Fusion of infrared and visible images based on nonsubsampled shearlet transform and block compressive sensing sampling

[1] Defa Hu and [2] Hailiang Shi

[1] Key Laboratory of Hunan Province for Mobile Business Intelligence, Hunan University of Commerce, Changsha 410205, Hunan, China, hdf666@163.com
[2] College of Mathematics and Information Science, Zhengzhou University of Light Industry, Zhengzhou 450002, Henan, China, 3812603@qq.com

Received: 05.05.2017

Abstract. Image fusion is aimed at combining source images of the same scene into a single fused image with more reliable and precise information, which is suitable for further image-processing tasks. Here we present a novel method for fusing infrared and visible images, which is based on nonsubsampled shearlet transform (NSST) and block compressive-sensing sampling (BCSS). Compressive-sensing approach is widely used in various fields and constitutes a basis of low-signal sampling compression method employed under conditions of sparsity of signals. We investigate the BCSS technique in image fusion and suggest a fusion method where the NSST is applied for decomposition and reconstruction, and two-measure forms of local region energy are presented for maximum-choosing fusion rules for the NSST subband coefficients. Our approach is tested on five standard datasets of infrared and visible images. The results are compared with those obtained using several common fusion methods. The experimental results illustrate validity and efficiency of our method. Both the efficiency and robustness of the fusion are analyzed using a number of image-quality measures.

Key words: image fusion, infrared images, nonsubsampled shearlet transform, compressive sensing, block sampling, local region energy

PACS: 42.30.Wb, 07.05.Pj
UDC: 004.932

1. Introduction

As two kinds of the most used image modalities, infrared and visible images contain complementary components of information on a given object [1]. Generally, infrared images reveal concealed objects of interest in poor environments but have low resolution. This makes it hard to recognize by a human since infrared sensors are sensitive only to temperature differences in a scene. Meanwhile, visible images have excellent resolution and can provide much more detailed information, since visual sensors are sensitive to frequency reflection of an object under inspection. As a result, fusion of infrared and visible images can provide a reliable and precise description of a scene, which is useful in subsequent information processing, e.g. target detection and intelligent surveillance.

There has been much research on the image fusion. The methods employed in the field are usually divided into two categories: spatial-domain and transform-domain ones. The fusion methods based on spatial domain rely directly on grey levels of pixels. The simplest is a spatial average method, which takes the average values pixel by pixel. Another fusion method based on the principal-component analysis uses a basic idea of 'component substitution'. Along with simplicity, there come several undesirable side effects, including contrast reduction. As a third-generation method based upon artificial neural network, a pulse-coupled neural network (PCNN)

can also be used in the image fusion. It can be qualified as a spatial-domain method since the PCNN-based fusion generally takes the source images as external stimuli. In our previous work [2], we have studied the PCNN and proposed a modified adaptive dual-channel PCNN for infrared-and-visible image fusion.

In the transform domain-based fusion methods, the source images are first decomposed into a sequence of subband images through a specific mathematical transformation structure, and then some fusion rules are applied to determine so-called subband coefficients [3]. At last, a resultant fused image is formed by the corresponding inverse transform. The mathematical transform is well known as a multiscale geometric analysis (MGA). The MGA includes Laplacian pyramid (LP), discrete wavelet transform (DWT), shift-invariant DWT (SIDWT), nonsubsampled contourlet transform (NSCT), shearlet transform, nonsubsampled shearlet transform (NSST), and some other varieties. The NSCT has initially been suggested by Da Cunha et al. [4] in 2006. It represents fully shift-invariant, multiscale and multidirectional transform scheme, which have been successfully used in the image fusion [5–7].

In 2008, Easley [8] has proposed a sparse directional-image representation of the NSST, which combines nonsubsampled Laplacian-pyramid transform with different shearing filters. In contrast to the NSCT, the NSST absorbs some recent developments in the MGA field and shows a satisfactory fusion performance, including better sparse-representation ability and much lower computational costs. In addition, the NSST manifests a shift-invariant property. It is hoped that further research in the area of image fusion using the NSST domain can be promising and competitive. To join together the advantages of different approaches, a number of fusion methods have been suggested, which are based on the combination of transforms [9]. For instance, Xiang et al. [10] have developed a fusion algorithm for the infrared and visible images by combining the NSCT with the dual-channel unit-linking PCNN. The experiments have testified that this hybrid method can significantly improve the fusion performance and outperforms the individual DWT- or NSCT-based methods.

Unlike the MGA tools, the approach of compressive sensing (CS) describes natural signals using a sparse linear combination of columns of an overcomplete dictionary. Being different from a limited dictionary in the multiscale transform, the CS uses an overcomplete dictionary in which a column is called as a signal atom. Overcompleteness is the most prominent characteristic of dictionaries in the CS theory. It indicates that the number of signal atoms is larger than the dimension of this signal, which ensures more meaningful and complete representation of source signals than that used by the traditional multiscale transforms. The CS theory reveals that the coefficients corresponding to the natural signals are sparse.

In the present work, we develop a novel fusion method of the infrared and visible images, which is based on the NSST and the CS theory. In frame of this method, the source images are first decomposed using the NSST, and then a fusion rule based on local-region energy maximum-choosing is used for low-frequency subbands. When choosing a maximum for the high-frequency subbands, we present another definition of local-region energy based on block compressive-sensing sampling (BCSS). After applying the inverse NSST, we arrive at a resultant fused image. Our experiments performed for a number of images and our computation results illustrate both the efficiency and robustness of our fusion method. A comparison with some common MGA-based fusion methods has also been carried out.

The article is organized as follows. In Section 2, the NSST and CS theories are briefly reviewed. In Section 3, we describe in detail our method for fusing infrared and visible images,

which is based on the NSST and the BCSS. Experimental results associated with five standard image datasets are documented and compared in Section 4. Finally, we draw conclusions in Section 5.

2. Related work

2.1. NSST theory

The NSST rests upon the nonsubsampled Laplacian-pyramid transform and several different combinations of the shearing filters. It is known that the NSST is a shift-invariant version of the shearlet transform. In order to eliminate the effects of upsampling and downsampling, the NSST utilizes nonsubsampled Laplacian-pyramid filters as a substitute for the Laplacian pyramid filters used in the shearlet transform. As a consequence, the NSST manifests shift-invariance, along with multiscale and multidirectional properties [11].

The NSST process involves two stages, a multiscale decomposition and a multi-directional decomposition. The nonsubsampled Laplacian pyramid is utilized to implement the multiscale property by using two-channel nonsubsampled filter bank. One low-frequency subband and one high-frequency subband are obtained at each nonsubsampled Laplacian-pyramid decomposition level. The subsequent nonsubsampled Laplacian-pyramid decompositions are implemented to decompose the low-frequency subband iteratively and capture singularities present in the image. The multidirectional property of the NSST is implemented using improved shearlet filters. The latter are formed through avoiding downsampling, to satisfy the property of shift-invariance. The shearlet filters allow multidirection decomposition for the high-frequency subband at each level. They produce 2^l directional subbands of the same size as the source image [11, 12]. More details on the subject matter can be found in Ref. [8].

2.2. CS theory

The CS theory [13] has recently attracted huge attention in the areas of applied mathematics and signal processing, since it can surpass the common limits of Nyquist sampling theory. The CS is built upon the fundamental fact that, in a suitable basis or dictionary, many signals can be represented using only a few non-zero coefficients. A nonlinear optimization can then enable recovering of such signals from the very few measurements. In this section, we will review in brief the basic CS theory.

Theoretical foundations of the modern signal processing have been described in well-known pioneering works by Kotelnikov, Nyquist, Shannon and Whittaker on sampling continuous-time band-limited signals. Their results gave demonstrated that any signals, images, videos or other data can be exactly recovered from a set of uniformly spaced samples taken at a so-called Nyquist rate, which is twice the highest frequency present in the signal under interest [13]. Unfortunately, for many important and emergence applications the resulting Nyquist rate is so high that we end up with far too many samples. Alternatively, building the devices capable of acquiring samples at a necessary rate can simply turn out to be too costly, or even physically impossible [13].

To solve the problem, one can employ the methods of transform coding or the CS. The transform coding relies typically on finding a basis that provides sparse or compressible representations for the signals. The CS enables potentially large reduction in the sampling and the computation cost for the case of sensing signals which have sparse or compressible representations. Rather than sampling at first at a high rate and then compressing the sampled data, we would like to find the ways to sense directly the data in a compressed form, i.e. at a lower sampling rate. It is proved that a finite-dimensional signal having a sparse or compressible representation can be

recovered from a small set of linear nonadaptive measurements [13].

Consider a real-valued and finite-length one-dimensional signal $x \in \mathbb{R}^N$, which can be regarded as an $N \times 1$ column vector. If x is K-sparse, it can be represented as

$$x = \Psi\alpha, \tag{1}$$

where Ψ is the $N \times N$ basis and α the $N \times 1$ vector containing only K non-zero elements. If $K \ll N$, the signal x is compressible. Then both x and α are equivalent representations of the signal, with x corresponding to the space domain and α to the Ψ domain.

In frame of the CS, we do not measure α directly, taking instead the compressive measurements

$$y = \Phi x, \tag{2}$$

where $y \in \mathbb{R}^M$ and Φ denotes the $M \times N$ measurement matrix which is fixed and does not depend on the signal x. Then, substituting x into Eq. (1), one can write y as

$$y = \Phi\Psi\alpha = A\alpha,$$

where $A = \Phi\Psi$ is the $M \times N$ sensing matrix [1]. The components of this matrix contain significant information for reconstructing the source signal. Since we have $M \ll N$, resolving α from y implies pursuing a solution of an under-determined equation, which represents an NP-hard problem. Nevertheless, since x is K-sparse, x can be reconstructed with high enough accuracy after solving a nonlinear optimization problem under certain conditions, such as restricted-isometry property and uniform-uncertainty principle [3].

After nonadaptive linear projecting, the major challenge in the CS is to reconstruct accurately the vector x from a set of incomplete and inaccurate samplings. At present, the relevant reconstruction algorithms can be divided into three types: combinatorial algorithms, convex-relaxation algorithms and greed-pursuit algorithms. The first algorithms are more efficient for complex systems though they require somewhat especial samples which are difficult to acquire. The second methods are often accompanied with a high computational complexity. Finally, the third algorithms reveal both proper computational complexity and commendable reconstruction performance. The researchers in the field have developed many greed-pursuit algorithms, including matching pursuit, orthogonal matching pursuit, stage-wise orthogonal matching pursuit and compressive-sampling matching pursuit algorithm. Since in this work we use the inverse NSST to reconstruct the images, we will not discuss the reconstruction algorithms in detail.

3. Our fusion method

In this section, we introduce the NSST into the image fusion and suggest a new method for fusing the infrared and visible images.

3.1. The framework of our fusion method

Let the fused image be F, and the source images be A and B, respectively. We assume that the source images have been well detected. Fig. 1 illustrates the framework of our fusion method. The main procedures of the method include the following steps:

(1) apply a J-level NSST to A and B, and obtain the corresponding subband coefficients $\{C_A^J(x,y), C_A^{j,\theta}(x,y)(1 \le j \le J, \theta = 1,2,\cdots,2^{l_j})\}$ and $\{C_B^J(x,y), C_B^{j,\theta}(x,y)(1 \le j \le J, \theta = 1,2,\cdots,2^{l_j})\}$, where $C^J(x,y)$ denote the low-frequency subband coefficients at the coarsest scale, and

$C^{j,\theta}(x,y)$ the directional high-frequency subband coefficients at the j th scale and in the θ th direction. Here l_j stands for the directional decomposition level of the NSST at the j th scale;

(2) use a maximum-choosing fusion rule based on local-region energy to determine the low-frequency and directional high-frequency subband coefficients. Then the fusion coefficients of the fused image F $\{C_F^J(x,y), C_F^{j,\theta}(x,y)(1 \le j \le J, \theta = 1,2,\cdots,2^{l_j})\}$ can be obtained. The fusion rule will be discussed in detail in the next subsections;

(3) perform the inverse NSST to the fused coefficients to obtain the fused image F.

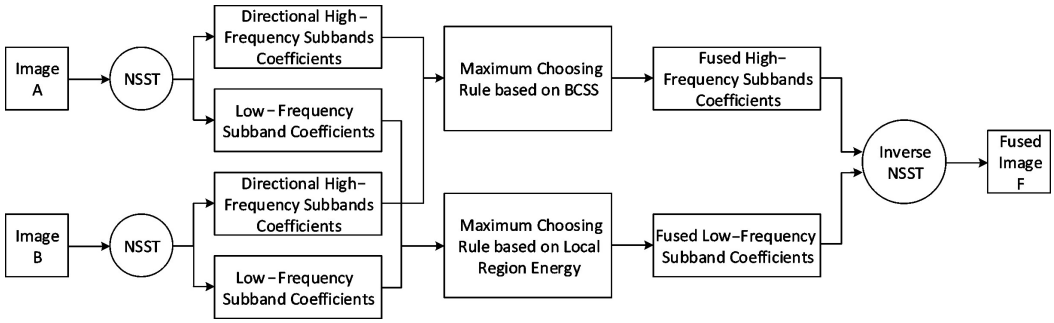

Fig. 1. Block diagram of our fusion method.

3.2. Fusion rule for the low-frequency subband

The low-frequency subband is an approximate version of the source image and includes usually the average greyscale information. Although the simplest way is to use a common averaging method, it often results in low contrast. Considering that the infrared sensor is sensitive only to the difference of temperatures, a human or any other heat sources are usually very prominent in the form of high grey values in the infrared images, and the other parts are usually dark. To capture more information from the objects in the infrared images, we employ a maximum-choosing rule based on local-region energy for fusion and using a standard sliding-window technique.

The local-region energy is defined by the L2-norm of the vectorized coefficients within the sliding window, and the sliding window contains $m \times n$ neighbourhoods of the current pixel:

$$E_1(x,y) = \left\| C^J(x+u, y+v) \right\|_2, \tag{3}$$

where $\|\cdot\|_2$ represents the L2-norm of the vector and (u,v) the location of the sliding window. Then the maximum-choosing fusion rule may by written as

$$C_F^J = \begin{cases} C_A^J(x,y), & if \quad E_1^A(x,y) \ge E_1^B(x,y) \\ C_B^J(x,y), & otherwise \end{cases}. \tag{4}$$

3.3. Fusion rule for the high-frequency subbands

The directional high-frequency subbands represent such details of the source images as edges, lines, contours and the region boundaries in sparse-presentation domain. For preserving more detailed information and overcoming the defects, we utilize the BCSS technique.

In the BCSS, each subband X is divided into blocks of size $b \times b$. Then the blocks are sampled using an orthogonal random matrix, or a so-called measurement matrix. Suppose that $X_{i,j}$ are vectorized neighbourhood coefficients representing the block centre at the location (i,j) in a raster-scan fashion. The corresponding measurement data $Y_{i,j}$ are then obtained as

$$Y_{i,j} = \Phi_B X_{i,j}, \tag{5}$$

where Φ_B is the $m \times b^2$ Gaussian random orthonormal measurement matrix with $m = \lfloor r \cdot b^2 \rfloor$, and $r \in (0,1)$ the sampling rate. The reason for utilizing the Gaussian random matrix is that this matrix is incoherent with any sparse basis and requires less number of measurements to acquire the salient information. Moreover, it can yield better results in the inverse solution and restrict the mean errors to project sparse coefficients [14].

Using the BCSS rather than the random sampling applied to the entire image has several advantages. First, the measurement matrix Φ_B is conveniently stored and employed due to its compact size. Second, the projection is more flexible since it can send each block after linear projection and needs no waiting until the entire image is measured [15].

Another definition of the local-region energy based on the BCSS is suggested in this study, which is expressed as

$$E_2(\mathrm{i,j}) = \left\| Y_{i,j} \right\|_1, \tag{6}$$

where $\|\cdot\|_1$ represents the L1-norm of a vector. Subsequently, the maximum-choosing fusion rule for the directional high-frequency subbands may be written as

$$C_{j,\theta}^F(\mathrm{x,y}) = \begin{cases} C_{j,\theta}^A(x,y), & if \;\; E_2^A(x,y) \geq E_2^B(x,y) \\ C_{j,\theta}^B(x,y), & otherwise \end{cases}. \tag{7}$$

4. Experimental results and their discussion

4.1. Experimental introduction

Five sets of fusion experiments have been carried out to show the validity of our fusion method. The experimental results have proved that the method achieves a better fusion quality than that typical for the common fusion methods, including those based on LP, DWT, SIDWT, NSCT and NSST (see Introduction).

The experimental details are as follows: (1) the NSST decomposition level is equal to 3, the directional numbers are 4, 9 and 16, and the corresponding local support of the shearing filters amounts to 32, 32 and 16, (2) the sizes of the sliding windows are typically equal to 5×5, and (3) the sampling rate is 0.3. In the other fusion methods based on the multiscale transform, the fusion rules all take averages for the low-frequency subbands and use the absolute maximum-choosing rule for the high-frequency subbands, whereas the decomposition levels are also 3. The standard test images used in our experiments are datasets 'quad', 'octec', 'UNcamp', 'trees' and 'dune'. They are available from the website http://www.imagefusion.org. All the experiments have been implemented in Matlab 2015a on an Intel Core i3-4000M 2.4GHz CPU PC with 4GB RAM.

In the fusion-performance comparison, both subjective and objective quantitative evaluations are utilized to analyze the quality of the fused images. The quantitative evaluation criteria are standard deviation, average gradient, Shannon entropy and mutual information:

(1) the standard deviation reflects how much variation or dispersion of the greyscale of pixels differs from the average value. High-standard deviations indicate that the fusion data points are spread out over a large range of values, and so is the fusion quality [3]:

$$SD(F) = \left[\frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (F(i,j) - u) \right]^{\frac{1}{2}}, u = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} F(i,j); \tag{8}$$

(2) the average gradient reflects the features of small-detail contrast and the texture changes.

It can characterize the clarity of an image [14]:

$$AG(F) = \frac{1}{(m-1)(n-1)} \cdot \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} ((F(i,j) - F(i+1,j))^2 + (F(i,j) - F(i,j+1))^2)^{\frac{1}{2}}; \quad (9)$$

(3) the Shannon entropy reflects directly the amount of average information contained in the fused image [11]. The larger the Shannon entropy, the more abundant information amount is:

$$SE = -\sum_{i=0}^{L-1} p_i \log_2 p_i , \quad (10)$$

where p_i indicates the probability of pixels whose grey value amounts to i over the total image pixels;

(4) the mutual information measures similarity of two images, thus reflecting the fused quality [16]. The larger the mutual information, the more abundant information is contained in the fused image:

$$MI(I,F) = \sum_{i,f} p_{I,F}(i,f) \log \frac{p_{I,F}(i,f)}{p_I(i) p_F(f)}, \quad (11)$$

where $p_{I,F}$ is the jointly normalized histogram of I and F, p_I and p_F are the normalized histograms of I and F, and i and f represent the pixel values of the images I and F, respectively. The mutual information used in this work is calculated by averaging the two corresponding values between the source image and the fused image.

In addition, we evaluate the computational complexity of the fusion methods using the time (in seconds) needed for computing the fusion procedure. It is acquired by averaging the corresponding values of 20 times of the fusion procedure.

4.2. Performance results

The first experiment has been performed with the infrared and visible 'quad' images of the size 496×632, as shown in Fig. 2a, b. These images are characterized by highly structured statistical properties, and there is apparent information lost for the human vision. A better image quality is expected for the natural environment. The corresponding fused images are displayed in Fig. 2c–h. They are obtained using the fusion methods based upon LP, DWT, SIDWT, NSCT, NSST and the method developed in this work.

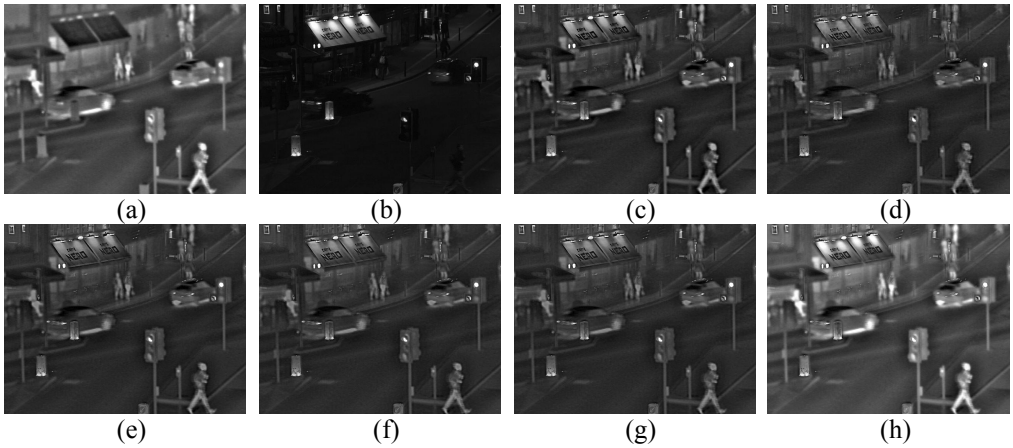

Fig. 2. 'quad' image dataset and fused images: infrared image (a), visible image (b), and fused images based on LP (c), DWT (d), SIDWT (e), NSCT (f), NSST (g) and our fusion method (h).

From the viewpoint of human visual system, the results of our method are superior to those derived with the other common fusing methods, because almost all of the useful information has been transferred into the fused image. In particular, the pedestrians and the cars in the image (h) are more prominent from a subjective viewpoint.

The objective performance characteristics are listed in Table 1, where the best results are indicated in bold. The quantitative values of the first four criteria clearly testify that the fusion method suggested in this work enables a better fusion quality in terms of the largest values of square deviation, average gradient, Shannon entropy and mutual information. However, our method does not achieve the shortest computing time. One of the reasons is that each subband image generated by the NSST has the same size as the source image, as follows from the NSST property of shift-invariance. The other reason is complicated BCSS calculations needed in frame of our method.

Table 1. Performance results obtained for the fused images.

Image dataset	Method	Standard deviation	Average gradient	Shannon entropy	Mutual information	Time, s
quad	LP	0.0664	0.0346	6.2486	0.6503	**0.0827**
	DWT	0.0632	0.0296	6.0317	0.6457	0.1057
	SIDWT	0.0635	0.0307	6.0780	0.6958	0.3411
	NSCT	0.0626	0.0292	6.0054	0.6809	143.4620
	NSST	0.0625	0.0291	6.0018	0.6847	6.1766
	Our method	**0.0781**	**0.0480**	**6.7739**	**1.3941**	17.1521
octec	LP	0.1077	0.0421	6.4417	0.9167	**0.0758**
	DWT	0.0919	0.0408	6.4018	0.9346	0.1030
	SIDWT	0.0952	0.0410	6.4020	0.9552	0.3222
	NSCT	0.0908	0.0406	6.3856	0.9619	141.2890
	NSST	0.0903	0.0406	6.3849	0.9618	4.8944
	Our method	**0.1412**	**0.0514**	**6.7532**	**1.4032**	15.9276
UNcamp	LP	0.1154	0.0401	6.5712	0.5151	**0.0270**
	DWT	0.1125	0.0365	6.4402	0.5010	0.0347
	SIDWT	0.1130	0.0369	6.4538	0.5211	0.0764
	NSCT	0.1123	0.0359	6.4032	0.5292	45.8403
	NSST	0.1122	0.0358	6.3982	0.5281	1.3348
	Our method	**0.1466**	**0.0438**	**6.7580**	**0.8223**	4.9475
trees	LP	0.1089	0.0228	5.9798	0.5011	**0.0277**
	DWT	0.0987	0.0215	5.9022	0.4861	0.0380
	SIDWT	0.1002	0.0215	5.9029	0.5077	0.0883
	NSCT	0.0972	0.0211	5.8775	0.5076	49.3519
	NSST	0.0969	0.0211	5.8727	0.5079	1.4873
	Our method	**0.1186**	**0.0241**	**6.0481**	**0.6969**	5.6890
dune	LP	0.0604	0.0223	6.0593	0.7136	**0.0336**
	DWT	0.0572	0.0213	6.0134	0.6950	0.0461
	SIDWT	0.0573	0.0214	6.0151	0.7335	0.0975
	NSCT	0.0563	0.0211	6.0004	0.7321	51.5926
	NSST	0.0562	0.0210	5.9990	0.7314	1.6313
	Our method	**0.0634**	**0.0268**	**6.2010**	**0.9124**	6.0572

The other four experiments have been carried out on the standard datasets 'octec' (480×640), 'UNcamp' (270×360), 'trees' (270×360) and 'dune' (270×360). The source images and the corresponding fused image are illustrated in Fig. 3 to 6. From the point of view of human visual system, one arrives at the conclusions similar to those formulated above for the first experiment. Our fusion method not only highlights the targets of interest in the scenario, but can also preserves the more detailed information on the edges and textures. The corresponding quantitative evaluation results are listed in Table 1. One can find that our method is characterized by the best performance parameters, except for the computing time.

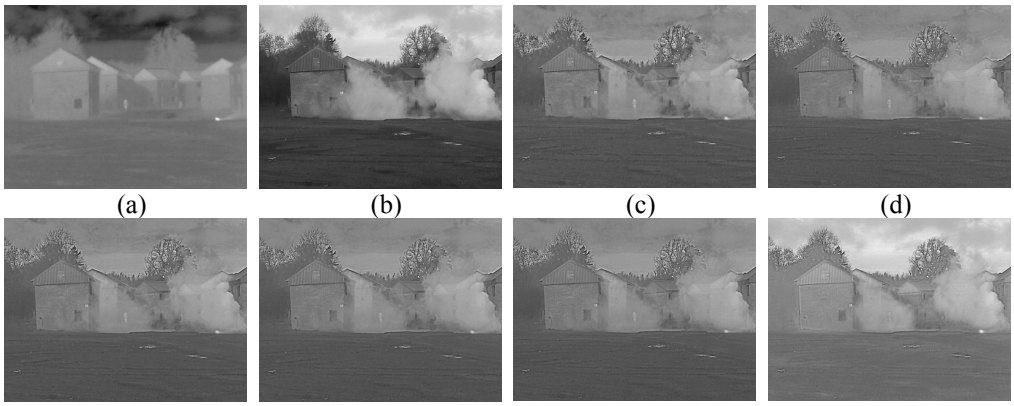

Fig. 3. 'octec' image dataset and fused images: infrared image (a), visible image (b), and fused images based on LP (c), DWT (d), SIDWT (e), NSCT (f), NSST (g) and our fusion method (h).

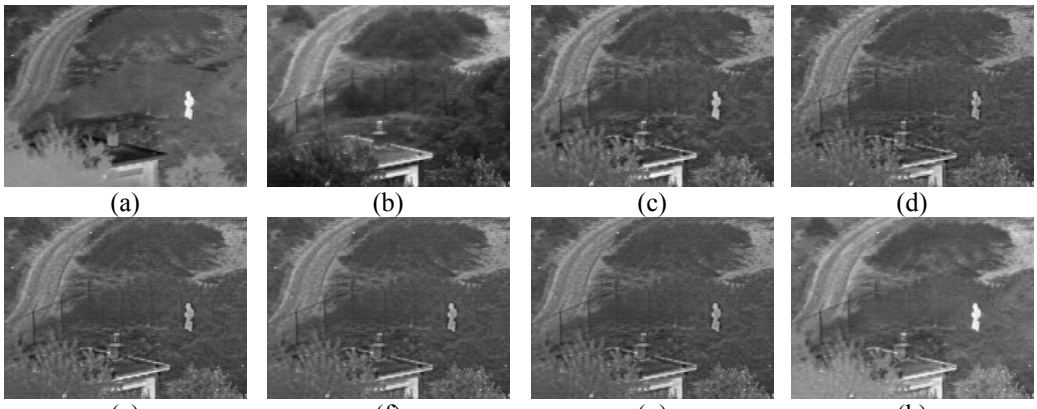

Fig. 4. 'UNcamp' image dataset and fused images: infrared image (a), visible image (b), and fused images based on LP (c), DWT (d), SIDWT (e), NSCT (f), NSST (g) and our fusion method (h).

4.3. Discussion

Following from the experimental data and their comparison, we conclude that our fusion method is superior to the traditional MGA-based methods according to both the subjective visual effect and the objective quantitative characteristics. As for comparison with the state-of-the-art, below we give a brief explanation, using some qualitative arguments. In our previous work [2] we have employed the method for fusing the infrared and visible images, which is based on the multiscale top-hat transform and the modified adaptive dual-channel PCNN. As mentioned in Introduction, it represents in fact a spatial domain method. On the one hand, it inherits the advantages peculiar for the spatial-domain methods, such as low computational complexity, fast speed and easy

implementation. On the other hand, it also inherits the disadvantage of low contrast, especially when comparing with the method suggested in the present work. As a result, the present method acquires a perfect fusion effect, failing only in the fusion speed.

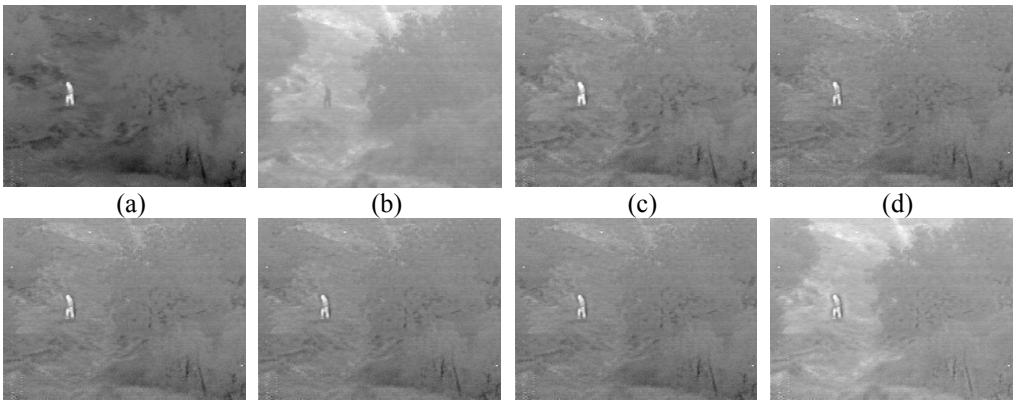

Fig. 5. 'trees' image dataset and fused images: infrared image (a), visible image (b), and fused images based on LP (c), DWT (d), SIDWT (e), NSCT (f), NSST (g) and our fusion method (h).

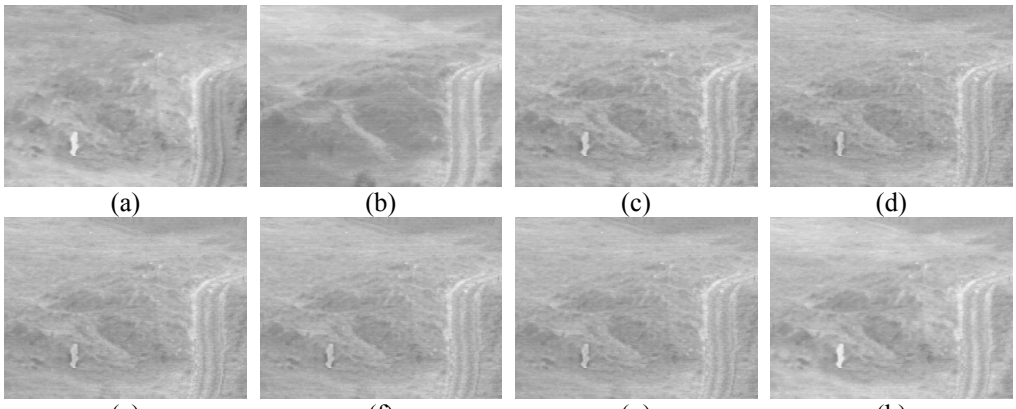

Fig. 6. 'dune' image dataset and fused images: infrared image (a), visible image (b), and fused images based on LP (c), DWT (d), SIDWT (e), NSCT (f), NSST (g) and our fusion method (h).

The fusion method proposed in Ref. [10] combines the characteristics of multiresolution and multidirection of the NSCT and the advantages of global coupling and pulse synchronization of the unit-linking PCNN. Similar to our method, both the experiments and the data show that the fusion effect is characterized with a high contrast and prominent target information. The quantitative evaluation yields in the optimal standard deviation, Shannon entropy and mutual information, which are superior to those of the individual DWT- or NSCT-based methods. However, the method of Ref. [10] involves the NSCT and the corresponding computing time is one order of magnitude larger than that of our method. This can be readily seen from our experimental results. Then the computing time of the whole method must be at least one order larger than that of our method. Therefore, we draw the conclusion that our method is slightly better.

5. Conclusion

In this work we have developed a novel infrared-and-visible image fusion method based on the NSST and the BCSS. As a recently suggested multiscale geometric analysis tool, the NSST is not only remarkably superior to the other common tool in terms of information capturing and

computational-costs saving, but also overcomes a lack of shift-invariance in the shearlet transform. The NSST is used to decompose and reconstruct the subband coefficients. Then the two measure forms of local-region energy are presented for maximum-choosing fusion rules, one of which is based on the L2-norm of the vectorized neighbourhood coefficients within a sliding window, and the other utilizes the L1-norm of the measurements of block CS sampling. This method is capable of merging the complementary information, highlights the targets of interest and improves image contrast and details. We have compared our method with several common fusion methods based on the multiscale transform, including the LP, the DWT, the SIDWT, the NSCT and the NSST. The experimental results illustrate the validity and efficiency of our method in terms of the standard deviation, the average gradient, the Shannon entropy and the mutual information. Visual human perception of the fused images also indicates that the method suggested in the present work is successful in transferring the complementary information.

Acknowledgement

This work has been supported by the National Natural Science Foundation of China (No 61202464) and the Natural Science Research Program of Henan Educational Committee (No 14A120012).

References

1. Li X and Qin S Y, 2011. Efficient fusion for infrared and visible images based on compressive sensing principle. IET Image Process. **5**: 141–147.
2. Hu D, Shi H and Jiang, W, 2016. Infrared and visible image fusion using multiscale top-Hat transform and modified adaptive dual-channel pcnn. Rev. Téc. Ing. Univ. Zulia. **39**: 173–180.
3. Zhang Q and Maldague X, 2016. An adaptive fusion approach for infrared and visible images based on NSCT and compressed sensing. Infrared Phys. Technol. **74**: 11–20.
4. Da Cunha A L, Zhou J and Do M N, 2006. The nonsubsampled contourlet transform: theory, design, and applications. IEEE Trans. Image Process. **15**: 3089–3101.
5. Bhatnagar G, Wu Q M and Liu Z, 2013. Directive contrast based multimodal medical image fusion in NSCT domain. IEEE Trans. Multimedia. **15**: 1014–1024.
6. Jianhua Adu, Jianhong Gan, Yan Wang and Jian Huang, 2013. Image fusion based on nonsubsampled contourlet transform for infrared and visible light image. Infrared Phys. Technol. **61**: 94–100.
7. Li H, Chai Y and Li Z, 2013. Multi-focus image fusion based on nonsubsampled contourlet transform and focused regions detection. Optik. **124**: 40–51.
8. Easley G, Labate D and Lim W Q, 2008. Sparse directional image representations using the discrete shearlet transform. Appl. Comput. Harmonic Analysis. **25**: 25–46.
9. Li S, Kang X, Fang L, Hu J and Yin H, 2017. Pixel-level image fusion: A survey of the state of the art. Inform. Fusion. **33**: 100–112.
10. Xiang T, Yan L and Gao R, 2015. A fusion algorithm for infrared and visible images based on adaptive dual-channel unit-linking PCNN in NSCT domain. Infrared Phys. Technol. **69**: 53–61.
11. Kong W, Zhang L and Lei Y, 2014. Novel fusion method for visible light and infrared images based on NSST–SF–PCNN. Infrared Phys. Technol. **65**: 103–112.
12. Luo X, Zhang Z and Wu X, 2016. A novel algorithm of remote sensing image fusion based on shift-invariant Shearlet transform and regional selection. AEU Int. J. Electron. Commun. **70**: 186–197.
13. Davenport M A, Duarte M F, Eldar Y C and Kutyniok G, 2011. Introduction to compressed

sensing. Preprint. **93**: 1–68.
14. Yin Hongpeng, Liu Zhaodong, Fang Bin and Li Yanxia, 2015. A novel image fusion approach based on compressive sensing. Opt. Commun. **354**: 299–313.
15. Mun S and Fowler J E, 2009. Block compressed sensing of images using directional transforms. Proc. 16th IEEE Int. Conf. Image Process: 3021–3024.
16. Yang B and Li S, 2012. Pixel-level image fusion with simultaneous orthogonal matching pursuit. Inform. Fusion. **13**: 10–19.

Defa Hu and Hailiang Shi. 2017. Fusion of infrared and visible images based on nonsubsampled shearlet transform and block compressive sensing sampling. Ukr.J.Phys.Opt. **18**: 156 – 167

Анотація. *Злиття зображень використовують для об'єднання вихідних зображень однієї і тієї ж сцени в єдине злите зображення з надійнішою і точнішою інформацією, яка годиться для подальших завдань обробки. Ми представляємо новий метод злиття інфрачервоних і видимих зображень, який базується на безнаповнених шеарлет-перетвореннях (БНШП) і блочній вибірці, заснованій на стиснутому зчитуванні (БВСЗ). Підхід вибірки зі стиснутим зчитуванням широко використовують у різних областях. Він представляє собою основу методу стиснення вибірки з низьким рівнем сигналу, який застосовують для рідкісних або розсіяних сигналів. Ми досліджуємо техніку БВСЗ для злиття зображень і пропонуємо метод злиття, в якому БНШП застосовують для розкладання та реконструкції зображення, а двопараметричні форми локальної енергії області представлено для правил злиття для піддіапазонних коефіцієнтів БНШП, заснованих на обранні максимуму. Наш підхід перевірено на п'ятьох стандартних наборах інфрачервоних і видимих зображень. Відповідні результати порівняно з даними, одержаними за допомогою декількох традиційних методів злиття. Результати експериментів ілюструють коректність і ефективність нашого методу. Проаналізовано ефективність та надійність злиття зображень за допомогою кількох мір якості зображень.*