,k}^{d}$ is the coefficient of the fused high-frequency If_{high} in the d-th high-frequency subband. $p1_{l,k}^{d}$ and $p2_{l,k}^{d}$ represent for the k-th pixel the products of characteristics to the high-frequency components $I1_{\text{high}}$ and $I2_{\text{high}}$ of window l in the d-th high-frequency subband, respectively. If_{high} denotes the obtained high-frequency component.

The above describes fusion rules for the high and low frequency components specifically. On one hand, the new top-hat transform has been applied to low frequency components and thus separate the bright-&-dark features of the low-frequency component from its background, and consequently different fusion rules are selected with the help of different feature attributes. On the other hand, for high-frequency component, the fusion rule is based upon the feature product in order to extract the accuracy information from the high frequency component. For infrared and visible image fusion, such a fusion strategy has enabled the fusion effect with objects being clear and features being rich.

4. Experimental results and analysis

The experimental data are three sets of infrared and visible images which have been accurately registered: the UNcamp group, the duck group, and the street group, which are shown in Fig. 4, where the top row figures are for the infrared and the bottom row figures are for the visible. The characteristics of the above three groups of source images are as follows: the first group is a single infrared target with a simple background; the second group is a single infrared target with a rich background; and the third group is a multi-target with a rich background. In order to show the superiority of the method in this study, results from various other methods are also presented and compared, viz. fusion methods which are based upon celebrated top-hat transforms TH1[19] and TH2,[20] classical SR[21] and JSR1,[22] the latest which is based upon salient map JSR2,[23] and three representative multi-scale transforms DWT,[11] DTCWT,[13] and NSCT.[18] The proposed method is denoted as MST+TH, in particular as DWT+TH, DTCWT+TH, and NSCT+TH. The fusion of infrared and visible images is mainly to make full use of the complementarity of the two pieces of information, and to retain as much as possible the target characteristics of the infrared and the details of the visible. So the widely-used Petrovic index[25] and Piella index[26] have been adopted as the objective evaluation indices. The Petrovic index, also called Q_{abf} index, is a gradient-based fusion index which mainly measures the gradient information passed on from the source image to the fused one. Piella index is based upon structural similarity and consists mainly of three indices Q_0, Q_{w}, and Q_{e}, where Q_0 focuses on correlation loss, brightness distortion, and contrast distortion to evaluate the distortion degree of the fused image. Q_{w} and Q_{e} both evaluate stark information passed on from the source image to the fused one while Q_{w} is on sensitive regions and Q_{e} on edges. The range of all four indices is [0,1], and closeness to the upper value one indicates better effect.

In the experimental design, all parameters for the TH1 and TH2 methods are taken exactly as the optimal ones in Refs. [19] and [20], respectively. For the SR, JSR1, and JSR2 methods, the training sample consists of the randomly-selected (from the source image) 5000 image blocks with size of 8×8. The number of iterations for the K-SVD algorithm is 20. For the OMP algorithm, the decomposition error is set to 0.1. The fusion rule for the SR method is taking the largest norm of l_1 and the weighted average for the JSR1 method. For the JSR2 method, the parameters for the low-frequency component are the same as the optimal ones in Ref. [23]. Three layers of "db1" wavelet basis have been applied for the DWT and DWT+TH methods. For the DTCWT and DTCWT+TH methods, the first stage filter is "FSfarras" and "dualfilt1" for other stage filters. For the NSCT and NSCT+TH methods, the "pyrexc" filter is adopted and the "vk" as the directional filter, and the number of decomposition is four and the directional series sequentially take the values of [1,2,3,4]. In order

to verify the role of the top-hat transform in fusing the low-frequency component, for the MST method, the simplest average fusion rule is utilized, viz. taking the mean of coefficients. For high-frequency components, the fusion rule for the MST and MST+TH methods is taking the largest of the products of characteristics.

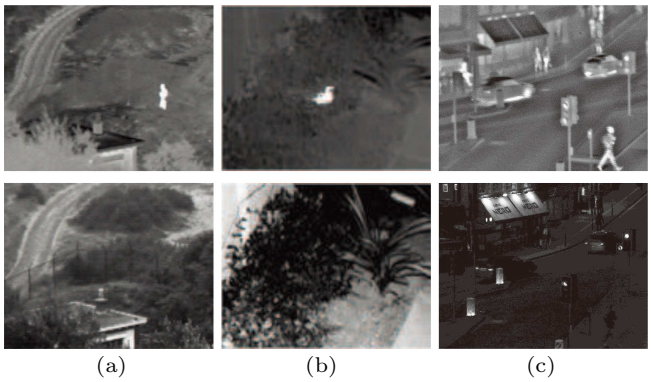


Fig. 4. Three sets of accurately registered infrared and visible images. Corresponding images are shown. (a) UNcamp, (b) duck, (c) street.

4.1. Discussion on the scale of elements

As mentioned earlier for the new-top hat transform, there exist structural elements B_1 and B_2 with scales $S_{B1} > S_{B2}$, where scale S_{B2} is for the extraction of bright features and S_{B1}–S_{B2} for dark ones. So the relative 'gravity' factor can be calculated as $M = (S_{B1} - S_{B2})/S_{B2}$. Thus, the scales of elements B_1 and B_2 have a direct impact on the relative proportion of the extracted bright-&-dark features and their finer details, thus affecting the fusion effect. In practical applications, both the bright-&-dark features shall be valued. Thus the gravity factor M is set to one in the study. The following discussion focuses on the impact of the structural scales on the fusion effect, which are given as pairs in the combination of (S_{B1}, S_{B2}). The smaller the scale, the finer the extracted bright-&-dark features. Thus, the scale S_{B2} increases from one to six with step-size of one. Consequently the scale S_{B1} is obtained from S_{B2} and from the gravity factor M, and has the minimum value of three. Therefore, the following pairs are obtained and used in the experiment: (3,1), (4,2), (6,3), (8,4), (10,5), and (12,6). Structural elements are all square flat ones. The analysis of impact of scales on the fusion effect is sequentially carried out for the DWT+TH, DTCWT+TH, and NSCT+TH methods.

Figure 5 shows the objective evaluation indices of fusion results for various studied methods with different scales. It is evident that for the DWT+TH method the evaluation indices have the largest decrease as scales increase, followed by the DTCWT+TH method, and the NSCT+TH method is immune to scale variation. Hence, for the DWT+TH and DTCWT+TH methods, smaller scales can accomplish better fusion results. However, the smaller the scale, the higher the complexity of the algorithm. Thus, in order to acquire better fusion results and also to balance the complexity of the algorithms, the scale of all structural elements takes uniformly the value in the pair (6,3).

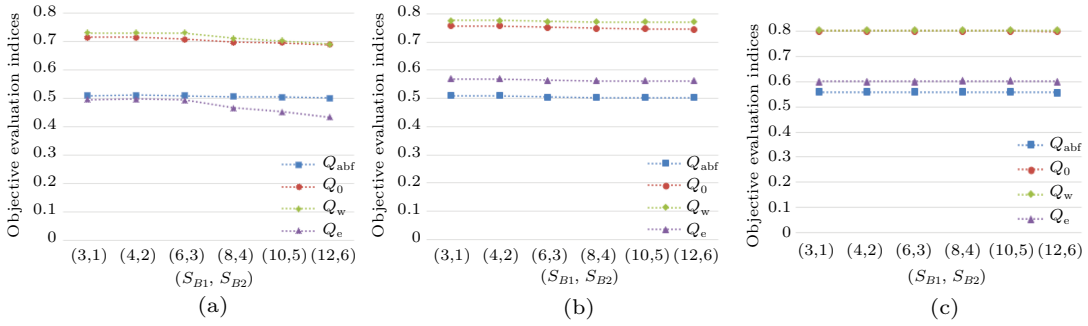


Fig. 5. (color online) Objective evaluation indices of fusion results for various methods with different scales. The horizontal coordinate represents scale and the vertical coordinate represents objective evaluation indices. Corresponding images are shown. (a) DWT+TH, (b) DTCWT+TH, (c) NSCT+TH.

4.2. Subjective and objective evaluation of fusion results

The first set of infrared and visible images is called the UNcamp group with size of 240×320, and its experimental results are shown in Fig. 6. The corresponding source image is given in Fig. 4(a). Comparing with Fig. 4(a), the human figure in the infrared source image is clearly seen but other information in the scene is rather weak. It is difficult to achieve target identification and location by a sole infrared image. Yet the visible image is full of rich information in the scene such as trees, huts, and fences but these objects are all hidden. Therefore, fusion of the two can generate a resultant image which has clear objects and rich scene information. In order to facilitate the observation of resultant images in the experiment, the target area is marked with a red box and enlarged while the outline texture area is marked with a green box and also enlarged. Figure 6(a) shows the fusion result with the TH1 method: the information on details (e.g., trees, road) is rich, indicating that the method can well retain details of the source image. However, from the enlarged image it is found that the method keeps only the outline information of the infrared target, and the energy loss of the target is severe. Figure 6(b) exhibits the fusion result with the TH2 method: evidently the

information on details is rich with the shortcoming of somewhat energy loss of the infrared target. For the SR, the JSR1, and the JSR2 methods, the fusion results all retain the complete target information but details of the scene such as leaves suffer a heavy loss. The JSR2 method is of relative richness in details as it combines local and global saliency. The fusion effect of the DWT, the DTCWT, and the NSCT methods is improved in turn. Certain degree of the Gibbs effect is associated with the DWT method, which shows that excellent multi-scale transforms can effectively improve fusion quality though the image tends to become smooth due to the average fusion rule for the low-frequency component, incurring reduced contrast. Compared to the corresponding multi-scale method, the herein proposed MST+TH method has tremendously improved the fusion quality: the infrared target is clear and the information on details is abundant. In particular, since the NSCT+TH method combines feature extraction in multiple scales and from diverse directions (from NSCT), translation invariance, and the separate processing of features and of background with the top-hat transform, in the resultant fusion image the infrared target is the clearest, the background is rich in details, the contrast is high, and the visual effect is far superior.

The second set of infrared and visible images has the size of 288×208, and is called the duck ground. Experimental results are shown in Fig. 7 and the corresponding source image is Fig. 4(b). It can be seen from the figure that the hot target duck in the infrared source image is very clear but the background is rather dark, and it is difficult to directly infer the exact location of the duck from a single image. In contrast, the visible source image has rich scene information but since the duck is hidden under water, thus, it is impossible to glean the hidden target information. Experimental results of different fusion methods are shown in Fig. 7: all methods can blend to different degrees the scene information of the infrared and the visible images. Further careful observations reveal that the fusion results of the TH1 and the TH2 methods are relatively good in terms of detail retention but the identifiability of the infrared target is rather poor; and in specific, the fusion result of the TH1 method keeps only the contour information of the infrared target. The SR method generates a high contrast but details are heavily lost: textures of grass and shrubs are somewhat lost and the visual effect is poor. For the JSR1 method, the fusion result is rich in detailed information but the brightness of the target and of the white wall in the background is obviously low, indicating that the energy in the low frequency is not well preserved. Since the JSR2 method has considered both the local and the global saliency, information on details is improved but the overall contrast is still not high enough. In the framework of MST, the DWT and the DTCWT methods have incurred dim overall brightness, low contrast, and loss of the brightness for the infrared target. Comparatively, the NSCT method has retained information of both the infrared

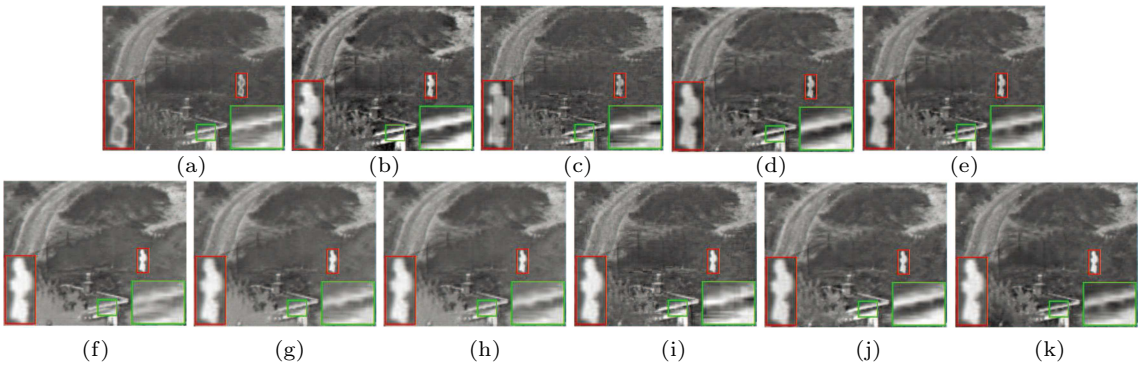


Fig. 6. (color online) Fusion results of the UNcamp group with different fusion methods. Corresponding images are shown. (a) TH1, (b) TH2, (c) DWT, (d) DTCWT, (e) NSCT, (f) SR, (g) JSR1, (h) JSR2, (i) DWT+TH, (j) DTCWT+TH, (k) NSCT+TH.

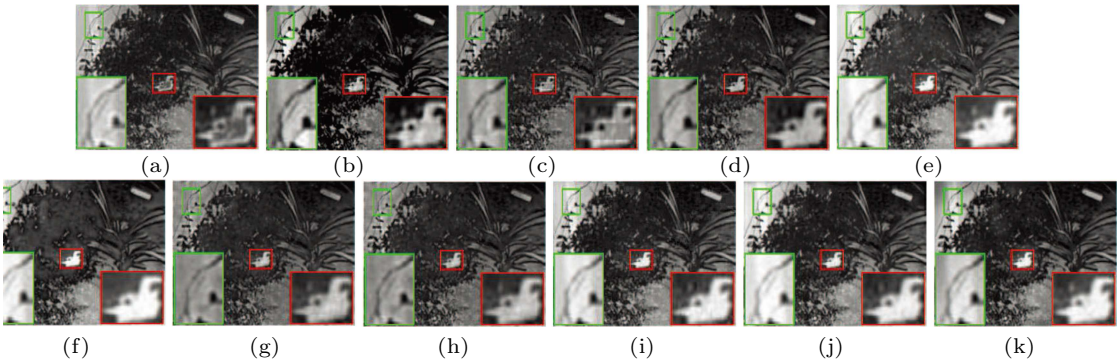


Fig. 7. (color online) Fusion results of the duck group with different fusion methods. Corresponding images are shown. (a) TH1, (b) TH2, (c) DWT, (d) DTCWT, (e) NSCT, (f) SR, (g) JSR1, (h) JSR2, (i) DWT+TH, (j) DTCWT+TH, (k) NSCT+TH.

target and the rich details. The fusion effect of the proposed MST+TH method has been starkly improved: the brightness and contour of the infrared target have both been preserved, detailed information and in particular on the edge and texture of the water grass is rich, and the overall visual effect is optimal. In order to facilitate observation of the fusion effect shown in Fig. 7, the infrared target is marked by a red frame and enlarged and the area of background details is marked by a green frame and also enlarged. It is evident that the MST+TH method not only retains a wealth of background details but also preserves the infrared target information, and has engendered good visual effects.

The third set of experimental images is called the street group. The infrared and visible source images have the size of 632×496, and the experimental results are shown in Fig. 8. The corresponding source image is given in Fig. 4(c). Pedestrians, vehicles, and street lights in the infrared source image are all clearly visible but information on the shops in the street cannot be distinguished at all; and if a traffic accident happens here, the policeman from the infrared image will not be able to acquire the intersection information at all. The shop information is available on the visible image but pedestrians and vehicles are not recognizable due to the restriction by the lighting condition, viz. the policeman is able to know which intersection it is but cannot grasp specific traffic information. By comparing fusion results from different fusion methods, it is revealed that the TH1 method generates low overall contrast, impairs brightness of the target, and incurs loss of details, which means that a single morphological top-hat transform cannot guarantee effective extraction of details. The detailed information with the TH2 method is well integrated but the overall brightness is rather weak, which is unfavorable to naked eyes. The SR method enables clear details and well-preserved objects in the fusion result, indicating that it is suitable for fusing such kind of images. However, the involved procedure of sparse representation is complicated and rather time consuming. In the case of generating similar fusion effects, a simple and efficient multi-scale-transform-based method shall be preferred. The JSR1 method does not well retain such information as on edges and details, and is not conducive to naked-eyes' observation. The JSR2 method gives much richer background details relative to the JSR1 method. Fusion results from the classical MST method are shown in Figs. 8(c)–8(e). The DWT and DTCWT methods incur severe information loss of the target and the detailed information is not good enough due to low contrast. Fusion results from the proposed MST+TH method are given in Figs. 8(i)–8(k). The overall brightness is moderate, the target information is clear, and the street information is well preserved, and the visual effect has been capitally improved. In order to elaborate the fusion effect, in Fig. 8, select the visible information (the trademark character), mark them with a green frame, and zoom in and place them in the lower-left corner on each graph. Furthermore, the infrared target information (pedestrians' legs) is marked with a red frame and zoomed in and placed in the lower-right corner. Evidently, the herein proposed fusion method is far superior to the traditional MST method: the result exhibits higher contrast and resolution, and enjoys better visual effect.

As can be seen from Table 1, compared to the celebrated top-hat transform, classical and the latest sparse methods, and the representative MST method, the evaluation indices $(Q_{\mathrm{abf}}, Q_0, Q_{\mathrm{w}}, Q_{\mathrm{e}})$ have all improved with the proposed method herein, which means that it performs well with respect to target retention and to integration of information details. Further comparison shows that compared to the celebrated top-hat method, the proposed method herein (i) preserves the advantage of effective integration of the bright-&-dark features of the image (inherited from the top-hat transform); (ii) efficiently extracts high-frequency features of the image in multiple scales and from diverse directions (via the multi-scale transform); (iii) enables clear infrared target, rich information details with the resultant image; and (iv) achieves better fusion effect. Compared with the representative MST method, the studied one has significantly improved all objective evaluation indices, which means that with the application of the top-hat transform, the bright-&-dark features extracted from the low-frequency component are further beneficial to details extraction from the low-frequency component. In particular for the NSCT+TH method, the combination of NSCT and TH has efficaciously improved the integration of edges

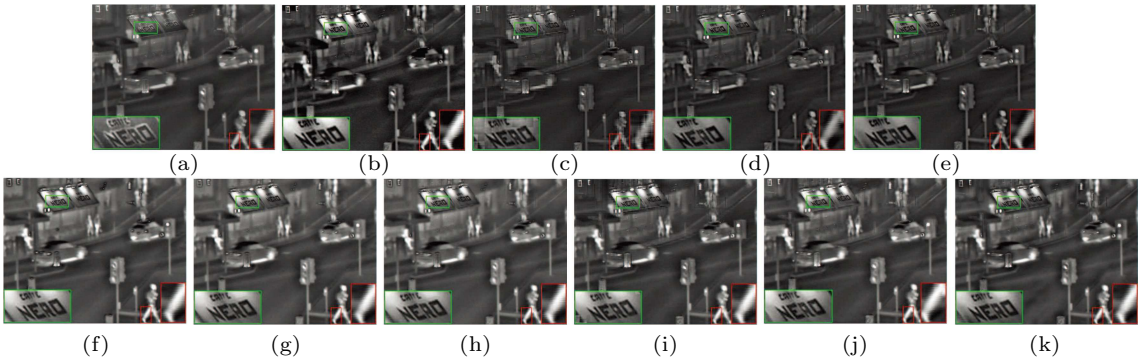


Fig. 8. Fusion results of the street group with different fusion methods. Corresponding images are shown. (a) TH1, (b) TH2, (c) DWT, (d) DTCWT, (e) NSCT, (f) SR, (g) JSR1, (h) JSR2, (i) DWT+TH, (j) DTCWT+TH, (k) NSCT+TH.

and other regions. Sparse methods are currently an excellent kind of fusion methods, and it is evident from the evaluation table that the Q_{abf} index shows obvious advantages, indicating that such a method can better retain the gradient information of the source image. In order to further compare the proposed method with the sparse method and taking computational running time as the index, experiments have been done with Matlab 2016b platform in a computer with 3.3 GHz Intel processor, 8 G RAM. Table 2 records the running time for each sparse method and for the proposed one. Data in the table show that sparse methods are heavily time-consuming since complicated solution processes are involved in dictionary training and sparse coding, which is in sharp contrast with the proposed one herein: less computational time, much simpler yet more efficient.

Table 1. Objective fusion evaluation indices (mean values) of the three sets of source images.

	Q_{abf}	Q_0	Q_w	Q_e
TH1	0.5911	0.6765	0.7466	0.5896
TH2	0.6281	0.8199	0.8213	0.6522
DWT	0.5311	0.7326	0.7405	0.5415
DTCWT	0.5558	0.7706	0.8008	0.6551
NSCT	0.6100	0.7953	0.8150	0.6808
SR	**0.6895**	0.7922	0.8429	0.6875
JSR1	0.6456	0.8096	0.84397	0.6779
JSR2	0.6464	0.8188	0.8560	70.7078
DWT+TH	0.6444	0.7974	0.8536	0.7276
DTCWT+TH	0.5923	0.8019	0.8644	0.7527
NSCT+TH	0.6737	**0.8376**	**0.8841**	**0.7767**

Table 2. Running time with different fusion methods.

Fusion method	TH1	TH2	DWT	DTCWT	NSCT	SR	JSR1	JSR2	DWT+TH	DTCWT+TH	NSCT+TH
Running time/s	2.07	2.24	3.29	5.79	39.15	4341	9497	9390	3.96	9.54	76.62

5. Conclusions

For the fusion of infrared and visible images, a novel method has been proposed to combine the multi-scale and top-hat transforms. Firstly, the source image is multi-scale transformed to obtain its low-frequency component and several high-frequency components. Next, different fusion rules are applied to integrate low- and high-frequency components, respectively. With the multi-scale transform, due to its limited capability of feature extraction, there still exist many important features in the low-frequency component, thus, the top-hat transform in mathematical morphology is applied to further extract background information and bright-&-dark features, and the latter is being fused via the rule of taking the maximum of the absolute value. Apply the rule of weighted factors to fuse the low-frequency background. For the low-frequency component, such means of handling separately the bright-&-dark features and the background information has effectively retained its many important features. For high-frequency components, taking into consideration of the single pixel and local regions, the product of characteristics is set as the fusion rule. Finally, the inverse multi-scale transform is applied to the high-&-low frequency components to yield the resultant fusion image. Experimental results have shown that the proposed method herein is superior to the classical and state-of-the-art methods. (i) Compared to MST methods, such as the DWT, the DTCWT, and the NSCT methods, separate treatment of features and background in the low-frequency component is helpful in preserving important features and for improving fusion performance. (ii) The proposed method is better than the one which is based singly upon the new top-hat transform, viz. the TH. in the sense that it can effectively overcome the drawback of incompleteness of feature extraction, which is inherent in top-hat transform. (iii) It is also much simpler and more efficient than those methods which are based upon sparse representation.

References

[1] Li S T, Kang X D, Fang L Y, Hu J W and Yin H T 2017 *Inform. Fusion.* **33** 100
[2] Zhang Q, Liu Y, Blum R S, Han J G and Tao D C 2018 *Inform. Fusion.* **40** 57
[3] Mishra D and Palkar B 2015 *Int. J. Comput. Appl.* **130** 7
[4] Ma J Y, Chen C, Li C and Huang J 2016 *Inform. Fusion.* **31** 100
[5] Cui G M, Feng H J, Xu Z H, Li Q and Chen Y T 2015 *Opt. Commun.* **341** 199
[6] Bavirisetti D P and Dhuli R 2016 *Ain Shams Engineering Journal*
[7] Sahu A, Bhateja V, Krishn A, *et al.* 2015 *IEEE International Conference on Medical Imaging*, 2015, p. 448
[8] Nooshyar M, Abdipour M and Khajuee M 2011 *Multi-focus Image Fusion for Visual Sensor Networks in Wavelet Domain* (New York: Pergamon Press) pp. 789–797
[9] Mehra I and Nishchal N K 2015 *Opt. Commun.* **335** 153
[10] Ye M and Tang D B 2015 *J. Electr. Measur. Instr.* **29** 1328 (in Chinese)
[11] Yang Y 2010 *IEEE International Conference on Bioinformatics and Biomedical Engineering*, 2010, p. 1
[12] Wang X H, Zhou Y and Zhou Y 2017 *Comput. Eng. Desig.* **38** 729 (in Chinese)
[13] Yu B, Jia B, Ding L, *et al.* 2016 *Neurocomputing.* **182** 1
[14] Sruthy S, Parameswaran L and Sasi A P 2013 *IEEE International Multi-Conference on Automation*, 2013, p. 160
[15] Liu S Q, Shao-Hai H U, Zhao J, *et al.* 2016 *J. Signal. Process.* (in Chinese)
[16] Chen X P, Yang X Z and Dong Z Y 2017 *Remot. Sens. Inf.* **32** 68 (in Chinese)
[17] Yang Y, Song T, Huang S, *et al.* 2015 *IEEE. Sens. J.* **15** 2824
[18] Zhao C, Guo Y and Wang Y 2015 *Infrared. Phys. Techn.* **72** 266
[19] Bai X, Zhou F and Xue B 2011 *Opt. Express* **19** 8444
[20] Chen T M, Wang J Q, Zhang X X, *et al.* 2016 *Laser. Infr.* **46** 357 (in Chinese)
[21] Yang B and Li S 2010 *IEEE. T. Instrum. Meas.* **59** 884
[22] Yu N, Qiu T, Bi F, *et al.* 2011 *IEEE. J-STSP.* **5** 1074
[23] Liu C H, Qi Y and Ding W R 2017 *Infrared. Phys. Techn.* **83** 94
[24] He G Q and Hao C Y 2007 *Comput. Eng. Appl.* **43** 71 (in Chinese)
[25] Xydeas C S and Petrovic V 2000 *Military Technical Courier* **56** 181
[26] Piella G and Heijmans H 2003 *IEEE International Conference on Image Processing* 2003, p. 173