Infrared Physics & Technology 74 (2016) 11–20

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

An adaptive fusion approach for infrared and visible images based on NSCT and compressed sensing

Qiong Zhang *, Xavier Maldague

MIVIM & Computer Vision and Systems Laboratory, Department of Electrical and Computer Engineering, Laval University, Quebec, QC G1V 0A6, Canada

HIGHLIGHTS

- Introduce the pre-enhancement step for the infrared image.
- Introduce the compressed sensing to sparsely sample the high-frequency subbands.
- Introduce the adaptive-Gaussian method to the fusion rule.
- Introduce the gradient descent iterative recovery algorithm to recover the fused high-frequency coefficients.

ARTICLE INFO

Article history:
Received 23 February 2015
Available online 2 December 2015

Keywords:
Image fusion
NSCT-AG-CSTV
Adaptive Gaussian
NSCT
Compressed sensing
Contrast enhancement

ABSTRACT

A novel nonsubsampled contourlet transform (NSCT) based image fusion approach, implementing an adaptive-Gaussian (AG) fuzzy membership method, compressed sensing (CS) technique, total variation (TV) based gradient descent reconstruction algorithm, is proposed for the fusion computation of infrared and visible images.

Compared with wavelet, contourlet, or any other multi-resolution analysis method, NSCT has many evident advantages, such as multi-scale, multi-direction, and translation invariance. As is known, a fuzzy set is characterized by its membership function (MF), while the commonly known Gaussian fuzzy membership degree can be introduced to establish an adaptive control of the fusion processing. The compressed sensing technique can sparsely sample the image information in a certain sampling rate, and the sparse signal can be recovered by solving a convex problem employing gradient descent based iterative algorithm(s).

In the proposed fusion process, the pre-enhanced infrared image and the visible image are decomposed into low-frequency subbands and high-frequency subbands, respectively, via the NSCT method as a first step. The low-frequency coefficients are fused using the adaptive regional average energy rule; the highest-frequency coefficients are fused using the maximum absolute selection rule; the other high-frequency coefficients are sparsely sampled, fused using the adaptive-Gaussian regional standard deviation rule, and then recovered by employing the total variation based gradient descent recovery algorithm.

Experimental results and human visual perception illustrate the effectiveness and advantages of the proposed fusion approach. The efficiency and robustness are also analyzed and discussed through different evaluation methods, such as the standard deviation, Shannon entropy, root-mean-square error, mutual information and edge-based similarity index.

1. Introduction

With the rapid development and improvement of sensor signal processing technology, infrared (IR) sensors have been extensively used in various fields, such as camera surveillance, medical imaging, and machine vision to detect the desired object or its related useful information. Infrared sensors have the capability of sensing and acquiring differences in the mode of thermal energy intensity reflected from targets in a scene, which is more effective than visible sensors under poor lighting condition, such as during the night.

There has been much research on vision enhancement, target detection and tracking from infrared images utilizing frequency and spatial domain algorithms. Nevertheless, the image data

* Corresponding author at: Department of Electrical and Computer Engineering, Laval University, 1065 avenue de la Médecine, Quebec City G1V 0A6, Canada.
E-mail addresses: qiong.zhang.1@ulaval.ca (Q. Zhang), maldagx@gel.ulaval.ca (X. Maldague).

http://dx.doi.org/10.1016/j.infrared.2015.11.003
1350-4495/

acquired by different sensors, with different physical and technological limitations, demonstrate diverse modalities and characteristics. By combining data sets of two or more different sensors, a surveillance system performs better than the condition when one single sensor is simply used. This technology is named image fusion [1,2], which transforms same-scene images into a single composite image with highly comprehensive information, thus leading to more suitable visual perception and better processing results. Currently, image fusion technology has been broadly applied in digital imaging, biomedical imaging, remote sensing, computer vision, etc.

The image fusion methods are usually divided into spatial domain and transform domain techniques [3]. Fusion methods in the spatial domain are directly functioned on pixel gray level or color space from the source images for the operation, thus the spatial domain fusion schemes are usually known as single-scale fusion methods. Regarding the transform domain based fusion methods, each source image is decomposed, at the beginning, into a sequence of subband images through a particular mathematical transformation structure. Afterwards, specific fusion rules are applied into the synthesis of corresponding subbands to obtain fused coefficients. Eventually, the fusion image is obtained by way of a mathematical inverse transformation. Therefore, the transform domain fusion methods are generally known as multi-scale fusion methods.

Generally, the transform domain fusion schemes include methods on the basis of Laplacian pyramid (LP), wavelet transform (WT), contourlet transform (CT), nonsubsampled contourlet transform (NSCT), etc.

The pyramid method was first proposed by Burt [4], who introduced the low-pass Laplacian pyramid for binocular fusion. The Laplacian pyramid has the bandpass equivalence of the Gaussian pyramid, and it's obtained by the subtraction between two successive low-pass Gaussian pyramid levels. The fusion application on visible, thermal and infrared images was early proposed by Rogers et al. [5] in 1989. After that, plenty of improved pyramid-based approaches, including filter subtract decimation, morphological and gradient pyramids, were proposed in some other image fusion works.

Traditional wavelet-based fusion methods have been studied and proved to be capable of providing image fusion with a better quality than the traditional Laplacian pyramid methods [6]. Mallet [7] first defined and proposed a mathematical model as a complete wavelet representation around the concept of multiresolution analysis (MRA). Hunstberger [8] introduced a wavelet-based image fusion approach in the early years. Li et al. [9] used wavelets as alternative basis functions for multisensor image fusion, which has the capability to overcome the limitations of pyramid-based schemes because of its directionality. After that, a lot of fusion techniques have been proposed to employ the characteristics of wavelet in the image fusion field.

However, neither the Laplacian pyramid methods nor the wavelet methods are able to rescue the loss of image information, such as the edges, contours and curves, which implies that a large amount of details are lost. Do and Vetterli [10] first proposed the Contourlet transform (CT), which gives the asymptotic optimal representation of contours, which can be successfully utilized for image fusion. It combines the Laplacian pyramid (LP) and the pyramid directional filter bank (PDFB) together, and considers two-dimensional filtering to decompose an image into several directional subbands in multiple scales. It has significant characteristics such as multi-resolution, localization and directionality, and it can represent edges and singularities along curves more efficiently than the wavelet-based transforms.

Nevertheless, due to the downsamplers and upsamplers acted in the process of decomposition and reconstruction, the Contourlet transform does not have enough shift invariance, which produces a pseudo-Gibbs effect and thus leads to distortion in image fusion processing. The non-subsampled contourlet transform (NSCT), proposed by da Cunha et al. in 2006 [11], is a fully shift-invariant, multi-scale and multi-directional transform scheme, which deals well with fusion calculation. It captures edges and textures of the input image, fit with rich details and multi-sensor image orientation information, meanwhile eliminates the ringing effect and shaking phenomenon. Yin et al. [12] proposed an infrared and visible image fusion method using NSCT and fuzzy logic, where a degree of membership to the target and the background for the pixel at its location was presented. Wang et al. [13] proposed a NSCT and sparse representation based image fusion method (NSCTSR), and this dictionary learning scheme could extract salient features of images effectively. Other more methods based on NSCT become an emerging force on the way of infrared and visible image fusion issues.

In this article, we focus on the fusion process of infrared and visible images, propose an advanced NSCT based IR-visible image fusion algorithm employing adaptive Gaussian (AG) method and compressed sensing (CS) method, where the total variation (TV) based gradient descent recovery algorithm is used to recover the compressively sampled signal. The Gaussian fuzzy membership function can redistribute the pixel weight adaptively according to the different characteristics of the infrared or visible images. The compressed sensing, proposed by Candes and Donoho [14,15], is indicated that signals or images can be sparsely represented and accurately reconstructed through far fewer measurements of data, which can be used to reduce the redundant information or even denoise the signal to a certain extent.

The entire proposed method in this paper can be named NSCT-AG-CSTV approach. Experimental implementations in images and computation results illustrate the effectiveness and robustness of the proposed approach. Comparative analysis has also been conducted among the wavelet-based methods, contourlet-based methods and traditional NSCT-based methods.

The rest of this article is organized as follows. Section 2 gives a brief introduction of the image fusion scheme, and presents the image fusion procedure. Section 3 describes the compressed sensing theory and gradient based recovery theory. Section 4 presents the specific image fusion method using adaptive-Gaussian fuzzy membership, sparse sampling and gradient descent iterative recovery algorithm. Section 5 provides different evaluation methods for the experimental applications. Section 6 provides the experimental and comparable results, and discusses the performance of various fusion methods. Section 7 presents the conclusions of this article.

2. Image fusion scheme

2.1. Image fusion scheme based on NSCT

The NSCT method is established upon the combination of the structures of a non-subsampled pyramid (NSP) and a non-subsampled directional filter bank (NSDFB). The specific diagrams of the NSCT transform can be illustrated as shown in Fig. 1 as the filter bank structure and in Fig. 2 as the idealized frequency partitioning structure, respectively.

The non-subsampled multiscale subband decomposition is obtained from the NSP through a shift-invariant filtering process, which captures point discontinuities but seems similar to that of the Laplacian Pyramid. After NSP, the source image is decomposed into lowpass and bandpass images.

The non-subsampled multiscale directional decomposition is derived from the NSDFB, which links point discontinuities into linear structures, while abandoning the downsamplers and

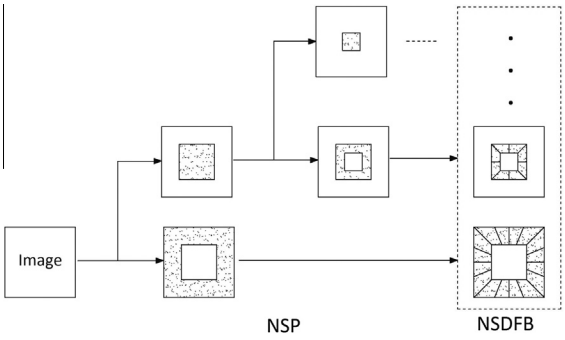

Fig. 1. NSCT: Non-subsampled filter bank structure.

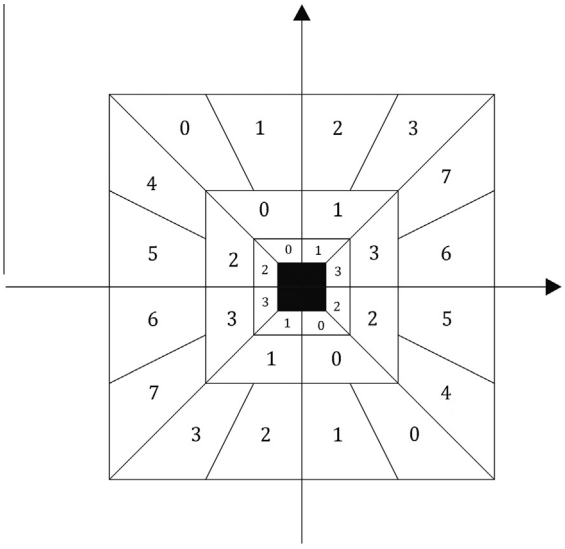

Fig. 2. NSCT: idealized frequency partitioning.

upsamplers. After NSDFB, the source image is decomposed into directions in subbands.

For each image, throughout the J levels NSCT decomposition calculation, 1 low-frequency subband coefficient image and $\sum_{j=1}^{J} 2^{l_j}$ high-frequency bandpass subband coefficient images can be obtained, where l_j is the directional decomposition series l of the jth scale.

2.2. Diagrams of fusion procedures

The fundamental fusion framework based on NSCT for the grayscale image fusion process is described in Fig. 3. The low-frequency and high-frequency coefficients images are produced for image A and image B after NSCT, respectively. In the fusion processing module, the low-frequency and high-frequency coefficients will be fused separately to obtain the fused low-frequency and high-frequency coefficients. Finally, the fused image can be obtained via the inverse NSCT from the fused coefficients.

In this article, an IR-visible image fusion scheme based on adaptive Gaussian method and compressed sensing technique is proposed. Fig. 4 presents the entire fusion procedure. First, the IR image is pre-enhanced, then the pre-enhanced IR image and the visible image will be decomposed using NSCT to produce low-frequency and high-frequency subbands respectively. The low-frequency coefficients from IR and visible image are fused using the adaptive regional average energy rule. The high-frequency coefficients from the two types of images will first be compressively sampled to produce sparse signals, then the maximum absolute selection rule will be employed to fuse the highest-frequency coefficients, the other high-frequency coefficients will be correspondingly fused using the adaptive-Gaussian regional standard deviation rule to produce sparse fused signals, followed by the application of total variation based gradient descent iterative recovery to transfer the sparse signals to general signals (high-frequency fused coefficients). In the end, the low-frequency fused coefficients and high-frequency fused coefficients will be processed by the inverse-NSCT to obtain the final fused image.

3. Compressed sensing and signal recovery

3.1. Principles of compressed sensing

As it is known that, the compressed sensing theory elaborates a truth that the source signal can be reconstructed precisely although it is sparsely represented way less than the number of observations acquired by the traditional Nyquist sampling theorem on the minimum rate of sampling.

In the CS principles, if considering a N-dimensional real signal $\mathbf{x} \in \mathbf{R}^{N\times 1}$ extended to be a superposition of spikes in the orthonormal basis $\mathbf{\Psi} = \{\boldsymbol{\psi}_i\}_{i=1}^{N}$ ($\boldsymbol{\psi}_i$ is a N-dimensional column vector), we can obviously have

$$\mathbf{x} = \mathbf{\Psi}\boldsymbol{\alpha} \tag{1}$$

where $\mathbf{\Psi}$ denotes the dictionary matrix and $\boldsymbol{\alpha}$ denotes the coefficients vector of an image in a tight-frame under the dictionary matrix.

It is assumed that $\boldsymbol{\alpha}$ is a k-sparse signal, which means that the number of non-zero coefficients keeps a relationship as $k \ll N$, thus we can sparsely observe the signal employing an observation matrix $\mathbf{\Phi} \in \mathbf{R}^{M\times N}$ incoherent with $\mathbf{\Psi}$, where $k < M \ll N$, and the

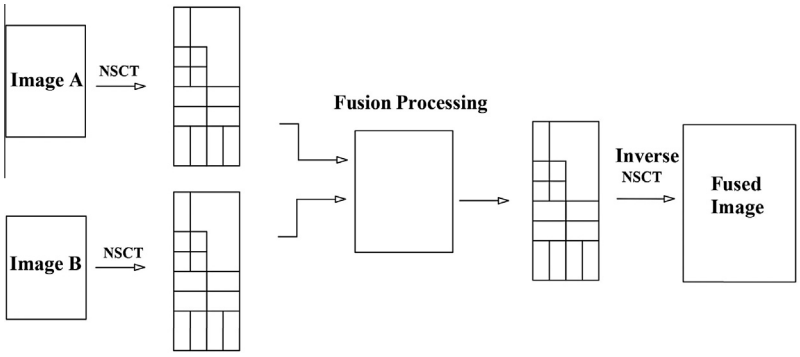

Fig. 3. General IR-visible image fusion scheme.

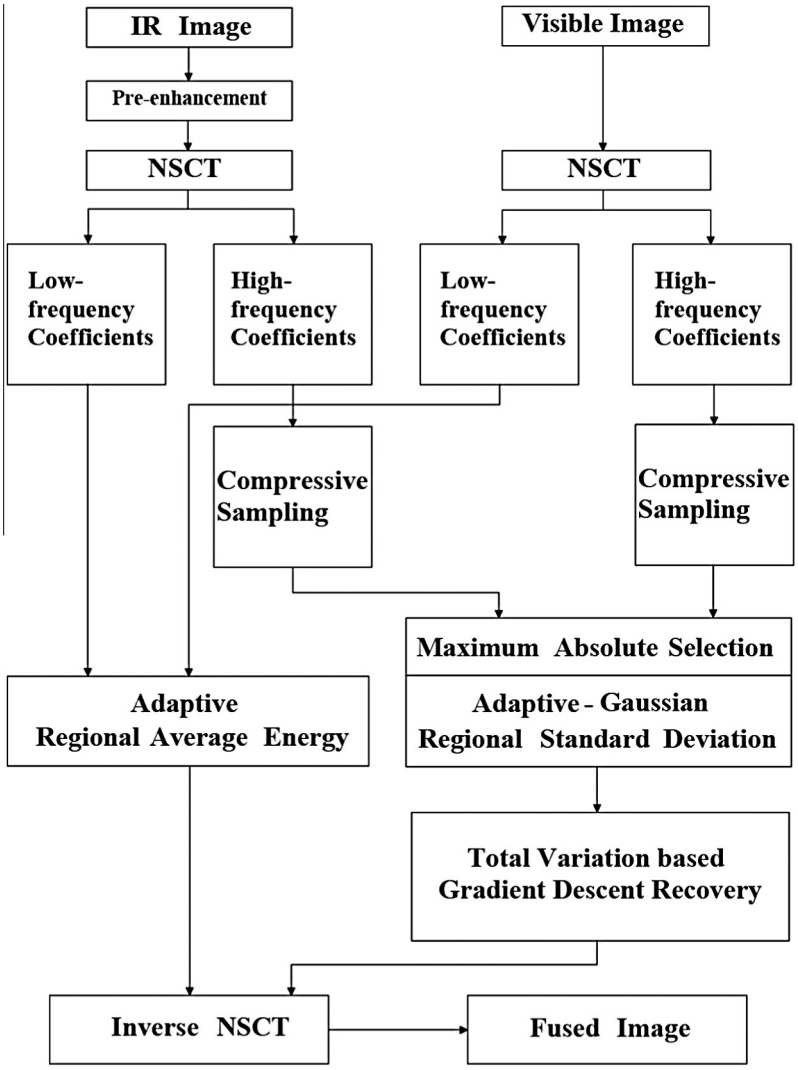

Fig. 4. Proposed NSCT-AG-CSTV IR-visible image fusion scheme.

linearly observed value is recorded to be $\mathbf{y} \in \mathbf{R}^{M\times 1}$. The equation of relationships can be described as

$$\mathbf{y} = \mathbf{\Phi x} = \mathbf{\Phi \Psi \alpha} = \mathbf{M}_0 \boldsymbol{\alpha} \tag{2}$$

in which, $\mathbf{M}_0$ is the sensing matrix established by the observation matrix and the dictionary matrix. The process of the sparsity measurement can be demonstrated as in Fig. 5.

The values in sensing matrix contain significant information for the reconstruction of the source signal. Due to the size of measurement maintaining $M \ll N$, to resolve $\boldsymbol{\alpha}$ from $\mathbf{y}$ is to pursue

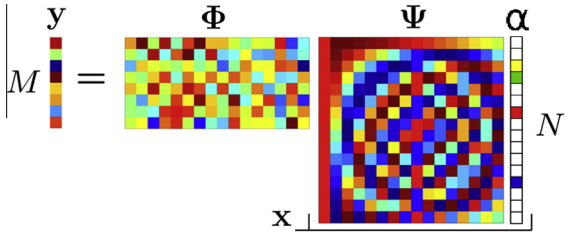

Fig. 5. The measurement scheme of compressed sensing.

a solution of an under-determined equation, which is a NP-hard problem. Nevertheless, the source signal/image $\mathbf{x}$ can be reconstructed highly accurately via solving a non-linear optimization problem, provided that the observation matrix $\mathbf{\Phi}$ obeys some necessary conditions [14,15], such as the restricted isometry property (RIP) and the uniform uncertainty principle (UUP).

3.2. Gradient descent reconstruction algorithm using total variation

To reconstruct the compressed sampling signal, traditional gradient-based recovery algorithms using total variation (TV) have been studied by related work in [16–19].

Here in this paper, to convert the convex optimization problem of traditional l_2-norm gradient descent based algorithm, the constraints of l_2-norm and l_1-norm are composed together to generate a gradient descent based Lagrange multiplier equation, where the total variation theory is employed into the l_1-norm signal optimization part to help resolve the modified convex optimization issue. Therefore, a finally improved gradient descent and total variation based reconstruction algorithm can be implemented.

Considering the linear matrix equation

$$\mathbf{M}_0\mathbf{f} = \mathbf{y} \tag{3}$$

the solution of sparse signal $\mathbf{f}$ can be achieved by the minimization problem as follows:

$$\min_f L(\mathbf{f}) = \frac{1}{2}\|\mathbf{M}_0\mathbf{f} - \mathbf{y}\|_2^2 \tag{4}$$

Taking into account of the l_1 norm strategy, this convex optimization issue can be converted into the form of

$$\min_f H(\mathbf{f}) = L(\mathbf{f}) + \lambda\|\mathbf{f}\|_1 \tag{5}$$

where λ is the Lagrange parameter, $\|\mathbf{f}\|_1 = \sum_{s,t}|f_{s,t}|$ is the l_1 norm for a general $N \times N$ image.

As for a general image, the aforementioned l_1 norm solution is not sparse in the time domain. Thus, the total variation strategy is applied for the sparse sampling case, and the optimization problem can be improved to be

$$\min_f H(\mathbf{f}) = L(\mathbf{f}) + \lambda TV(\mathbf{f}) \tag{6}$$

Here, $TV(\mathbf{f})$ means the total variation of the image $\mathbf{f}$. The solution of $TV(\cdot)$ can be solved by the summation of the gradient magnitude of each pixel, which can be expressed as

$$TV(\mathbf{f}) = \sum_{s,t}\sqrt{\left(D_{s,t}^v f\right)^2 + \left(D_{s,t}^h f\right)^2} = \sum_{s,t}|\nabla_{s,t} f| \tag{7}$$

where the derivatives of $\mathbf{f}$ along the vertical and horizontal directions can be given by

$$D_{s,t}^v f = f_{s,t} - f_{s+1,t} \ (1 \leqslant s < N) \tag{8}$$

$$D_{s,t}^h f = f_{s,t} - f_{s,t+1} \ (1 \leqslant t < N) \tag{9}$$

Eventually, the gradient of TV can be calculated by

$$\nabla_{s,t}(TV(\mathbf{f})) = \frac{D_{s,t}^h f}{|\nabla_{s,t} f|} - \frac{D_{s,t-1}^h f}{|\nabla_{s,t-1} f|} + \frac{D_{s,t}^v f}{|\nabla_{s,t} f|} - \frac{D_{s-1,t}^v f}{|\nabla_{s-1,t} f|} \tag{10}$$

4. NSCT-based adaptive image fusion algorithm

4.1. General fusion process

The entire fusion approach proposed in this article is assumed under the condition that the source images have already been strictly registered before the fusion process. Here we set the infrared and visible input images as **A** and **B** respectively, and the final fused image is set to be **F**.

(a) Decompose **A** and **B**, respectively, into J levels NSCT. Therefore, we are able to acquire the NSCT coefficients described as $\left\{C_{j_0}^A(x,y), C_{j,l}^A(x,y)\right\}$ and $\left\{C_{j_0}^B(x,y), C_{j,l}^B(x,y)\right\}$, where $j \geqslant j_0, C_{j_0}(x,y)$ refers to the low-frequency subband coefficient, and $C_{j,l}(x,y)$ denotes the high-frequency coefficient of direction l under scale j. There are J scales in the high-frequency decomposition.

(b) Adopt the corresponding image fusion rules to fuse the low-frequency subbands and high-frequency subbands, respectively, and thus obtain all of the NSCT coefficients for the inverse-NSCT process, where the fused coefficients can be given as $\left\{C_{j_0}^F(x,y), C_{j,l}^F(x,y)(j \geqslant j_0)\right\}$.

(c) Do the inverse-NSCT transform employing low-frequency and high-frequency fused coefficients to acquire the final fused image **F**.

The low-frequency subbands reflect the approximate characteristics of the source image, because it aggregates most of the information and energy of the image, and can lead to selecting the fusion coefficients guided by the regional energy index. The high-frequency coefficients represent the corresponding details of an image, such as edge details, contours and curves, and several research studies [20–22] have been conducted on the fusion of high-frequency image fusion.

The main purpose, or in other words, the meaning of fusion, is to perserve or intensify the characteristics between the foreground targets and the background details. Regarding the issue of infrared-visible image fusion, even though the low-frequency subbands and high-frequency subbands can be straightforwardly fused separately via different fusion rules, the determination of whether a pixel or a region belongs to the foreground targets or the background details is a pervasive problem of uncertainty. However, fuzzy logic [23,24] provides a path of admitting intermediate states assuming that each status remains with a degree of truth ranging from 0 to 1, making it reasonable in dealing with uncertainty problems and controlling the processing conditions.

Before the fusion rules applied, we can pre-enhance the infrared image in order to obtain better results in the process of subband coefficients fusion. Besides, herein we propose a series of specific fusion methods for the low-frequency subbands case and the high-frequency subbands case, respectively.

For the low-frequency coefficients fusion, the adaptive regional energy rule is used, because this fusion rule is able to keep detailed information better from the low-frequency decomposition. The high-frequency coefficients have an impact on edges and contours of the image, thus the selection of coefficients from a source image is more significantly influential on the human visual quality of the fused image. Empirically, the classical maximum absolute selection fusion rule is chosen for the highest-frequency coefficients fusion, because larger absolute values usually correspond to sharper brightness changes and thus to the salient features. For the other high-frequency coefficients fusion, the adaptive Gaussian fuzzy membership function is utilized to better balance the assignment of weights of high-frequency coefficients. Inside the fusion, the compressed sensing (CS) technique has been proposed to compressively sample the signals from the infrared and visible input coefficients. The high-frequency signal has a lot of redundancy, and it can be sparsely sampled to filter distortions or singularities to some extent. After the fusion of CS-applied high-frequency coefficients, the signal will be recovered by the total variation (TV) based nonlinear conjugate gradient descent algorithm with backtracking line search method, and the high-frequency fused coefficients will be obtained from the recovery.

4.2. Pre-enhancement of IR image

Before fusing the IR image and visible image, we can apply the S-function [25] to enhance the contrast of IR image, in order that we can obtain much better fusion effect. The adaptive contrast stretching S-function can be defined as:

$$S(i,j) = \frac{255}{k_A * \exp\left[-10 * A(i,j)/255 - 5\right] + 1}$$

$$k_A = \frac{255/\mu - 1}{\exp\left[-(10 * \mu/255 - 5)\right]} \tag{11}$$

where $A(i,j)$ is the pixel gray value of the IR image, μ is the mean pixel value of the IR image, $S(i,j)$ is the pixel gray value after contrast stretching, and k_A is the inflection point parameter. From this function, it can be known that the pixel values larger than μ will be enlarged and the smaller pixels will be decreased, which can effectively improve the contrast of the IR image.

4.3. Adaptive Gaussian fuzzy membership

We introduce an adaptive Gaussian fuzzy membership function into the high-frequency coefficients fusion. This function was presented by Yin et al. [12] to distribute the degree of membership respectively to the background and the target for a specific pixel location (i,j). Considering the translation invariance of NSCT transform, the frequency coefficients can be assigned to calculate the degree of membership to adjust the weights between infrared image decomposition coefficients and visible image decomposition coefficients.

Here, the adaptive Gaussian membership can be described as:

$$\eta_B(x,y) = \exp\left[-\frac{1}{2}\left(\frac{C_j^A(x,y)-\mu}{k\sigma}\right)^2\right] \tag{12}$$

$$\eta_A(x,y) = 1-\eta_B(x,y) \tag{13}$$

where there is an empirical value of adaptivity $k \in [1,\ 5], C_j^A(x,y)$ means the frequency subband coefficient of source image A, μ and σ are image A's mean value and standard deviation, respectively.

4.4. Low-frequency coefficients fusion

The adaptive weight rule based on regional average energy for the low-frequency subbands fusion is chosen here in the approach. The regional energy of the low-frequency coefficients for images **A** and **B** can be represented as follows:

$$E_{j_0}^A(x,y) = \sum_{m\in K,n\in L} w(m,n)\left[C_{j_0}^A(x+m,y+n)\right]^2 \tag{14}$$

$$E_{j_0}^B(x,y) = \sum_{m\in K,n\in L} w(m,n)\left[C_{j_0}^B(x+m,y+n)\right]^2 \tag{15}$$

where w is the regional window mask, its size $K\times L$ can be chosen as 3×3 or 5×5; here we choose a 3×3 mask as $w=[1,2,1;\ 2,3,2;\ 1,2,1]/15$.

The weight factors are defined as:

$$W_{j_0}^A = \frac{E_{j_0}^A(x,y)}{E_{j_0}^A(x,y)+k_0E_{j_0}^B(x,y)} \tag{16}$$

$$W_{j_0}^B = 1-W_{j_0}^A \tag{17}$$

where k_0 is a parameter for the adjustment of the weight factors, thus the low-frequency coefficients after fusion can be written as:

$$C_{j_0}^F(x,y) = W_{j_0}^A C_{j_0}^A(x,y) + W_{j_0}^B C_{j_0}^B(x,y) \tag{18}$$

4.5. High-frequency coefficients fusion

In the high-frequency subband decomposition, the lower level reflects coarser information, and we choose the regional standard deviation fusion rule. The highest level reflects more detailed information, which indicates more independency, and we choose the absolute maximum selection rule.

The fusion coefficients for the highest decomposition scale J can be written as follows:

$$C_{J,l}^F(x,y) = \begin{cases} C_{J,l}^A(x,y), \left|C_{J,l}^A(x,y)\right| \geqslant \left|C_{J,l}^B(x,y)\right| \\ C_{J,l}^B(x,y), \left|C_{J,l}^A(x,y)\right| < \left|C_{J,l}^B(x,y)\right| \end{cases} \tag{19}$$

4.5.1. CS-applied high-frequency coefficients fusion

For the other high-frequency scales, we calculate the weighting factors using the regional standard deviation to form the subband coefficients. The standard deviation is a significant indicator for the image fusion and contrast enhancement; here we also choose the window template with 3×3.

In accordance with the definition of standard deviation, we can express the regional standard deviation as follows:

$$\sigma_{j,l}^A(x,y) = \sqrt{\frac{1}{KL}\sum_{m\in K}\sum_{n\in L}\left[C_{j,l}^A(x+m,y+n)-\overline{C_{j,l}^A}\right]^2} \tag{20}$$

$$\sigma_{j,l}^B(x,y) = \sqrt{\frac{1}{KL}\sum_{m\in K}\sum_{n\in L}\left[C_{j,l}^B(x+m,y+n)-\overline{C_{j,l}^B}\right]^2} \tag{21}$$

If we define the weighting factors as:

$$W_{j,l}^A = \frac{\sigma_{j,l}^A(x,y)}{\sigma_{j,l}^A(x,y)+k_l\sigma_{j,l}^B(x,y)} \tag{22}$$

$$W_{j,l}^B = 1-W_{j,l}^A \tag{23}$$

where k_l is also an adjustment parameter for the weight factors, the general high-frequency coefficients after fusion can be expressed as:

$$C_{j,l}^F(x,y) = W_{j,l}^A C_{j,l}^A(x,y) + W_{j,l}^B C_{j,l}^B(x,y) \tag{24}$$

If taking Gaussian self-adaptivity into consideration, the high-frequency weighting factors after adaptive fusion can be presented as:

$$W_{j,l}^A = \frac{\eta_A(x,y)\sigma_{j,l}^A(x,y)}{\eta_A(x,y)\sigma_{j,l}^A(x,y)+\eta_B(x,y)k_l\sigma_{j,l}^B(x,y)} \tag{25}$$

$$W_{j,l}^B = 1-W_{j,l}^A \tag{26}$$

where the fused coefficients can still have the corresponding form similar as Eq. (24).

4.5.2. Recovery algorithm for CS-applied signal

The fusion process for the most high-frequency coefficients is established on the signals after compressive sampling has been applied. The fused high-frequency coefficient sparse signal needs to be recovered to be the regular signal for the reconstruction of the final fusion image. Here, we basically employ the nonlinear conjugate gradient descent and line search algorithm [19] and improve it a bit to better obtain the recovery results catering for our issue.

Algorithm. Gradient Descent based Iterative Algorithm.

Inputs:
1: y : κ-space measurements
2: y_0: initial measurement value for iteration
3: λ: a weight threshold constant for tuning iterations
4: *eps*: convergence criterion of the gradient (default: 10^{-4})
5: *MaxIter*: convergence condition by number of iterations
6: α,β: line search parameters (defaults: $\alpha=0.05$ and $\beta=0.6$)
7: η: weight threshold shrinkage factor
8: k: general iteration number
9: t: line search parameter
10: m: line search numerical approximation parameter

Algorithm. Gradient Descent based Iterative Algorithm.

Outputs:
% Initialization
1: $k = 0; m = 0; t = 1; g_0 = \nabla f(m_0); \Delta m_0 = -g_0$
% Iterations
1: while ($\|g_k\|_2 > eps$ and $k < MaxIter$){
2: % Backtracking line search
1: while ($f(m_k + t\Delta m_k) > f(m_k) + \alpha t \Re(g_k^* \Delta m_k)$)
2: $\{t = \beta t; m = m + 1\}$
3: $m_{k+1} = m_k + t\Delta m_k$
4: $g_{k+1} = \nabla f(m_{k+1})$
5: $\gamma = \frac{\|g_{k+1}\|_2^2}{\|g_k\|_2^2}$
6: $\Delta m_{k+1} = -g_{k+1} + \gamma \Delta m_k$
7: $k = k + 1; \lambda = \lambda\eta$

After the recovery processing, the recovered signal stemming from the input measurement signal will be the fused high-frequency coefficients for the further step, the NSCT reconstruction. When all the fused coefficients for NSCT reconstruction, including low-frequency and high-frequency fused coefficients, are obtained, the final fused image **F** can be resolved through the inverse-NSCT processing.

5. Fusion evaluation

To better evaluate the effectiveness of the fusion approach, we apply some indexes, such as the global standard deviation, the Shannon entropy, the root-mean-square error, the mutual information, and the edge-based similarity measurement, to quantitatively assess the results and qualities of the proposed algorithm.

5.1. Global standard deviation

In the assessment of the quality of an image, the standard deviation reflects how much the variation or dispersion of the grayscale of image pixels differs from the average value. A high standard deviation indicates that the fusion data points are spread out over a large range of values, and so is the fusion effect. Deviation generally reflects the influence of the high frequency in an image, and the value is usually proportional to the image contrast effect. The standard deviation is defined as:

$$\sigma = \sqrt{\frac{1}{MN}\sum_{x\in M}\sum_{y\in N}\left[F(x,y) - \overline{F}\right]^2} \tag{27}$$

where $\overline{F}$ is the average grayscale value of the final fused image. The larger quantity of standard deviation, the better image quality.

5.2. Shannon information entropy

The information entropy, also known as Shannon entropy, quantifies the expected value of the information contained in an image. The definition of Shannon entropy is expressed as follows:

$$H = -\sum_{g=0}^{G-1} P_g \log P_g \tag{28}$$

where P_g is the distribution probability of the grayscale g, which maintains a relationship of $g \in [0, 1, \ldots, G-1]$. Similar to the standard deviation, an image with bigger entropy usually has better quality.

5.3. Root mean square error

The root-mean-square error (RMSE) is a measure of the differences between values predicted by a model and the values actually observed. It can be used to evaluate the difference of cumulative squared error between two images or measure the similarity of two images.

RMSE can be obtained by the equation written as

$$\mathrm{RMSE} = \sqrt{\frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[y(i,j) - \widehat{y}(i,j)\right]^2} \tag{29}$$

where y and $\widehat{y}$ denote the infrared image and the fused image, respectively. Generally, the larger calculation value of RMSE means the better image fusion effect.

5.4. Mutual information

The mutual information (MI) [26] between fusion image **F** and the source images **A** and **B** can be given by

$$\mathrm{MI} = \mathrm{MI}^{AF} + \mathrm{MI}^{BF} \tag{30}$$

in which

$$\mathrm{MI}^{AF} = \sum_{f=0}^{L}\sum_{a=0}^{L} p^{AF}(a,f)\log_2\left(\frac{p^{AF}(a,f)}{p^A(a)p^F(f)}\right) \tag{31}$$

$$\mathrm{MI}^{BF} = \sum_{f=0}^{L}\sum_{b=0}^{L} p^{BF}(b,f)\log_2\left(\frac{p^{BF}(b,f)}{p^B(b)p^F(f)}\right) \tag{32}$$

Here, MI^{AF} and MI^{BF} respectively denote the normalized MI between the fused image and the source image. $p^A(a), p^B(b)$ and $p^F(f)$ are the normalized gray level histograms of source images and the fused image; a, b and $f \in [0, L]$. $p^{AF}(a,f)$ and $p^{BF}(b,f)$ are the joint gray level histograms between the fused image and the source images. MI is able to indicate how much information the fused image **F** can convey from the source images, and the greater value, the better fusion effect.

5.5. Edge based similarity metrics

The edge-based similarity metrics ($Q^{AB/F}$), proposed by Xydeas and Petrovic [27], demonstrates the similarity between the edges transferred from the source images to the fused image. The definition equation can be defined as follows:

$$Q^{AB/F} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left[Q^{AF}(i,j)w^A(i,j) + Q^{BF}(i,j)w^B(i,j)\right]}{\sum_{i=1}^{M}\sum_{j=1}^{N}\left[w^A(i,j) + w^B(i,j)\right]} \tag{33}$$

Here, $Q^{AF}(i,j)$ and $Q^{BF}(i,j)$ are edge preservation values for the strength and orientation, which are weighted by coefficients $w^A(i,j)$ and $w^B(i,j)$. $w^A(i,j)$ and $w^B(i,j)$ reflect the perceptual significance of the corresponding edge elements within source images, and can be calculated via the gradient strengths. Thus, these parameters can be expressed as:

$$w^A(i,j) = \left|\sqrt{s_i^A(i,j) + s_j^A(i,j)}\right|^L \tag{34}$$

$$w^B(i,j) = \left|\sqrt{s_i^B(i,j) + s_j^B(i,j)}\right|^L \tag{35}$$

$$Q^{AF}(i,j) = Q_a^{AF}(i,j)Q_g^{AF}(i,j) \tag{36}$$

$$Q^{BF}(i,j) = Q_a^{BF}(i,j)Q_g^{BF}(i,j) \tag{37}$$

where $s_i(i,j)$ and $s_j(i,j)$ respectively denote the horizontal and vertical Sobel edge detectors at a pixel for a source image. $Q_a^{xF}(i,j)$ and

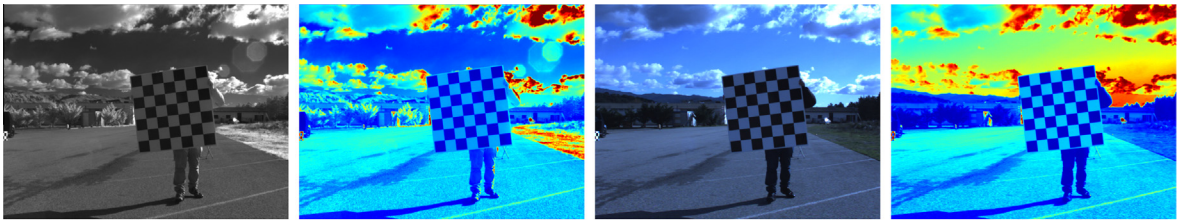

Fig. 6. Source images. (a) IR image. (b) IR image in HSV space. (c) Visible image. (d) Visible image in HSV space.

$Q_g^{xF}(i,j)$ respectively denote the edge strength and orientation preservation values at a pixel location for a source image.

$Q^{AB/F}$ indicates how much strength and orientation information can be captured at each pixel among the source images and the fused image by a Sobel detector. Likewise, the larger value, the better effect of fusion.

6. Implementation and discussion

6.1. Experiments

A set of multi-focus data including infrared and visible images has been processed employing the proposed algorithm in this article.

The IR-visible data set with image dimension of $M \times N = 768 \times 1024$ has been chosen to make the experiments implemented. The data set was acquired from the field experiments in Corte, France in November 2013, by the Stereovision system [28] (University of Corsica, France), which was composed of two industrial multispectral near-infrared/visible cameras, permitting the capture of NIR (700 nm $\sim$ 900 nm) and visible images with pixel resolution of 1024×768.

The WT (Wavelet transform) based fusion method, CT (Contourlet transform) based fusion method, general NSCT (Nonsubsampled Contourlet transform) based fusion method, and our proposed NSCT-AG-CSTV (NSCT, Adaptive Gaussian, Compressed Sensing with Total Variation) fusion method are respectively implemented to make analysis and comparisons. The experiments have been executed by MATLAB-2014 under CPU-3.1 GHz Windows 7 GUI system.

In the proposed NSCT-AG-CSTV fusion approach, the infrared image is pre-enhanced before the NSCT decomposition. For the low-frequency subbands fusion, the regional adaptive weighting rule is used; whereas, for the high-frequency subbands fuison, the absolute maximum selection rule and the adaptive-Gaussian maximum-deviation rule based on compressed sensing technology are respectively utilized for the fusion operation.

For our proposed hybrid fusion approach, we have conducted all of the decomposition experiments in 2^1–2^2 directions to take into account the balance of implementation results and computation speed.

Fig. 6 displays the input infrared and visible source images and their mapped images in HSV colorspace, respectively. Figs. 7 and 8 illustrate the fusion images when the adaptivity parameter maintains $k = 3$. Fig. 7 displays the fused image in grayscale and RGB colorscale, while their mapped HSV images are also listed, respectively. Before the final fused images are obtained, the fused coefficients of low-frequency and high-frequency subbands can be acquired. Fig. 8 exhibits the low-frequency coefficients image, the first-level and second-level fused high-frequency coefficients images. From the coefficients images, it is clear to see the directionality from the NSDFB decomposition.

6.2. Discussion

In addition to the experiments for the presented approach, as mentioned before, the experiments of wavelet transform based, contourlet transform based and nonsubsampled contourlet transform based fusion methods have also been conducted for comparison and analysis.

Table 1 presents the quantitative assessment for different fusion schemes in terms of varieties of evaluation metrics of standard deviation, Shannon entropy, root mean square error, mutual information and edge based similarity index.

From the quantitative values shown in the table, it is evident that the proposed NSCT-based Adaptive-Gaussian fusion approach using Compressed Sensing and Total Variation (NSCT-AG-CSTV) generally has much better quality and robustness than the traditional fusion methods on the basis of wavelet transform, contourlet transform, nonsubsampled contourlet transform, etc.

One also notes that the proposed method evidently has much better Standard Deviation, RMSE and Shannon Entropy values than the other three conventional methods. But comparing with the typical NSCT-based method, the proposed scheme does not have better calculation performance in *MI* and $Q^{AB/F}$. Because the compressed sensing (CS) method has been introduced to the entire approach, it compressively samples the high-frequency coefficients in a certain sampling rate, thus has to lose some information of edges and curves. Here in this experiment, we choose the random sampling rate as 45%. The higher sampling rate, the larger *MI* and $Q^{AB/F}$ approximate to traditional NSCT-based scheme's. But, if we choose much higher sampling rate, such as more than 80%, the

Fig. 7. Fused images. (a) Fused image in grayscale. (b) Grayscale fused image in HSV space. (c) Fused image in colorscale. (d) Colorscale fused image in HSV space. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

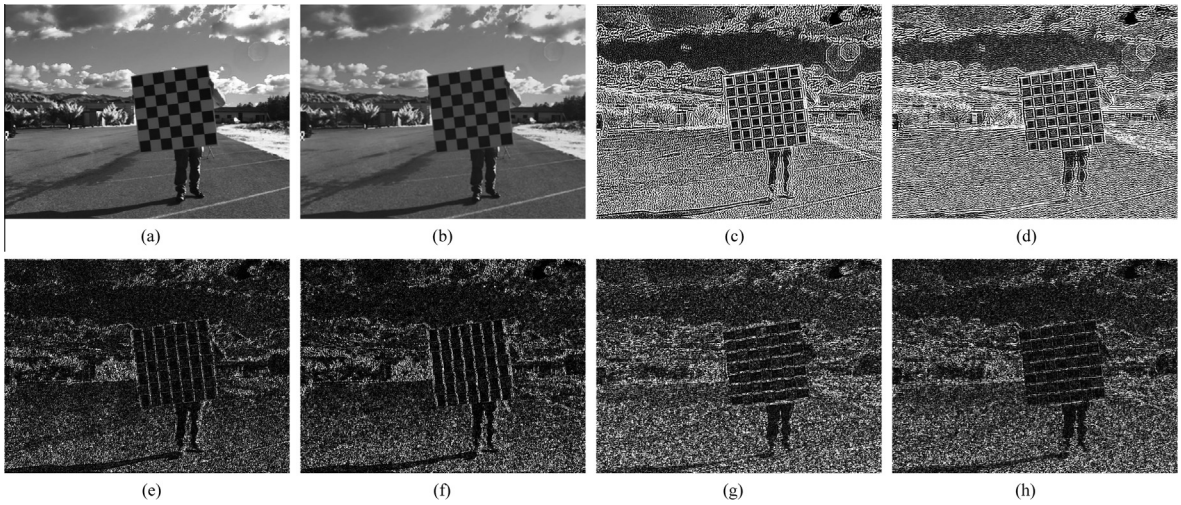

Fig. 8. NSCT-AG-CSTV fused coefficients images. (a) Fused image in grayscale. (b) Low-frequency coefficients fused image. (c and d) First-level high-frequency coefficients fused images. (e)–(h) Second-level high-frequency coefficients fused images.

Table 1
Performance of different fusion methods.

	SD	*H*	*RMSE*	*MI*	$Q^{AB/F}$
IR image	41.1175	6.9767	N/A	N/A	N/A
Visible image	53.2545	7.3542	N/A	N/A	N/A
WT	42.2954	7.0014	23.4291	5.1348	0.6964
CT	45.7947	7.2814	34.0181	5.3615	0.5735
NSCT	45.8153	7.2610	34.1048	6.1301	0.6970
NSCT-AG-CSTV ($k = 1$)	55.5205	7.5039	40.0403	5.8287	0.6031
NSCT-AG-CSTV ($k = 2$)	55.5204	7.5039	40.0413	5.8292	0.6039
NSCT-AG-CSTV ($k = 3$)	55.5210	7.5040	40.0427	5.8258	0.6045

CS method will become meaningless to the approach, and the redundancy will be more with more computation time. The CS technique can effectively discard some redundant or noisy information, in order that the more valuable information can be better depicted through the deployment.

On account of the complexity of the proposed approach, it costs about 290s to obtain the grayscale fused image and calculate all the performance results, whereas, the general NSCT-based fusion approach only costs about 95s in comparison. Besides, the adaptivity value k can be usually adopted as 1, 2 and 3, where the best case is when $k = 3$ has been chosen. When $k > 3$, the performance will be decreased gradually along with the increase of the value of k. That is to say, the NSCT-AG-CSTV ($k = 3$) basically has the best performance results than the other compared fusion methods.

Some other data sets have also been experimented using the presented approach. All the implementations prove the superiority of the proposed NSCT-AG-CSTV scheme.

7. Conclusion

An advanced hybrid approach for the IR-visible image fusion based on NSCT, adaptive Gaussian method and compressed sensing technique has been proposed.

There are some advantages from the presented approach: (1) NSCT transform is selected to decompose and reconstruct the coefficients, due to its superiority of shift-invariance, multi-resolution, multi-direction, etc. (2) The pre-enhancement of IR source image can promote the contrast of the IR image, which benefits the following fusion effect. (3) The adaptive regional average energy fusion rule is employed to fuse the low-frequency coefficients, because the low-frequency subbands contain most information from source images, and this fusion rule has the outstanding performance to keep and reflect the low-frequency information. (4) The highest-frequency coefficients fusion usually adopts the maximum absolute selection rule, but the other high-frequency coefficients convey information in the fused image evidently according to different selected fusion rules. We choose to compressively sample these high-frequency coefficients, from IR and visible source images, fuse them using a proposed adaptive-Gaussian regional standard deviation rule, and recover the sparse fused coefficients to be regular coefficients employing an improved total variation based gradient descent iterative recovery algorithm.

The entire proposed fusion approach can be named NSCT-AG-CSTV. This approach is capable of increasing complementary information of fusion, strengthening selected targets or defects, improving image contrast and details, etc. Experiments and calculations prove the advantages of our proposed approach, in comparison of standard deviation, Shannon entropy, root-mean-square error, mutual information and edge-based similarity metrics with the wavelet transform based fusion method, contourlet transform based fusion method and general nonsubsampled contourlet transform based fusion method. Human visual perception also indicates that the proposed approach is advantageous to contain and convey all the information from source images.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

The authors would like to appreciate the support of data set from Mr. Tom Toulouse at the University of Corsica, France, and are grateful of the partial support obtained from the Canada Research Chair program, Chair MIVIM (Multipolar Infrared Vision Infrarouge Multipolaire) and from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] C. Pohl, J.L.V. Genderen, Multisensor image fusion in remote sensing: concepts, methods and applications 19 (1998) 823–854.
[2] G. Piella, A general framework for multiresolution image fusion: from pixels to regions, Inf. Fus. 4 (4) (2003) 259–280. doi:http://dx.doi.org/10.1016/S1566-2535(03)00046-0.

[3] Y. Yang, S. Tong, S. Huang, P. Lin, Multifocus image fusion based on NSCT and focused area detection, Sens. J., IEEE 15 (5) (2015) 2824–2838, http://dx.doi.org/10.1109/JSEN.2014.2380153.
[4] P. Burt, E. Adelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (4) (1983) 532–540, http://dx.doi.org/10.1109/TCOM.1983.1095851.
[5] S.K. Rogers, C.W. Tong, M. Kabrisky, J.P. Mills, Multisensor fusion of ladar and passive infrared imagery for target segmentation, Optical Eng. 28 (8) (1989) 288881, http://dx.doi.org/10.1117/12.7977051.
[6] B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, I. Pippi, Multispectral fusion of multisensor image data by the generalized Laplacian pyramid, in: Proceedings of the IEEE 1999 International Geoscience and Remote Sensing Symposium, 1999, IGARSS '99, vol. 2, 1999, pp. 1183–1185, doi:http://dx.doi.org/10.1109/IGARSS.1999.774572.
[7] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 674–693, http://dx.doi.org/10.1109/34.192463.
[8] T.L. Huntsberger, B.D. Jawerth, Wavelet-based sensor fusion, 1993, doi:http://dx.doi.org/10.1117/12.150252.
[9] H. Li, B. Manjunath, S. Mitra, Multi-sensor image fusion using the wavelet transform, in: Proceedings of the IEEE International Conference on Image Processing, 1994, ICIP-94, vol. 1, 1994, pp. 51–55, doi:http://dx.doi.org/10.1109/ICIP.1994.413273.
[10] M. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (12) (2005) 2091–2106, http://dx.doi.org/10.1109/TIP.2005.859376.
[11] A. da Cunha, J. Zhou, M. Do, The nonsubsampled contourlet transform: theory, design, and applications, IEEE Trans. Image Process. 15 (10) (2006) 3089–3101, http://dx.doi.org/10.1109/TIP.2006.877507.
[12] S. Yin, L. Cao, Q. Tan, G. Jin, Infrared and visible image fusion based on NSCT and fuzzy logic, in: 2010 International Conference on Mechatronics and Automation (ICMA), 2010, pp. 671–675, doi:http://dx.doi.org/10.1109/ICMA.2010.5588318.
[13] J. Wang, J. Peng, X. Feng, G. He, J. Wu, K. Yan, Image fusion with nonsubsampled contourlet transform and sparse representation, J. Electron. Imag. 22 (4) (2013) 043019, http://dx.doi.org/10.1117/1.JEI.22.4.043019.
[14] E.J. Candes, Compressive sampling, in: Proceedings of the International Congress of Mathematicians, Madrid, August 22–30, 2006 (Invited Lectures, 2006).
[15] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306.
[16] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inf. Theory 53 (12) (2007) 4655–4666.
[17] T. Blumensath, M.E. Davies, Gradient pursuits, IEEE Trans. Signal Process. 56 (6) (2008) 2370–2382.
[18] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D 60 (1–4) (1992) 259–268.
[19] M. Lustig, D. Donoho, J.M. Pauly, Sparse MRI: the application of compressed sensing for rapid mr imaging, Magnet. Reson. Med. 58 (6) (2007) 1182–1195.
[20] L. Tang, Z.-G. Zhao, A new method for infrared weak small target enhancement based on image fusion, in: Eighth International Workshop on Image Analysis for Multimedia Interactive Services, 2007, WIAMIS '07, 2007, pp. 22–22, doi:http://dx.doi.org/10.1109/WIAMIS.2007.7.
[21] X. Su-xia, L. Xiao-feng, C. Tian-hua, X. Hongbing, Image fusion method based on NSCT and robustness analysis, in: 2011 International Conference on Computer Distributed Control and Intelligent Environmental Monitoring (CDCIEM), 2011, pp. 346–349, doi:http://dx.doi.org/10.1109/CDCIEM.2011.511.
[22] Q. Zhang, B.-L. Guo, Research on image fusion based on the nonsubsampled contourlet transform, in: IEEE International Conference on Control and Automation, 2007, ICCA 2007, 2007, pp. 3239–3243, doi:http://dx.doi.org/10.1109/ICCA.2007.4376961.
[23] S.K. Pal, A note on the quantitative measure of image enhancement through fuzziness, IEEE Trans. Pattern Anal. Mach. Intell. PAMI-4 (2) (1982) 204–208, http://dx.doi.org/10.1109/TPAMI.1982.4767227.
[24] C.-W. Tao, W. Thompson, J. Taur, A fuzzy logic approach to multidimensional target tracking, in: Second IEEE International Conference on Fuzzy Systems, 1993, vol. 2, 1993, pp. 1350–1355, doi:http://dx.doi.org/10.1109/FUZZY.1993.327589.
[25] Z. Fu, X. Dai, Y. Li, H. Wu, X. Wang, An improved visible and infrared image fusion based on local energy and fuzzy logic, in: 2014 12th International Conference on Signal Processing (ICSP), 2014, pp. 861–865, doi:http://dx.doi.org/10.1109/ICOSP.2014.7015126.
[26] G. Qu, D. Zhang, P. Yan, Information measure for performance of image fusion, Electron. Lett. 38 (7) (2002) 313–315, http://dx.doi.org/10.1049/el:20020212.
[27] C. Xydeas, V. Petrovic, Objective image fusion performance measure, Electron. Lett. 36 (4) (2000) 308–309, http://dx.doi.org/10.1049/el:20000267.
[28] L. Rossi, T. Toulouse, M. Akhloufi, A. Pieri, Y. Tison, Estimation of spreading fire geometrical characteristics using near infrared stereovision, 2013, doi:http://dx.doi.org/10.1117/12.2001624.