A Remote Sensing Image Fusion Method based on adaptive dictionary learning

Tongdi He¹*, Zongxi Che²

¹College of physical and mechanical & electrical Engineering, Hexi University, Gansu Zhangye, China
²Nature Reserves Bureau of Qilian Mountain, Gansu Zhangye, China

*Corresponding author e-mail: hetongdi@126.com

Abstract. This paper discusses using a remote sensing fusion method, based on adaptive sparse representation (ASP), to provide improved spectral information, reduce data redundancy and decrease system complexity. First, the training sample set is formed by taking random blocks from the images to be fused, the dictionary is then constructed using the training samples, and the remaining terms are clustered to obtain the complete dictionary by iterated processing at each step. Second, the self-adaptive weighted coefficient rule of regional energy is used to select the feature fusion coefficients and complete the reconstruction of the image blocks. Finally, the reconstructed image blocks are rearranged and an average is taken to obtain the final fused images. Experimental results show that the proposed method is superior to other traditional remote sensing image fusion methods in both spectral information preservation and spatial resolution.

1. Introduction
Remote sensing image fusion refers to certain algorithms used for generation of new information or a synthetic image from two or more images within the same scene that have been obtained from different sensors or the same sensor with different scales, in order to improve the clarity and legibility of the images, and to obtain characteristic information which cannot be provided from a single image [1]. Remote sensing image fusion multi-spectral and pan-chromatic images uses complementary features and redundancy between the images to increase the spatial details of the fused image as much as possible while maintaining the original spectral characteristics of the image. The aim is to obtain an optimal target scene information description, which will not only improve the quality of the remote sensing image, but can also be beneficial for subsequent processing of the remote sensing image, such as classification and object recognition [2]. Remote sensing image fusion plays a decisive role in aerospace, military reconnaissance, disaster prediction and many military and civilian fields. It can provide abundant and valuable information for resource investigation and environmental monitoring, and is used for global change research, environmental monitoring, resource survey and disaster prevention, and many other control aspects of applications [3]. Currently, the main fusion algorithms in the spectral domain are the Brovey transform, the IHS transform [4], the PCA transformation [5, 6], and the wavelet analysis method is also a popular method, which was recently proposed [7]. The IHS transform, PCA transform and wavelet transform can improve the spatial resolution of the fusion
image. However, the IHS fusion method distorts the spectral characteristics of the original image and produces spectral degeneration or color aberrations; the PCA fusion method also leads to spectral degradation; and the wavelet transform cannot effectively capture any directional image information, and also the lack of translation invariance causes image jitter around the sharp edges of the reconstructed image near any uniform local areas [8]. Inspired by the sparsity of human visual cortical neuron’s response, Olshausen and Field [9] suggested an effective signal representation sparse representation. This method is a type of adaptive data representation method that can adaptively select the most relevant atoms according to the characteristics of the image. The multi-scale dictionary used is better than a single-scale dictionary at matching the structure of each component in the image, with more sparse representation capability [10].

This article builds on research over the past few years on signal sparse representation theory to propose an adaptive dictionary learning remote sensing image fusion algorithm. Due to the sparsity and the energy concentration of the decomposition coefficient, the optimization process of the energy function can reduce the noise in the image block, meaning that the new algorithm is robust with improved adaptability. Experiments are performed using IKONOS data for visual effects and objective evaluation of the fusion results respectively. Results show that the proposed algorithm can improve the spatial details of the fused image while retaining multispectral image spectrum information.

2. Image sparse representation

2.1. Image sparse representation

A study is undertaken on an image Y, with image size, where the image length is and is the image width. If the image is decomposed into a set of complete orthogonal bases, then the size of the set should be. Due to orthogonality of the bases, the distribution of bases across the space formed by the image is sparse, so after decomposition, the energy of the image will be distributed across different mediums. This decentralized distribution of energy results in a linear combination of orthogonal bases of an image when the expression is not simple, it isn’t a sparse image representation. A non-sparse representation is not conducive to image processing, such as recognition and compression. In order to get a sparse representation of the image, the structure of each basis must be dense enough within the image space. As a result, the orthogonality of the bases will no longer be guaranteed, so the term ‘base’ is no longer accurate, and is renamed the ‘original’. The collection space composed of these atoms is complete, known as the Over-Complete Dictionary of Atoms. The image decomposition results in the Over-Complete Dictionary of Atoms must be sparse [11].

2.2. The structure of the adaptive multi-scale dictionary

Design of the redundant dictionary D is the key challenge in the theory of sparse representation, since dictionary selection affects the sparse vectors’ iteration results and convergence of the algorithm [12]. The dictionary atoms can be structured using the learning method, which can effectively capture the spectrum characteristics of the remote sensing image pixel so that the dictionary can efficiently represent the pixels. An adaptive multi-scale dictionary can be constructed by setting the training samples as a normalized vector with their starting point at the origin. Therefore, different training samples are simply vectors corresponding to each different direction, and a set of hyper-planes can be used to approximate the spherical cap to achieve sparse representation.

Implementation of ASP algorithm: N = input training samples of the training set, Y=[y(1), y(2), ..., y(N)]∈R*mN, where atoms of each layered signal are given by k(1), k(2), ..., k(l), l for each hierarchical number. The output Y is the sparse representation matrix α in the dictionary D:

**Step 1:** The column structure within the random image block can be represented by the sample set Y = {Y; i = 1, >, ..., N}. Y has size M×N, where M is the sample size (i.e. the block size) and N is the number of samples. The N samples of Y are divided into k(l) different categories using K-means clustering method, represented as θ(a), i = 1, 2, ..., k(l). When the observation vector
satisfies: 

\[ d(y(t), \hat{a}_i) = \min \left\{ y(t), \hat{a}_i, i = 1, 2, \ldots, k(t) \right\} \]

we consider \( y(t) \in \theta(\hat{a}_i) \cdot d(y(t), \hat{a}_i) \) is the distance between \( y(t) \) and the cluster centers \( \hat{a}_i \), where the clustering center \( [\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_{k}] \) is the first layer of atoms.

**Step 2:** Set \( [p_1, p_2, \ldots, p_N] \) as the \( [y_1, y_2, \ldots, y_N] \) projection in the respective cluster center, so \( e = y - p_i \) \( i = 1, 2, \ldots, N \);

**Step 3:** For any \( j \) and \( s \), if \( j \neq s \), \( \|e_j - e_s\| < \varepsilon \) or \( \|e_j + e_s\| < \varepsilon \). Then remove \( e_r \) to retain only \( e_s \), and obtain \( e_s' \);

**Step 4:** For \( e_s' \), K-means clustering, get \( \theta_{(s)}, \hat{a} \) Dictionary

\[ D = [\theta_{(0)}(a), \theta_{(1)}(a), \ldots, \theta_{(J)}(a)] \]

if \( i = j \), \( E = Y - DX \) Stop iteration. Otherwise, go to step 5.

**Step 5:** The OMP algorithm is used to calculate the sparse representation matrix \( X \) using the dictionary \( D \), so that \( X \) satisfies:

\[ \begin{align*}
\min_{D, X} & \quad \frac{1}{2} \|Y - DX\|_F^2, \quad \|x_i\|_0 \leq T_0 \\
\end{align*} \]

Where \( x_i \) is the \( i \)-th column of \( X \) and \( T_0 \) is the sparse degree

**Step 6:** Calculate the residual \( E = Y - DX \), where \( e_i \) is the \( i \)-th column of \( E \). Return to step 3.

### 2.3. Sparse reconstruction of image

The physical world of atoms and image sparse decomposition using the atom have a common characteristic: atomic energy is very concentrated and the vast majority of energy is distributed in the center of an atom, being sparsely distributed in all other locations [13]. Through image sparse decomposition, a linear image can be described by:

\[ Y = Y_m + Y_r = \sum_{k=0}^{n} < R^*Y, d_k > d_k + R^*Y \]

Where the residual components are small because \( \|R^*Y\| \) has attenuation characteristics. Since there are a small number of atoms (compared with the size of the image), the main components of the image can be described by:

\[ Y \approx Y_m + Y_r = \sum_{k=0}^{n} < R^*Y, d_k > d_k \]

The image reconstruction process uses the residual \( R^*Y \), an image set of \( N \) atomic parameters or the \( Y \) image in the corresponding atomic component to restore the image [14].

### 3. Image fusion based on adaptive sparse representation

The iterative computation process for over-complete sparse representation of the training samples can simultaneously obtain the over-complete dictionary and sparse coefficients. For the image fusion task, selection of the sparse coefficient determines the final fusion result. The image sparse points are determined from two remote sensing images which have been registered. The specific fusion process is as follows:

**Step 1:** Estimate the maximum noise variance of the input image \( \sigma^2 \), and set the sparse threshold \( T = \lambda \sigma^2 \).
**Step 2 (Image Block Creation):** If the atomic size is set to $M$, where $M = n \times n$, then the fused image $a, b$ needs to be divided atom-by-pixel $P_1$ and $P_2$ into $n \times n$ block sizes (the two images to be fused are set up to have the same size, $P_1 = P_2 = P$). The sample matrix consists of block rows and column vectors $Y_1$ and $Y_2$.

**Step 3 (Sparse coding):** Sparse decomposition of $Y_1$ and $Y_2$ is performed using the ASP algorithm to obtain the sparse coefficient matrix $\alpha_1$ and $\alpha_2$, where each column corresponds to an image block.

**Step 4 (Data fusion):** $\alpha_1$ and $\alpha_2$ are made to use certain fusion rules for feature selection of the significant coefficients as the fusion coefficient $\alpha$. The method of selection of the fusion rules is the important aspect here, and adaptive selection of weighted coefficients of regional energy is used in this paper. Determine point $(m, n)$ firstly as the central location of the local energy $E_{JA}$ and $E_{JB}$. The following formula (4) is used to calculate the fused image sparse representation coefficient:

$$\alpha = \frac{E_{ab}(m,n)}{E_{ab}(m,n)+E_{ab}(m,n)} \alpha_1 + \frac{E_{ab}(m,n)}{E_{ab}(m,n)+E_{ab}(m,n)} \alpha_2$$

(4)

The characteristics of the original image affect the characteristics of the weighting coefficient: larger regional energy in the obtained image A corresponds to a larger weighted coefficient; smaller regional energy corresponds to a smaller weighted coefficient.

**Step 5 (Image block reconstruction):** The fusion coefficient and over-complete dictionary’s convolution block are reconstructed as $\hat{y} = DA$, $y = \{y_i, i = 1, \ldots, N\}$, where $\hat{y}$ is the reconstructed image block vector after fusion.

**Step 6 (Image reconstruction):** Use $y_i$ recovery as the image block data and reconstruct the image by reordering the blocks based on their order plus the mean. The overlapped block average can be used to realize the image reconstruction to obtain the final fusion image $Y$.

**4. Experiment and Analysis**

In this paper, some high spatial resolution full-color and low spatial multispectral image are fused. To verify the performance of this algorithm, three groups of fusion algorithm performance experiments with full-color images and multispectral images were conducted. The results from the three experiments are consistent. Experimental source images are shown in Fig.1, (a) and (b) with $512 \times 512$ pixels and 256 grayscale level.

Figure 1. Original remote sensing image: (a) IKONOS panchromatic image; (b) IKONOS multispectral.
Five traditional experiment methods (HIS transformation, Principle Component Analysis, weighted fusion, high-pass filtering transformation, and wavelet transformation) are proposed and compared. The fused results of six fusion methods are shown in Fig. 2.

![Fused images](image1)

**Figure 2.** Fused images: (a) IHS; (b) PCA; (c) Weighted Average; (d) HPF; (e) Wavelet; (f) Proposed method.

The purpose of image fusion is to retain the important information of the two images to a maximum degree. Therefore, for the multispectral and high resolution multi-sensing image, surface spectral information of original multispectral image should be kept, and the detailed texture information of high resolution image should be brought in as well. To estimate the image fusion effects objectively, the above two aspects should be considered comprehensively. The objective indicators, for example, experimental mean value, information entropy, definition, deviation index, spectral distortion degree, and Correlation coefficient, should be used to analyze the performance and advantages and disadvantages of various fusion methods. For a certain image, if the mean value is moderate (with gray-level around 128), the visual effect is excellent. The comparison of image information entropy can reflect the detailed performance of images. The value of entropy reflects the amount of information that image holds. The larger the entropy, the more information that image holds. Sharpness can reflect the small detail contrast performance of images and the texture change features, which can be used to evaluate the fuzziness degree of images. The higher the sharpness, the clearer the image will be. The size of deviation index reflects the spectrum degrees that fusion results retain. The larger the deviation index, the greater the spectral distortion of fusion image, and the worse the fusion effects will be. Spectral distortion reflects the spectrum degree that fusion results retain. The smaller the distortion degree is, the smaller the spectral distortion will be and the better the fusion effect will be. The cross-correlation between the images before and after fusion shows the change degree of image spectral information. As multi-spectral wavebands have a large number, so the waveband mean value is taken for each index. Objective evaluation indices of the fusion results on IKONOS experimental data in Tab.1.
Table. 1 Objective evaluation indices of the fusion results on IKONOS experimental data.

|                        | IHS method | PCA method | Weighted average | HPF method | Wavelet method | Proposed method |
|------------------------|------------|------------|------------------|------------|----------------|-----------------|
| Mean value             | 65.432     | 78.199     | 69.398           | 70.298     | 70.298         | 69.175          |
| Bias index             |    —      | 0.0660     | 0.2401           | 0.2078     | 0.2290         | 0.2371          |
| Correlation coefficient|    —      | 0.7332     | 0.9312           | 0.8561     | 0.8518         | 0.8733          |
| Entropy of information | 4.3837    | 5.2366     | 5.1681           | 5.1681     | 5.4577         | 5.2402          |
| Spectral distortion degree | —    | 4.9478     | 8.1512           | 16.7367    | 10.3462        | 12.0255         |
| Definition             | 1.9717    | 2.6012     | 2.3384           | 2.3170     | 4.3961         | 3.3362          |

It can be seen from Tab.1 that the IHS transform method has a smaller spectral distortion and deviation index and a larger correlation coefficient compared with other methods, indicating that the IHS transform obtains the fusion image with a relatively small deviation from the original multispectral image. There is more of the multi-spectral information retained, but it is well maintained spatial information. The direct average fusion method obtains the fusion image with a small improvement on the spatial resolution, but there exists obvious spectral variation. By high-pass filtering, this fusion method can obtain an enhanced fusion image, but the spectrum of the fusion image exhibits serious distortion. The PCA fusion image retains the high frequency information of the original image, and the details of the target are clearer, but the image boundary has a little deficiency after PCA fusion and the geometric structure information of the fused images suffer certain losses. The wavelet transform fusion effect is higher than any of the three methods above. The adaptive sparse representation of the fusion image fusion method obtains the maximum entropy indicating that it retains more abundant spectral information. The definition and the spatial frequency is highest, showing good retention of the spatial information and the image is also clearer, whether the spatial information fusion quality or spectral quality is better than the other resulting image.

5. Conclusion
Image fusion technology is one of the key technologies in image processing. Image fusion technology has been widely used in the military, remote sensing, robotics and computer vision fields. Remote sensing image fusion algorithm based on sparse representation is proposed which is suited to the characteristics of different spatial resolution remote sensing images in the same scene. This algorithm uses both a high and a low resolution dictionary respectively for sparse representation of the multispectral image and panchromatic image. Adaptive fusion rules are used for selection of regional energy weighted coefficients, which can effectively represent significant features of source images, as well as detailed multi scale fusion image details. The experimental results show that the method retains multi-scale and directionality at the same time and that the sparse representation can effectively extract the characteristics, with lower spectral distortion. A high quality fusion image is obtained which renders the fusion image scene clearly and displays a large amount of information, which is more conducive to human visual observation. Both the subjective visual effect and objective evaluation index obtained better fusion effects, demonstrating that the proposed method is an effective and feasible method of fusion. Selection of the sparse similarity threshold \( T \) was mainly obtained experimentally in this study, therefore, the next step is to research adaptive selection methods for the optimal threshold, to enhance the algorithm usability.

When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. Should authors use tables or figures from other Publications, they must ask the corresponding publishers to grant them the right to publish this material in their paper.
Acknowledgments
This research work has been partially supported by Hexi University Foundation under grant No. XZ2014-21 and National Science Foundation of China under grant No. 41371338.

References
[1] T.D.He. Remote sensing fusion using wavelet based on contourlet transform [J]. Computer Engineering and Design, 2008, 29(24): 6302-6304.
[2] W. Strunk Jr., E.B. White, The Elements of Style, third ed., Macmillan, New York, 1979. G.Simone, A.Farina, F.C.Morabito. Image fusion techniques for remote sensing applications [J]. Information Fusion,2002, 3(1): 3-15.
[3] W.X.Yu, S.W.Yong, G.R.Guo. Review on multi-sensor information fusion technology [J]. Journal of Nation University of defense Technology, 2008, 16(3) :1-11. G.R. Mettam, How to prepare an electronic version of your article, in: B.S. Jones, R.Z. Smith (Eds.), Introduction to the Electronic Age, E-Publishing Inc., New York, 1999, pp. 281-304.
[4] P.R.Chavez, S.C.Sides, A.Anderson. Comparison of three different methods to merge multi-resolution and ultraspectral data [J]. Photogrammetric Engineering and Remote Sensing, 57(3) :2009, 295-303.
[5] H.L.Shi, T.Wei, X.J.Xin. Fusion of multispectral and panchromatic images based on PCA and NSCT [J]. Computer Engineering and Applications,2012, 48(10): 212-216.
[6] J.W.Zhang, H.Yang, J.R.Bao. Improved bayesian compressive sensing algorithm with Laplace priors [J]. Journal of Circuits and Systems, 2012, 17(1) :34-40.
[7] J.X.Zhang, W.B.Niu, K.W.Zhang. Improved image fusion algorithm based on wavelet transform [J]. Journal of Chongqing University of Technology(Natural Science), 2012, 26(1) :61-65.
[8] T.Liu, J.Chen. Remote image fusion based on wavelet and sparse representation [J]. Journal of Image and Graphics, 18(8) :2013, 1045-1053.
[9] H.L.Shi, T.Wei, X.J.Xin. Fusion of multispectral and panchromatic images based on PCA and NSCT [J]. Computer Engineering and Applications,2012, 48(10): 212-216. B.A.Olshausen, D.J.Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images[J], Nature, 381(83): 2009, 607-609.
[10] M.G.Xue, Liu C C, M.G.Xue. Infrared and light level fusion based on multi-scale dictionary [J]. Infrared Technology, 35(11) :2013, 296-701.
[11] J.Wang, J.Y.Peng, G.Q.He. Image fusion algorithm based on multi- scale dictionary learning [J]. Journal of Northwestern Polytechnical University, 31(5) :2013, 793-796. H.L.Shi, T.Wei, X.J.Xin. Fusion of multispectral and panchromatic images based on PCA and NSCT [J]. Computer Engineering and Applications,2012, 48(10): 212-216.
[12] Z.K.Ying, J.Y.Wang. A fast algorithm for image reconstruction based on sparse decomposition [J]. Journal of University of Electronic Science and Technology of China, 2006, 35(4) :448-449.
[13] C.M.Yang, B.L.Guo, M.Yi. Multi-focus image fusion using adaptive dictionary learning method [J]. Journal of Image and Graphics, 17(9):2012, 1145-1149.
[14] T.D.He, J.W.Li. Hyperspectral remote sensing image classification based on adaptive sparse representation dictionary [J]. System Engineering and Electronics, 2013, 35(9): 1194-1198.