IJAMML 15:1 (2021) 1-29 September 2021
ISSN: 2394-2258
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/ijamml_7100122215

A SPARSE REPRESENTATION METHOD BASED ON QUATERNION FOR MULTI-FOCUS IMAGE FUSION

Xiuming Sun[a], Weina Wu[b], Peng Geng[c] and Lin Lu[c]

[a]Zhangjiakou University, Zhangjiakou, P. R. China

[b]Shijiazhuang No.24 High School, Shijiazhuang, P. R. China

[c]Shijiazhuang Tiedao University, Shijiazhuang, P. R. China

Abstract

In order to achieve the multi-focus image fusion task, a sparse representation method based on quaternion for multi-focus image fusion is proposed in this paper. Firstly, the RGB color information of each pixel in the color image is represented by quaternion based on the relevant knowledge of computational mathematics, and the color image pixel is processed as a whole vector to maintain the relevant information between the three color channels. Secondly, the dictionary represented by quaternion and the sparse coefficient represented by quaternion are obtained by using the our proposed sparse representation model. Thirdly, the coefficient fusion is carried out by using the "max-L1" rule. Finally, the fused sparse coefficient and dictionary are used for image reconstruction to obtain the quaternion fused image, which is then converted into RGB color multi-focus fused image. Our method belongs to computational mathematics, and uses the relevant knowledge in

*Corresponding author.
E-mail address: sunxiuming@zjku.edu.cn (Xiuming Sun).

2020 Mathematics Subject Classification: 15-XX.
Submitted by Jianqiang Gao.
Received July 15, 2021

the field of computational mathematics to help us carry out the experiment. The experimental results show that the method has achieved good results in visual quality and objective evaluation.

Keywords: multi-focus image fusion, computational mathematics, sparse representation, quaternion, vector, sparse coefficient, max-L1.

---

1. Introduction

Image fusion refers to the technology that the image data about the same target collected by multi-source channels are processed by image processing and computer technology to extract the favourable information in each channel to the maximum extent, and finally integrated into a high-quality image. Image fusion, also known as multi-sensor image fusion, can make use of different information acquired from different sensors in a simultaneous interpreting scene, and obtain a more comprehensive, accurate and reliable fusion image through a certain fusion algorithm. It can overcome the limitation and difference of single sensor image in geometry, spectrum and spatial resolution, and at the same time improve the image definition and information packet content, and obtain the fusion result more consistent with human visual perception. As is known, it is often difficult to use a camera, such as a digital single-lens reflex camera, to get a full focus picture in which all objects or scenes are in focus at one shot. Generally speaking, only the objects in the focal distance of optical lens can retain sharp appearance information, and that beyond the focal distance will be blurred. Generally, a focused camera can only obtain the clear image of the target within its depth of field. The common method getting full focus images is the so called multi-focus image fusion that fuses multiple same scene observations taken with different focus distance into one full focus image, whose key thought is to preserve the information involved in each observation as complete as possible. In the development of multi-focus

image fusion, multi-focus image fusion algorithms can be generally divided into two categories, namely, transform domain method and spatial domain method.

The transform domain method mainly refers to the method based on multi-resolution (MR) analysis [1] and the method based on multi-scale geometric transformation. An obvious advantage of using MR method to deal with image fusion is that the fusion result accords with the perception of human vision. Wavelet transform is a kind of classical transform domain method. Not only has discrete wavelet transform (DWT) a good performance in maintaining spatial and frequency domain locality, but also it has high algorithm efficiency. Although DWT has the above advantages, its fusion result will be significantly worse if the source image is not strictly registered. In addition to DWT, Stationary Wavelet [2], Complex Wavelet [3], Double Tree Complex Wavelet [4], and other multi-resolution analysis theories have also been used in image fusion and achieved good results. In addition, some multi-scale geometric analysis theories developed in recent years have also been used in the study of image fusion, such as Curvelet [5], Contourlet [6], NSCT [7, 8], etc. Compared with the traditional wavelet transform, multi-scale geometric transform has the advantage of multi-direction and can obtain better fusion results. The NSCT-based method decomposes the image using contourlet or shearlet transform without subsampling [9], which outperforms the traditional discrete wavelet transform (DWT) based image fusion methods. Li and Yang [10] proposed a combination method to fuse multi-focus images by comprehensively utilizing the directionality of curvelet and the multi-resolution characteristics of wavelet, and achieved a good effect.

The spatial domain methods usually adopt a block-based fusion strategy, in which the source images are decomposed into blocks and each pair of block is fused with a designed activity level measurement like spatial frequency and summodified-Laplacian [11]. Since the size of the block has a great impact on the quality of the fusion, some improved methods are put forward to solve this problem [12, 14]. Beyond that,

image segmentation based fusion methods that rely on the accuracy of segmentation are presented in [15, 16]. In recent years, novel gradient information based methods [17, 18] have achieved impressive fusion results.

However, due to the large redundancy of multi-scale geometric analysis, it is easy to produce "information overload" in the process of fusion, and brings a large computational complexity. In recent years, signal sparse representation theory has been widely used in all kinds of image processing problems, and has achieved great success, which has become a new research direction. Some scholars have proposed an image fusion method based on sparse representation [19], and the fusion results are better than many methods. Yang and Li [19] proposed to use the absolute value of sparse coefficient to measure the focusing degree of each image sub block, which realized the fusion of multi-focus images, and achieved a good fusion effect compared with wavelet transform and Curvelet transform. Zhang and Levine [20] replaced the sliding window in the traditional sparse representation method by combining the neighbourhood sub block information of the image sub block, and used the multi-task robust sparse representation method to fuse the image. Generally speaking, the sparse representation method is better than the traditional fusion method based on multi-scale decomposition.

Most of the traditional image fusion methods focus on gray image fusion. Because the three color channels of a color image are closely related, the traditional gray image fusion method cannot be directly applied to the color image fusion. For the traditional color image fusion problem, the main methods adopted at present mainly include the fusion methods of RGB, IHS, YIQ, HSV and other spatial models. These methods often cut the correlation between the three color channels of the color image, making the fusion image appear a certain degree of color distortion. For example, Jin et al. [21] adopts HSV color model to conduct non-subsampled shearlet decomposition for H, S, and V channels, and uses different fusion rules to fuse the non-subsampled shearlet

coefficients of each channel respectively, so that the correlation of the three color channels of the color image is separated by the way of processing each channel respectively. Quaternion theory can process color image pixels as a whole vector. In the process of image fusion, each pixel of color image is represented by quaternion and processed as a whole vector, which can effectively maintain the correlation among the three color channels of color image. Zeng et al. [22] extended principal component analysis to quaternion principal component analysis and realized color image classification by using quaternion principal component analysis. Xu et al. [23] proposed a quaternion sparse representation method of color image based on quaternion singular value decomposition (QSVD) and quaternion orthogonal matching pursuit (QOMP) for image denoising, super-resolution reconstruction and image in painting.

Aiming at the disadvantages of traditional color image fusion, this paper proposes an image fusion method based on quaternion theory and sparse representation theory. We use a pure quaternion to model the three color components of the multi-focus image to be fused. Combined with the quaternion theory and sparse representation theory, we use quaternion to process the color image pixels as a whole vector, and use quaternion to represent the non-local similarity of the image to train the quaternion dictionary. Finally, the RGB color fusion image is reconstructed.

The rest of this paper is arranged as follows. The Section 2 introduces quaternion theory, sparse representation theory and sparse representation model based on quaternion. The Section 3 introduces the sparse representation method based quaternion for multi-focus image fusion. The Section 4 shows the experimental results of this method and compares with other methods. The Section 5 is the conclusion of this paper.

2. Related Works

2.1. Quaternion theory

Quaternions were first proposed by Hamilton in 1843 [24]. They are generalization of complex numbers. The concept of quaternions is widely used in many fields. A quaternion q is composed of a real part and three imaginary parts. The q is shown in formula (1).

$$q = a + bi + cj + dk, \tag{1}$$

where a, b, c and d are all real numbers. The real part of quaternions is a and the imaginary part is the combination of the remaining three parts. The i, j and k in these three parts are imaginary number unit, and they are orthogonal to each other and obey the rules as below:

$$i^2 = j^2 = k^2 = -1,\ ij = -ji = k,\ ki = -ik = j,\ jk = -kj = i. \tag{2}$$

If the real part $a = 0$, q is called pure quaternion. More detailed information about quaternion is introduced in [24] and [25].

The second representation of quaternions is amplitude phase representation, which can be expressed as:

$$q = |q|e^{i\phi}e^{i\theta}e^{k\psi}(\phi,\ \theta,\ \psi) \in [-\pi,\ \pi) \times [-\pi/2,\ \pi/2) \times [-\pi/4,\ \pi/4], \tag{3}$$

where $|q|$ is the amplitude of q, and ϕ, θ, ψ are the phase angle of q which can be obtained from the following formula:

$$\begin{cases} \phi = \arctan\left(\dfrac{2(ac + \mathrm{bd})}{a^2 + b^2 + c^2 - d^2}\right), \\ \theta = \arctan\left(\dfrac{\mathrm{ab} + cd}{a^2 - b^2 + c^2 - d^2}\right), \\ \psi = \dfrac{1}{2}\arctan(2(ad - bd)). \end{cases} \tag{4}$$

According to the model in paper [26], the first representation of quaternion can be used to represent the three channels of color image as a pure quaternion.

$$q(x,\ y) = R(x,\ y)i + G(x,\ y)j + B(x,\ y)k, \tag{5}$$

where $R(x, y), G(x, y),$ and $B(x,\ y)$ are respectively the $R, G,$ and B components corresponding to the pixel at position $(x,\ y)$ in the color image. In this way, a color image can be represented by a quaternion matrix. Compared with the traditional method of sub channel processing or converting to gray image before processing, quaternion method can better reflect the integrity of color image.

2.2. Sparse representation theory

Sparse representation (SR) is a technique in which most or all of the original signals are represented or approximated by a linear combination of fewer fundamental signals selected from an over-complete dictionary. The basic signal is called “atom”, and the over-complete dictionary is made up of atoms whose number exceeds the dimension of the signal. After sparse signal representation, the more sparse the reconstructed signal is, the higher the accuracy will be.

Assume that the input signal $x \in R^n$ is an image vector and $D \in R^{n\times m}(n < m)$ is an over-complete dictionary, where n represents the signal dimension and m represents the number of prototype signals called atoms, which are contained in the dictionary D. This relationship of $n < m$ means that the number of atoms in the dictionary is more than the dimension of the image vector [27]. Thus there are adequate atoms number with homologous over-complete dictionary to execute flexible and meaningful representation operation. The signal x can be approximately represented by fewer atoms in dictionary D, i.e., $x \approx D\alpha$, where $\alpha \in R^m$ is the sparse coefficient vector. There is no unique solution for an uncertain system. The goal of SR is to use the given dictionary D to

calculate the sparsest α with the least nonzero terms in all solutions. The mathematical model of SR is as follows:

$$\min_{\alpha} \|\alpha\|_0, \text{ s.t. } \|x - D\alpha\|_2^2 \leq \varepsilon, \tag{6}$$

where $\|\alpha\|_0$ stands for the count of nonzero values in sparse coefficient $\alpha \cdot \varepsilon \geq 0$ is a preset approximation error tolerance.

2.3. Sparse representation model based on quaternion

A color image contains three channels R, G and B. At present, the sparse representation model of color image processes each channel independently with possibly different dictionaries. Mairal et al. [28] improved the sparse representation model of color image to deal with the connection of three channels. However, the experimental results show that the trained dictionary tends to be monochromatic, and the relationship among the three color channels is not well preserved. The combination of quaternion theory and sparse representation can put a series of constraints on the dictionary and coefficient, and can well retain the correlation among the three color channels of color image. The patch of a color image can be represented by pure quaternion, i.e., $y_q = 0 + y_r i + y_g j + y_b k$, where $y_q \in H^n$ is an n-dimension quaternion vector. The quaternion dictionary and its corresponding coefficients are respectively expressed as $D_q = 0 + D_r i + D_g j + D_b k$ and $\alpha_q = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$. We propose the quaternion-based sparse representation model as

$$\min_{\alpha_q} \|\alpha_q\|_0, \text{ s.t. } y_q = D_q \alpha_q. \tag{7}$$

Here $D_q \in H^{n\times T}$, $\alpha_q \in H^T$. is the number of pure quaternion atoms. Formula (7) can be extended to formula (8).

$$\{\hat{A}_q, \hat{D}_q\} = \arg\min_{D_q} \|Y_q - D_q A_q\|_2 + \lambda \|A_q\|_0. \tag{8}$$

Here formula (8) is the training process of quaternion dictionary.

Where the first term is the reconstruction error, and the second term is the sparsity penalty function. The parameter λ represents the tradeoff relationship between data reconstruction and sparsity. $Y_q^{n\times N} = \left\{Y_q^l, 1 \le l \le N\right\}$ represents a set of sample image blocks. Where $D_q^{n\times T} = \left\{d_q^l \middle| d_q^l = d_r^l i + d_g^l j + d_b^l k\right\}$ and the value range of l is $1 \le l \le T$. And $A_q^{T\times N} = \left\{\alpha_q^l \middle| \alpha_q^l = \alpha_r^l i + \alpha_g^l j + \alpha_b^l k\right\}$ is the coefficient matrix. The value range of its corresponding l is $1 \le l \le N$.

Over-complete dictionary selection and optimization problem solution are two main issue of sparse representation. The approaches of acquiring over-complete dictionary can be roughly classified into two kinds: (1) fixed transform bases (discrete cosine transform, wavelets, curvelets, etc.) [29]; (2) dictionary training algorithm (MOD [30], K-SVD [31], OLD [32], and K-QSVD [23]). Because the optimization problem is an NP-hard problem, it can only be solved by approximation approach, which mainly include:

(1) Greed tracking algorithm (MP [35], OMP [36], SOMP [37], and QOMP [23]); (2) l_p norm regularization (BP [33], FUOCUSS [34]). The color of the dictionary trained by K-SVD method tends to be monochromatic, which can not show a satisfactory result in terms of color diversity. However, the K-QSVD method shows superior performance in training quaternion dictionary. Not only it can keep the correlation among the three color channels of the color image, but also it can keep the spatial consistency. K-QSVD can update atoms and coefficients faster in quaternion system, which shows its high efficiency. QOMP is also

efficient in sparse representation. So we use QOMP to obtain the optimal sparse representation coefficients, and use K-QSVD to train the over-complete dictionary.

3. Proposed Fusion Method

In the field of digital image processing, different imaging devices acquire different information from the same scene. The optical image is only clear in the part of the scene that is focused in the lens range, and the rest is a blurred defocused image. Generally speaking, a focused camera can only obtain the image of local focus. Multi-focus image fusion technology can process two or more locally focused images to extract the information of their respective focus areas to the maximum extent, and finally obtain a full focus image with better quality than the original image. The fused image contains more scene information, which is more suitable for imaging features of the human eye and is also convenient for later computer processing. For multi-focus color image fusion, the traditional fusion method for gray image can not be directly used for color image fusion. Otherwise the correlation among the three color channels of color image will not be well maintained, which makes the fusion result appear color deviation and other bad effects. The traditional sparse representation model usually processes the three channels among the color image separately or connects the three channels for processing, but such processing cannot maintain the connection between the three channels of the color image well.

The fusion method in this paper combines quaternion theory and sparse representation model. The image fusion framework is shown in Figure 1. Firstly, quaternion is used to represent the source color multi-focus image as a whole. Secondly, sparse representation theory is introduced in the fusion process, and quaternion is used to represent the non-local similarity of image and train quaternion dictionary. According to the constructed over-complete dictionary, the input image to be fused is

divided into blocks to obtain the column vectors composed of reference image blocks and similar image blocks, and then the sparse representation coefficients of each column vector under the dictionary are calculated. Then the sparse coefficients are fused according to the "max-L1" rule. Finally, the quaternion multi-focus fusion image is reconstructed by combining the over-complete dictionary and the fused sparse coefficient, and then converted into RGB color multi-focus fusion image.

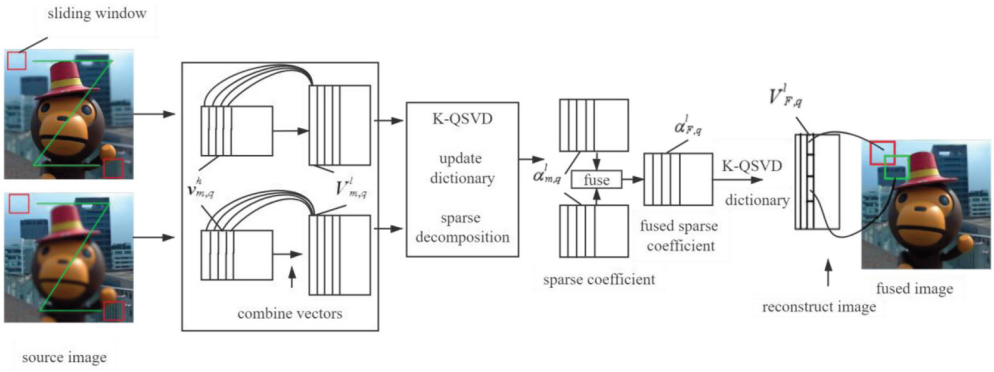


Figure 1. The architecture of proposed method.

3.1. Detailed fusion scheme

Step 1: Using quaternion to represent source multi-focus color image

The source color multi-focus image $I_m(m = 1, 2, \ldots, M)$ is represented by pure quaternion, and the R, G and B color components of the color image are expressed as three imaginary parts of quaternion. The expression is as follows:

$$I_q(x, y) = R(x, y)i + G(x, y)j + B(x, y)k. \tag{9}$$

Then two quaternion images I_A and I_B are obtained.

Step 2: After the image is represented by quaternion, sparse representation theory is introduced into the fusion process. According to the constructed over-complete dictionary, the input images to be fused are segmented into blocks and vectorized in the form of columns, and then the sparse representation coefficients of each column vector under the dictionary are calculated. The process mainly includes the following two parts.

(I) Column vectorization representation of images

The research results of BM3D principle show that there are local similar blocks and global similar blocks in the image, and the combination of similar blocks is beneficial to image processing. Apply the sliding window technique to divide source color image into image patches of the same size. L image blocks are selected randomly in the source color image, and the size of the image block is $n \times n$.

We suppose that the reference block at pixel r is $P_r \in Q^{n\times n}$(Q is quaternion space). Sliding any window in the whole image and selecting an image block of the same size. This image block is $P_s \in Q^{n\times n}$. The similarity η_q between the reference image block P_r and other arbitrary image block P_s is calculated by the similarity measurement criterion.

For multi-focus images, Euclidean distance method is usually used to measure the similarity between image blocks. Euclidean distance method means that after image blocks are grouped, Euclidean distance between image blocks is calculated to represent their similarity. However, for the color multi-focus image represented by quaternion, the similarity between image blocks should be represented by Euclidean distance represented by quaternion. The expression is as follows:

$$\eta_q(P_r,\ P_s) = \|P_r - P_s\|_F = \sqrt{\sum_{x=1}^{M}\sum_{x,\ y=1}^{N}\left|\eta_{r,s}(x,\ y)\right|}, \tag{10}$$

where $\eta_q(P_r, P_s)$ represents the quaternion Euclidean distance of the image blocks P_r and P_s. According to the quaternion modular operation rules, the quaternion distance of pixels corresponding to x row and y column of two image blocks can be expressed as

$$\begin{aligned}\eta_{r,s}(x, y) &= |(R_r(x, y)i + G_r(x, y)j + B_r(x, y)k) \\ &\quad - (R_s(x, y)i + G_s(x, y)j + B_s(x, y)k)|^2 \\ &= (R_r(x, y) - R_s(x, y))^2 + (G_r(x, y) - G_s(x, y))^2 \\ &\quad + (B_r(x, y) - B_s(x, y))^2. \end{aligned} \tag{11}$$

All blocks are arranged in descending order according to the size of η_q value, and h image blocks with the smallest η_q value are finally selected as the most similar blocks. This process can be illustrated in Figure 2. The reference image block and similar image block are all expressed as the form of $n^2 \times 1$-dimensional vector $v_{m,q}^h$ in the form of column priority. Then according to the similarity, the heads and tails of these $n^2 \times 1$-dimensional vectors are connected to form $(h+1) \times n^2 \times 1$-dimensional vector $V_{m,q}^l$ in turn. $V_{m,q}^l$ represents the vector composed of the l-th reference image block and its similar image block. And the value range of L is $1 \le l \le L$.

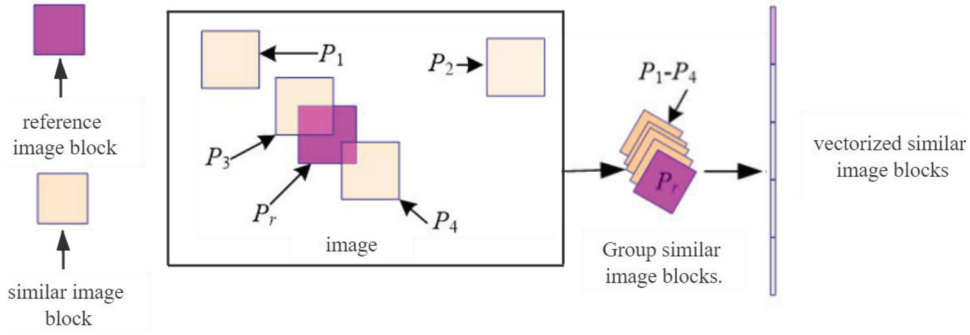


Figure 2. Schematic diagram of vectorization of reference image block and similar image blocks.

(II) Sparse representation

Repeat the previous process L times, so that you get $L(h+1)\times n^2 \times 1$-dimensional vectors, which are successively arranged from left to right to form matrix V with L columns and $(h+1)\times n^2$ rows. So you can get the matrices $V_{A,q}$ and $V_{B,q}$. Each column is used as the training sample, and the dictionary matrix D_q is obtained by using K-QSVD method to train the matrices $V_{A,q}$ and $V_{B,q}$. The K-QSVD method is as follows:

(1) Initializing dictionary matrix with over-complete quaternion dictionary D_q.

(2) The next step is the sparse coding. The sparse representation coefficients of each signal are solved by QOMP algorithm. Calculate the sparse coefficient vectors $\alpha_{A,q}$ and $\alpha_{B,q}$ of $V_{A,q}$ and $V_{B,q}$ by using QOMP.

$$\begin{aligned} \alpha_{A,q}^{l} &= \arg\min_{\alpha}\|\alpha\|_0, \quad \text{s.t.} \quad \left\|V_{A,q}^{l} - D_q\alpha\right\|_2 \le \varepsilon, \\ \alpha_{B,q}^{l} &= \arg\min_{\alpha}\|\alpha\|_0, \quad \text{s.t.} \quad \left\|V_{B,q}^{l} - D_q\alpha\right\|_2 \le \varepsilon, \end{aligned} \tag{12}$$

where $\alpha_{A,q}^{l}$ and $\alpha_{B,q}^{l}$ are the l-th column vector in $\alpha_{A,q}$ and $\alpha_{B,q}$. $V_{A,q}^{l}$ and $V_{B,q}^{l}$ are the l-th column vector in $V_{A,q}$ and $V_{B,q}$.

(3) Then the dictionary matrix D_q will be updated. Update the t-th column d_q^t in the dictionary to minimize the mean square error (MSE).

$$E_q^t = Y_q - \sum_{l \neq t} d_q^l A_{q,l},\ l = 1, 2, \ldots, t-1, t+1, \ldots, L. \tag{13}$$

$A_{q,l}$ is the l-th row of the sparse matrix A_q. Suppose $\omega_t = \{i, 1 \leq i \leq N, A_q(t, i) \neq 0\}$. We compute E_q^t and select partial columns of E_q^t according to the flag element in ω_t to get $E_q^{R,t}$. Then it is decomposed into $E_q^{R,t} = U_q \nabla V_q^H$ by QSVD method. Update the t-th column of d_q^t as the first column vector of U_q. The sparse coefficient of the t-th column atom is obtained by multiplying the first column of V_q^H by $\nabla(1, 1)$.

(4) If the number of iterations of the experiment is reached, the final redundant dictionary D_q will be obtained.

Here D_q is as follows: $D_q = \left\{ d_q^l \middle| d_q^l = d_r^l i + d_g^l j + d_b^l k \right\}$. Otherwise, go to the step (2).

Step 3: Fusion of coefficients

The sparse representation coefficients of the corresponding reference image blocks and their similar image blocks of each image are fused according to certain fusion rules to obtain the sparse coefficients of images waiting to be reconstructed. In this paper, the fused coefficient vector $a_{F,q}^l$ is calculated according to the "max-L1" rule.

$$\alpha_{F,q}^l = \begin{cases} \alpha_{A,q}^l & \left\|\alpha_{A,q}^l\right\|_1 > \left\|\alpha_{B,q}^l\right\|_1, \\ \alpha_{B,q}^l & \text{else.} \end{cases} \tag{14}$$

Step 4: Reconstruction of image

Reconstruction of the image is the inverse process of sparse decomposition, which combines the over-complete dictionary with the fused coefficients. The learned dictionary D_q is multiplied by the fused coefficient $\alpha_{F,q}^{l}$ to obtain the fused vector $V_{F,q}^{l}$. It is expressed as follows: $V_{F,q}^{l} = D_q \alpha_{F,q}^{l} \cdot V_{F,q}^{l}$ is the fused column vector of reference image blocks and similar image blocks. The inverse process of the vectorization process in Step 2 is used to obtain the fused reference image blocks and similar image blocks, and restore them to their initial positions.

Finally, all the image blocks restored to the same initial position are arithmetically averaged to get the fused image $I_{F,q}$ represented by quaternion. Then, the quaternion image is converted into a color fusion image I_F in the form of R, G and B.

4. Experiments and Analysis

In this subsection, in order to prove the effectiveness of the proposed algorithm, multi-focus color image datasets named Lytro [38], which contains 20 pairs of color multi-focus images, are used for experiments. A part of Lytro are shown in Figure 3. At the same time, some popular image fusion methods such as the transform domain methods, SR [19], the spatial domain methods MWG [18], DSIFT [42], and deep learning based CNN [41] are set as the contrast methods. The parameters of each method are set strictly according to the recommended values mentioned in the relevant literatures for the optimal performance. And the experiment results told that the proposed method has a superior performance and can achieve a better visual effect for multi-focus image fusion from both intuitive visual perception and objective perception.

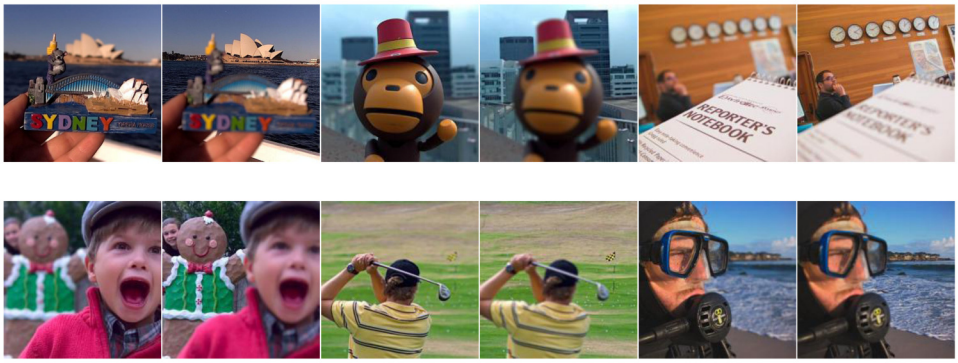


Figure 3. Examples of the source image pairs from the "Lytro" multi-focus dataset.

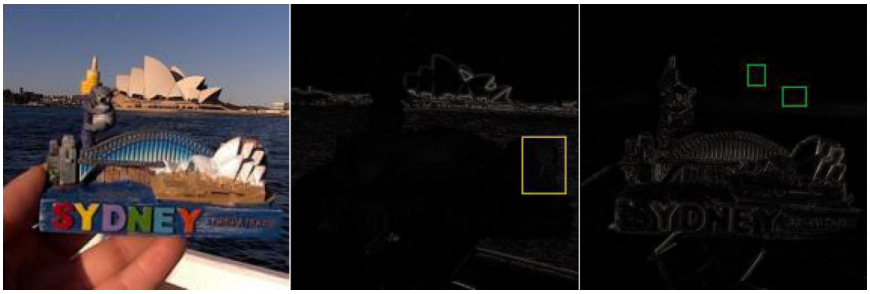
4.1. Subjective visual effect

In this subsection, the fusion results of three pairs of testing images are respectively exhibited in Figures 4 and 5 to assess the performance of different fusion methods from the perceptive of visual quality. In Figure 4, from the difference image between the fusion result of SR method and the source image, we can observe that the boundary of the focused area and the defocused area marked by the yellow box produces artifacts. The places marked with green boxes showed different degrees of residues. The MWG method produces a certain degree of residue on the boundary between the focused area and the defocused area, which can be shown in the area marked by the yellow box in the difference image. DSIFT method has a good performance in distinguishing the boundary between the focused area and the defocused area. CNN method fails to deal with the rich and detailed boundaries between the focused area and defocused area marked out with red box. It can't deal with the gap between the focused finger and the defocused background well, which can be reflected in the part of the yellow box marked finger and the background in the difference image. DSIFT method is also not good in dealing with the gap between the finger and the background. Our method can distinguish the gap between the focused finger and the unfocused background, such as

the area marked by the yellow box in the difference image. Our method also shows a good result in dealing with focused and unfocused regions, and the fusion results are more consistent with human visual perception subjectively.

In Figure 5, for the SR method, we can see that there is a certain degree of residue in the defocused region in the difference image between the fusion result and the source image. It can be shown in the area marked by the yellow box. The MWG method does not deal with the focused area and the defocused boundary well, and produces artifacts at the boundary, which can be reflected in the area marked by the yellow box. It also has not achieved satisfactory results in dealing with children's neckline, which can be shown in the red box marked area. DSIFT method also has residues in the defocused area, which can be reflected in the area marked by the yellow box. It also failed to deal with the boundary of children's clothes, and failed to maintain the continuity and integrity of the clothing boundary, which can be reflected in the area marked by the red box. CNN method also produces unsatisfactory results in dealing with the boundary of children's clothes, such as the area marked by the yellow box in the difference image between the fusion result and the source image. Not only can our method keep the continuity and integrity of the boundary of children's clothes, but also have no residue in the focused area and defocused area. Our method also produces no artifacts at the boundary of the focused area and the defocused area.

Source 1 Source 2



SR SR-Source 1 SR-Source 2

MWG MWG-Source 1 MWG-Source 2

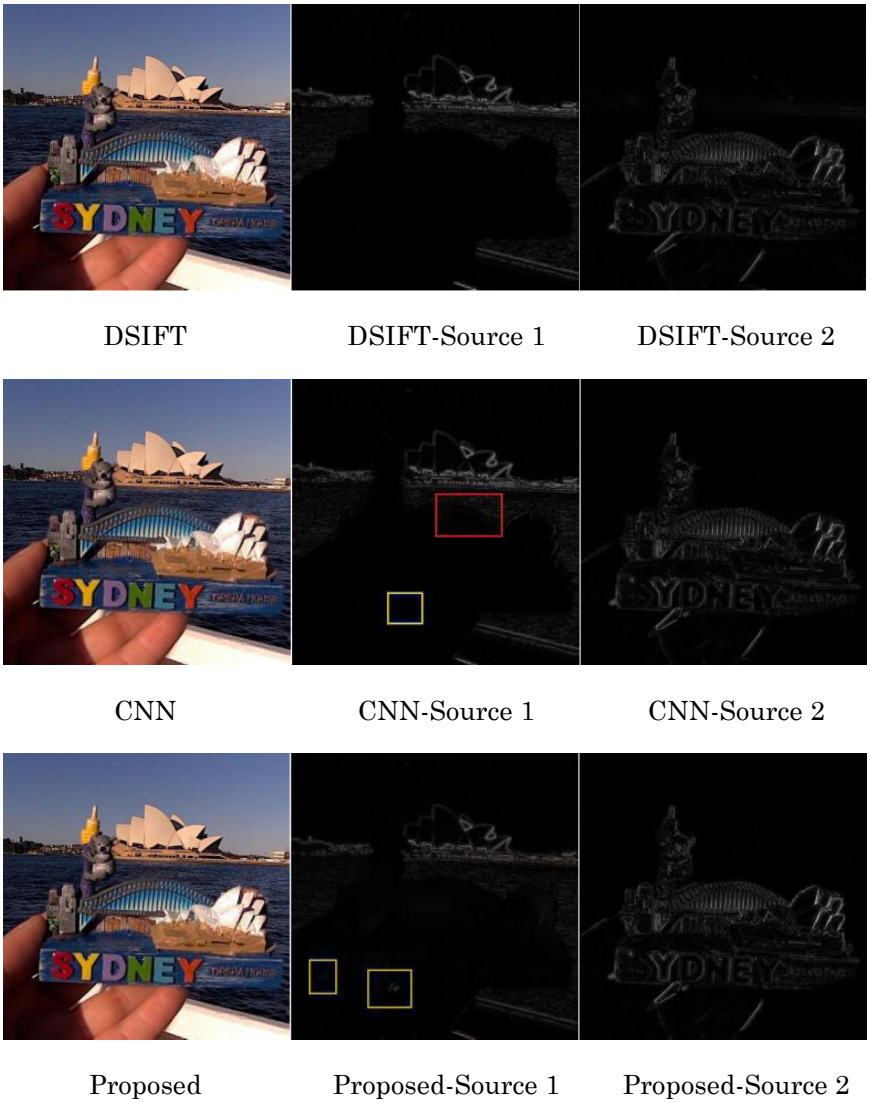
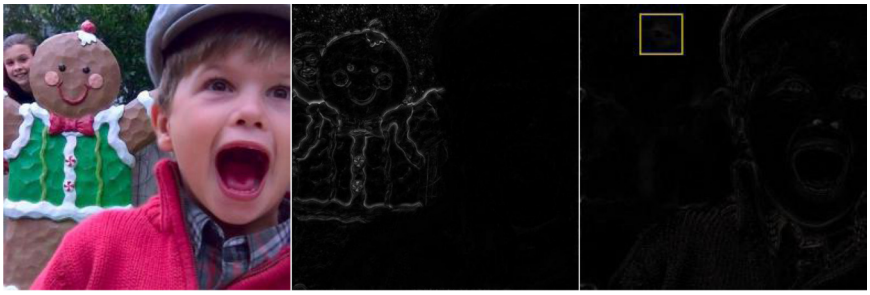
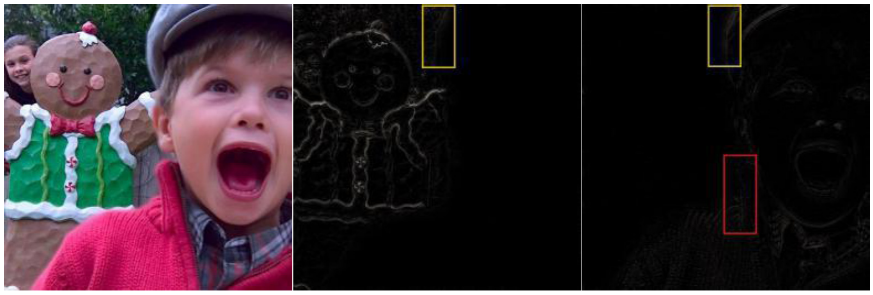


Figure 4. Fusion results of the 1st image pair in Figure 3 using different methods. The first column is the fused images obtained by SR, MWG, DSIFT, CNN and our proposed method. The other two columns are respectively the difference between the fused image and two source images.

Source 1 Source 2



SR SR-Source 1 SR-Source 2



MWG MWG-Source 1 MWG-Source 2

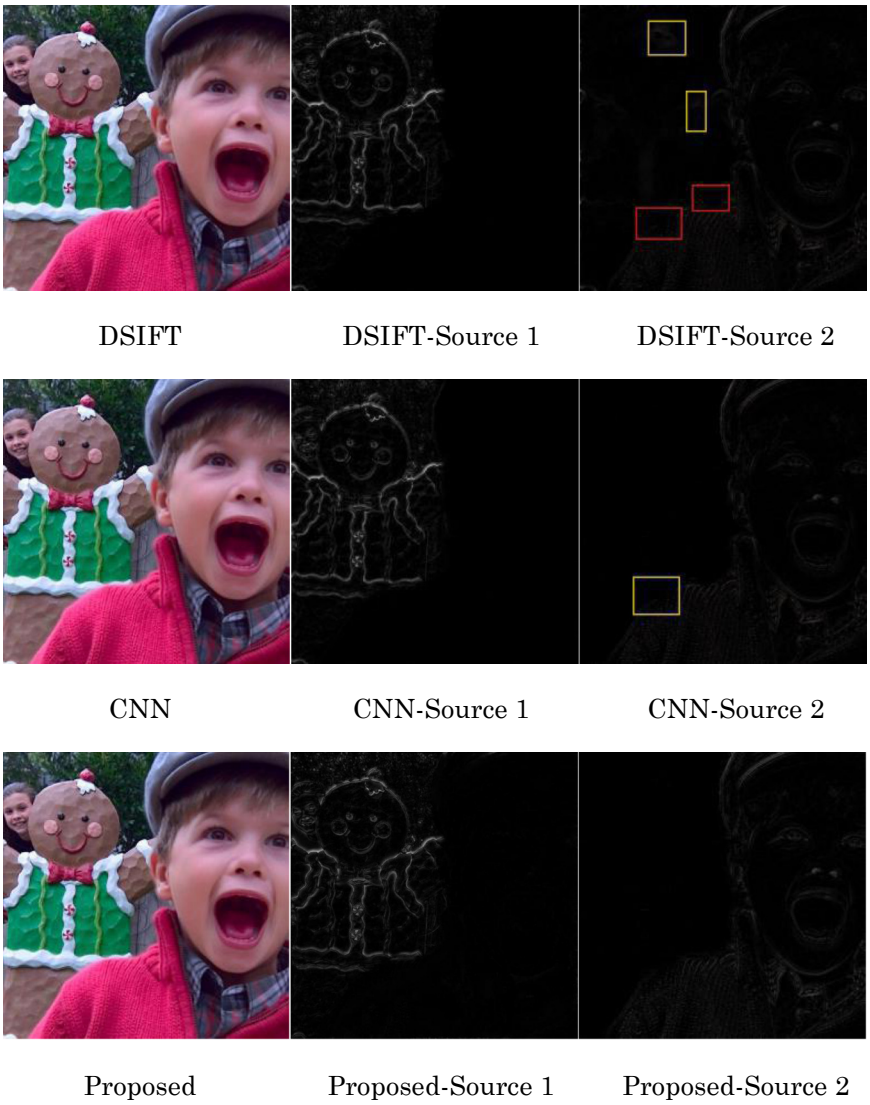


Figure 5. Fusion results of the 4th image pair in Figure 3 using different methods. The first column is the fused images obtained by SR, MWG, DSIFT, CNN and our proposed method. The other two columns are respectively the difference between the fused image and two source images.

4.2. Objective evaluation indexes

In order to make a more comprehensive analysis and comparison, we select several quantitative metrics to evaluate the fusion performance of these methods in various aspects from an objective perspective. In our experiments, we exploit six evaluation metrics to objectively evaluate the performance of fused images obtained by different algorithms, which are phase congruency Q_P [39], SSIM-based metric Q_Y [40], nonlinear correlation information entropy Q_{NCIE} [43], human perception-based metric Q_{CB} [44], mutual information Q_{MI} [45], tsallis entropy-based metric Q_{TE} [46]. For all the six evaluation metrics, a larger value indicates a better fused result. For quantitatively assess the performance of different fusion methods, we carry out the fusion upon twenty pairs of source images with different scene and list the average scores belong to the six metrics in Table 1, where the best result for each assessment is highlighted in bold. From the comparison of objective assessment, we can find that most performance scores of the former three methods comprehensively fall behind the later four methods. Our proposed method gets the best score in all six metrics. To sum up briefly, the validity and superiority of our proposed method are demonstrated by both of the subjective visual effects and objective performance assessment.

Table 1. Objective assessment of different fusion methods

Method	Q_P	Q_Y	Q_{NCIE}	Q_{CB}	Q_{MI}	Q_{TE}
SR	0.8168	0.9653	0.8374	0.7756	1.0642	0.8133
MWG	0.8240	0.9786	0.8396	0.7893	1.0918	0.8138
DSIFT	0.8320	0.9789	0.8415	0.8014	1.1292	0.8162
CNN	0.8326	0.9781	0.8402	0.7995	1.1095	0.8170
Proposed	**0.8464**	**0.9857**	**0.8436**	**0.8054**	**1.1341**	**0.8216**

5. Conclusion

In this paper, we propose a multi-focus image fusion algorithm based on quaternion theory and sparse representation theory. By representing the source color image in the form of quaternion, the pixels of the color image can be processed as a whole vector, and the fusion image represented by quaternion can be obtained by combining with the sparse representation model. Finally, the fusion image of color RGB can be obtained after transformation. Unlike the traditional sparse representation model, this method does not process the three channels of color image separately or connect them together for processing. Our method can overcome the limitation of splitting the connection among the channels of color image and effectively maintain the connection among the three channels of color multi-focus image. And the information of the focus areas in the source images can be obtained effectively, and the fusion image with rich focus information can be obtained, which is more in line with the perception of human vision. From the experimental results, not only is the method superior to other competitive methods in subjective vision, but also it better than other competition methods in objective evaluation.

References

[1] G. Piella, A general framework for multiresolution image fusion: From pixels to regions, Information Fusion 4(4) (2003), 259-280.

DOI: https://doi.org/10.1016/S1566-2535(03)00046-0

[2] O. Rockinger, Image sequence fusion using a shift-invariant wavelet transform, Proceedings of International Conference on Image Processing, Washington DC, USA: IEEE Computer Society, (1997), 288-291.

DOI: https://doi.org/10.1109/ICIP.1997.632093

[3] L. A. Ray and R. R. Adhami, Dual tree discrete wavelet transform with application to image fusion, Proceedings of the 38th Southeastern on System Theory, Huntsville, Albania: IEEE Computer Society (2006), 430-433.

DOI: https://doi.org/10.1109/SSST.2006.1619105

[4] S. Ioannidou and V. Karathanassi, Investigation of the dual-tree complex and shift-invariant discrete wavelet transforms on quickbird image fusion, IEEE Geoscience and Remote Sensing Letters 4(1) (2007), 166-170.

DOI: https://doi.org/10.1109/LGRS.2006.887056

[5] F. Nencini, A. Garzelli, S. Baronti and L. Alparone, Remote sensing image fusion using the curvelet transform, Information Fusion 8(2) (2007), 143-156.

DOI: https://doi.org/10.1016/j.inffus.2006.02.001

[6] M. N. Do and M. Vetterli, The contourlet transform: An efficient directional multiresolution image representation, IEEE Transactions on Image Processing 14(12) (2005), 2091-2106.

DOI: https://doi.org/10.1109/TIP.2005.859376

[7] A. L. Da Cunha, J. P. Zhou and M. N. Do, The nonsubsampled contourlet transform: Theory, design, and applications, IEEE Transactions on Image Processing 15(10) (2006), 3089-3101.

DOI: https://doi.org/10.1109/TIP.2006.877507

[8] B. Yang, S. T. Li and F. M. Sun, Image fusion using nonsubsampled contourlet transform, Proceedings of the 4th International Conference on Image and Graphics, Chengdu, China: IEEE Computer Society (2007), 719-724.

DOI: https://doi.org/10.1109/ICIG.2007.124

[9] Q. Zhang and B. L. Guo, Multifocus image fusion using the nonsubsampled contourlet transform, Signal Processing 89(7) (2009), 1334-1346.

DOI: https://doi.org/10.1016/j.sigpro.2009.01.012

[10] Shutao Li and Bin Yang, Multifocus image fusion by combining curvelet and wavelet transform, Pattern Recognition Letters 29(9) (2008), 1295-1301.

DOI: https://doi.org/10.1016/j.patrec.2008.02.002

[11] W. Huang and Z. Jing, Evaluation of focus measures in multi-focus image fusion, Pattern Recognition Letters 28(4) (2007), 493-500.

DOI: https://doi.org/10.1016/j.patrec.2006.09.005

[12] V. Aslantas and R. Kurban, Fusion of multi-focus images using differential evolution algorithm, Expert Systems with Applications 37(12) (2010), 8861-8870.

DOI: https://doi.org/10.1016/j.eswa.2010.06.011

[13] I. De and B. Chanda, Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure, Information Fusion 14(2) (2013), 136-146.

DOI: https://doi.org/10.1016/j.inffus.2012.01.007

[14] X. Bai, Y. Zhang, F. Zhou and B. Xue, Quadtree-based multi-focus image fusion using a weighted focus-measure, Information Fusion 22 (2015), 105-118.

DOI: https://doi.org/10.1016/j.inffus.2014.05.003

[15] M. Li, W. Cai and Z. Tan, A region-based multi-sensor image fusion scheme using pulse-coupled neural network, Pattern Recognition Letters 27(16) (2006), 1948-1956.

DOI: https://doi.org/10.1016/j.patrec.2006.05.004

[16] S. Li and B. Yang, Multifocus image fusion using region segmentation and spatial frequency, Image and Vision Computing 26(7) (2008), 971-979.

DOI: https://doi.org/10.1016/j.imavis.2007.10.012

[17] S. Li, X. Kang and J. Hu, Image fusion with guided filtering, IEEE Transactions on Image Processing 22(7) (2013), 2864-2875.

DOI: https://doi.org/10.1109/TIP.2013.2244222

[18] Z. Zhou, S. Li and B. Wang, Multi-scale weighted gradient-based fusion for multi-focus images, Information Fusion 20 (2014), 60-72.

DOI: https://doi.org/10.1016/j.inffus.2013.11.005

[19] B. Yang and S. T. Li, Multifocus image fusion and restoration with sparse representation, IEEE Transactions on Instrumentation and Measurement 59(4) (2010), 884-892.

DOI: https://doi.org/10.1109/TIM.2009.2026612

[20] Qiang Zhang and Martin D. Levine, Robust multi-focus image fusion using multi-task sparse representation and spatial context, IEEE Transactions on Image Processing 25(5) (2016), 2045-2058.

DOI: https://doi.org/10.1109/TIP.2016.2524212

[21] Xin Jin, Rencan Nie, Dongming Zhou, Quan Wang and Kangjian He, Multifocus color image fusion based on NSST and PCNN, Journal of Sensors 2 (2016); Article ID 8359602, pp. 1-12.

DOI: https://doi.org/10.1155/2016/8359602

[22] Rui Zeng, Jiasong Wu, Zhuhong Shao, Yang Chen, Beijing Chen, Lotfi Senhadji and Huazhong Shu, Color image classification via quaternion principal component analysis network, Neurocomputing 216 (2016), 416-428.

DOI: https://doi.org/10.1016/j.neucom.2016.08.006

[23] Yi Xu, Licheng Yu, Hongteng Xu, Hao Zhang and Truong Nguyen, Vector sparse representation of color image using quaternion matrix analysis, IEEE Transactions on Image Processing 24(4) (2015), 1315-1329.

DOI: https://doi.org/10.1109/tip.2015.2397314

[24] W. R. Hamilton, On Quaternions, Proceeding of the Royal Irish Academy, 1844.

[25] L. L. Kantor and A. S. Solodovnikov, Hypercomplex Numbers: An Elementary Introduction to Algebras, Springer-Verlag, 1989.

[26] S.-C. Pei and C.-M. Cheng, A novel block truncation coding of color images by using quaternion-moment-preserving principle, IEEE Transactions on Communications 45(5) (1997), 583-595.

DOI: https://doi.org/10.1109/26.592558

[27] Pan Zhu, Lu Liu and Xinglin Zhou, Infrared polarization and intensity image fusion based on bivariate BEMD and sparse representation, Multimedia Tools and Applications 80(3) (2021), 4455-4471.

DOI: https://doi.org/10.1007/s11042-020-09860-z

[28] J. Mairal, M. Elad and G. Sapiro, Sparse representation for color image restoration, IEEE Transactions on Image Processing 17(1) (2008), 53-69.

DOI: https://doi.org/10.1109/TIP.2007.911828

[29] Y. Liu, S. P. Liu and Z. F. Wang, A general framework for image fusion based on multi-scale transform and sparse representation, Information Fusion 24 (2015), 147-164.

DOI: https://doi.org/10.1016/j.inffus.2014.09.004

[30] K. Engan, S. O. Aase and J. H. Husoy, Multi-frame compression: Theory and design, Signal Process 80(10) (2000), 2121-2140.

DOI: https://doi.org/10.1016/S0165-1684(00)00072-4

[31] A. Michal, M. Elad and A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Transactions on Signal Processing 54(11) (2006), 4311-4322.

DOI: https://doi.org/10.1109/TSP.2006.881199

[32] J. Mairal, F. Bach, J. Ponce and G. Sapiro, Online learning for matrix factorization and sparse coding, Journal of Machine Learning Research 11 (2010), 19-60.

[33] S. S. Chen, D. L. Donoho and M. A. Saunders, Atomic decomposition by basis pursuit, SIAM Review 43(1) (2001), 129-159.

DOI: https://doi.org/10.1137/S003614450037906X

[34] I. F. Gorodnitsky and B. D. Rao, Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm, IEEE Transactions on Signal Processing 45(3) (1997), 600-616.

DOI: https://doi.org/10.1109/78.558475

[35] S. G. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Transactions on Signal Processing 41(12) (1993), 3397-3415.

DOI: https://doi.org/10.1109/78.258082

[36] Y. C. Pati, R. Rezaiifar and P. S. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, Proceedings of 27th Asilomar Conference on Signals, Systems and Computers 1 (1993), 40-44.

DOI: https://doi.org/10.1109/ACSSC.1993.342465

[37] B. Yang and S. T. Li, Pixel-level image fusion with simultaneous orthogonal matching pursuit, Information Fusion 13(1) (2012), 10-19.

DOI: https://doi.org/10.1016/j.inffus.2010.04.001

[38] M. Amin-Naji, A. Aghagolzadeh and M. Ezoji, Fully convolutional networks for multi-focus image fusion, 9th International Symposium on Telecommunications (2018), 553-558.

DOI: https://doi.org/10.1109/ISTEL.2018.8660989

[39] Q. Wang, Y. Shen and J. Jin, 19-Performance evaluation of image fusion techniques, Image Fusion: Algorithms and Applications, Great Britain 19 (2008), 469-492.

DOI: https://doi.org/10.1016/B978-0-12-372529-5.00017-2

[40] C. Yang, J. Q. Zhang, X. R. Wang and X. Liu, A novel similarity based quality metric for image fusion, Information Fusion 9(2), (2008), 156-160.

DOI: https://doi.org/10.1016/j.inffus.2006.09.001

[41] Y. Liu, X. Chen, H. Peng and Z. Wang, Multi-focus image fusion with a deep convolutional neural network, Information Fusion 36 (2017), 191-207.

DOI; https://doi.org/10.1016/j.inffus.2016.12.001

[42] Y. Liu, S. Liu and Z. Wang, Multi-focus image fusion with dense SIFT, Information Fusion 23 (2015), 139-155.

DOI: https://doi.org/10.1016/j.inffus.2014.05.004

[43] Y. Chen and R. S. Blum, A new automated quality assessment algorithm for image fusion, Image and Vision Computing 27(10) (2009), 1421-1432.

DOI: https://doi.org/10.1016/j.imavis.2007.12.002

[44] M. Hossny, S. Nahavandi and D. Creighton, Comments on: Information measure for performance of image fusion, Electronics Letters 44(18) (2008), 1066-1067.

DOI: https://doi.org/10.1049/el:20081754

[45] N. Cvejic, C. N. Canagarajah and D. R. Bull, Image fusion metric based on mutual information and Tsallis entropy, Electrons Letters 42(11) (2006), 626-627.

DOI: https://doi.org/10.1049/el:20060693

[46] Y. Zheng, E. A. Essock, B. C. Hansen and A. M. Haun, A new metric based on extended spatial frequency and its application to DWT based fusion algorithms, Information Fusion 8(2) (2007), 177-192.

DOI: https://doi.org/10.1016/j.inffus.2005.04.003

■