

Image Features Extraction and Fusion Based on Joint Sparse Representation

Nannan Yu, Tianshuang Qiu, Feng Bi, and Aiqi Wang

Abstract—In this paper, a novel joint sparse representation-based image fusion method is proposed. Since the sensors observe related phenomena, the source images are expected to possess common and innovation features. We use sparse coefficients as image features. The source image is represented with the common and innovation sparse coefficients by joint sparse representation. The sparse coefficients are consequently weighted by the mean absolute values of the innovation coefficients. Furthermore, since sparse representation has been significantly successful in the development of image denoising algorithms, our method can carry out image denoising and fusion simultaneously, while the images are corrupted by additive noise. Experiment results show that the performance of the proposed method is better than that of other methods in terms of several metrics, as well as in the visual quality.

Index Terms—Features extraction, image fusion, joint sparse representation, K-SVD.

I. INTRODUCTION

IMAGE fusion is a process of combining several source images captured by multiple sensors into a fused image that contains all important information. Multisensor data often presents complementary information about the region surveyed, so image fusion provides an effective method to enable comparison and analysis of such data [1]. Image fusion is widely used in remote sensing for combining high-resolution panchromatic and low-resolution multispectral images, in medical diagnosis for obtaining images having both soft issue and hard issue information, and in target recognition for combining visible and infrared images [2].

From the 1980s to now, there have been many well-known fusion algorithms proposed. Most of these methods consist of three steps, including extracting image features, fusing these features with a certain rule, and constructing a fused image. Nikolov *et al.* [3] proposes a classification of image fusion algorithms into spatial domain and transform domain techniques. The spatial domain method uses the source image itself (or partial image) as image features while the transform coefficients

Manuscript received October 14, 2010; revised December 24, 2010; accepted January 20, 2011. Date of publication February 04, 2011; date of current version August 17, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Michael Elad.

N. Yu is with the Electronic and Information Engineering College, Dalian University of Technology, Dalian 116024, China, and also with the Electronic and Information Engineering College, Northeast Petroleum University, Daqing 163318, China (e-mail: yunnann1981@yahoo.com.cn).

T. Qiu, F. Bi, and A. Wang are with the Electronic and Information Engineering College Dalian University of Technology, Dalian 116024, China (e-mail: qutsh@dlut.edu.cn).

Digital Object Identifier 10.1109/JSTSP.2011.2112332

of the source image in some bases [such as the discrete Fourier transform (DFT), the mother wavelet and cosine bases of the discrete cosine transform (DCT)] are used as image features in the transform domain method. The fusion rules mainly include “choose-max” and “weighted average.” The wavelet-based image fusion method “wavelet maxima” [4] selects the wavelet coefficients with the largest activity level at each pixel location. Mahbubur Rahman *et al.* [2] calculate the weighted mean of wavelet coefficients of the source image for image fusion, and the weights depend on the contrast of the source images. The authors in [5] and [6] use the mean absolute values of the ICA coefficients to determine the weights, and calculate the weighted mean of the ICA coefficients for image fusion. The image fusion method based on sparse representation [7] fuses sparse coefficients with the “choose-max” rule.

In the “choose-max” rule, only the features with the largest activity level are transferred into the fused image and other features are discarded. The fused image with this rule tends to be oversharpened and less smooth. In the “weighted average” rule, all input images are taken into consideration, and some images with more information contribute more than others. Usually, the fused images using the “weighted average” rule have better performance in a particular aspect [8]. However, this rule has a serious drawback which is difficult to overcome. The input images represent the same region with different sensors, so they can be expected to possess some correlation. Each image contains common and innovation features [2], [9], [10]. With the “weighted average” rule, the fused image is constructed by the weighted mean of both the common and innovation features. In the fused image, the ratio of weights between each innovation and the common features drops. So the innovation features in the fused image appear less visible than in the source images. A detailed state is given in Section II. Intuitively, it is seen that the common and innovation features should be processed separately.

In this paper, inspired by distributed compressed sensing [11], we propose a novel image fusion method based on joint sparse representation (JSR). It can extract and separate the common and innovation features of source images and fuse these features separately. Furthermore, since sparse representation has strong ability to denoise, the proposed algorithm can simultaneously carry out denoising and fusion of the images while corrupted by additive noise [12]. The rest of the paper is organized as follows. Section II gives a detailed state of image fusion with different fusion rules. Section III presents the description of the proposed method, whereas Section IV contains experimental results obtained by using the proposed method and a comparison with the state-of-the-art methods.

TABLE I
AVERAGE VALUES OF THE CORRELATIONS BETWEEN THE LOCAL NEIGHBORING PIXELS OF THE SOURCE IMAGES

Image	Multisensor images 114	Multisensor images 049	Multisensor images 051	CT and MR	UnCamp	SAR	Trees
Correlation	0.4625	0.6204	0.4501	0.1420	0.3324	0.5492	0.2361

II. IMAGE FUSION WITH THE “CHOOSE-MAX” AND “WEIGHTED AVERAGE” RULES

The image fusion process consists of three main steps: extracting features of each input image, following a specific rule, combining them into composite features, and constructing a fused image.

Let $x_i (i \in [1, K])$ represent the source image and θ_i represent image features. x_i is a function of θ_i , that can be written as

$$x_i = f(\theta_i). \quad (1)$$

In the spatial domain image fusion, θ_i is the i th source image itself (or partial image), while in the transform domain image fusion, θ_i is the coefficient sequence of source image in some bases. With the “choose-max” rule, the fused image can be constructed from the features with largest activity level. This rule is prone to make the fused image oversharpen and less smooth, and a lot of information is discarded, because only the features with the largest activity level are transferred into the fused image. The “weighted average” rule is an extension of the “choose-max” rule. Let $w_i (\sum_{i=1}^K w_i = 1)$ represent the corresponding weight, then with the “weighted average” rule, the fused image x_f can be obtained by

$$x_f = f \left(\sum_{i=1}^K w_i \theta_i \right). \quad (2)$$

Since the sensors presumably observe related phenomena, the ensemble of signals they acquire can be expected to possess some joint structure, or correlation [10], [11]. Table I shows the average values of the correlations obtained from the local neighboring pixels using a 8×8 window for a few source images. It is seen from Table I that a significant correlation exists among the source images.

Because of the correlation among the ensemble of images, the features of each image are generated as combination of two components: the common component θ^C , which is present in all images, and the innovation component θ_i^U , which is unique to each image. So

$$\theta_i = \theta^C + \theta_i^U. \quad (3)$$

The (2) can be expressed as

$$x_f = f \left(\sum_{i=1}^K w_i (\theta^C + \theta_i^U) \right) = f \left(\theta^C + \sum_{i=1}^K w_i \theta_i^U \right) \quad (4)$$

According to (1), (3), and (4), between the innovation and common components of image features, the ratio of the weights is smaller than 1 (equal to 1 only if other weights are 0) in the fused image while the ratio is 1 in the source image. So the innovation features in the fused image are more blurred than in the source image. In addition, the weight w_i should not be determined by θ_i . It should be determined by θ_i^U . In order to over-

come the drawback, the common and innovation features must be separated before calculating the weighted mean of the source image features. JSR is verified to be a good method to solve this problem, while the source images are sparse or compressible [13].

III. FUSION OF NOISY IMAGES WITH JOINT SPARSE REPRESENTATION

Let the pixels of the ideal image x_i to be fused be corrupted by an additive zero-mean white and homogeneous Gaussian noise n_i with known variances σ_i^2 . The measured image y_i is thus

$$y_i = x_i + n_i. \quad (5)$$

The authors in [13] for the first time gave the term “Joint Sparsity”—the sparsity of the entire signal ensemble. Three joint sparsity models (JSMs) that apply in different situations are presented: JSM-1 (sparse common component + innovations), JSM-2 (common sparse supports) and JSM-3 (nonsparse common + sparse innovations) [10], [11]. JSM-1 is more suitable to solve the problem of image fusion. Because of existing of joint sparsity among the source images, we can use JSR to extract the common and innovation features, then combine these features separately for image fusion. The proposed method starts with training the overcomplete dictionary from the entire signal ensemble by K-SVD [14]. Then, the common and innovation coefficients are found and denoised simultaneously by JSR, and the fused coefficients are evaluated from these denoised coefficients by using a suitable rule. Finally, the fused coefficients are inverse transformed to obtain the fused image.

A. Dictionary Training With K-SVD

There are two methods to choose an overcomplete dictionary. The first one is using a prespecified transform matrix, such as overcomplete wavelets, curvelets, contourlets, and short-time Fourier transforms, which leads to simple and fast algorithms for the evaluation of the sparse representation. The success of such dictionaries depends on how suitable they are to describe the signals sparsely. The second one is designing dictionaries based on training, such as PCA [15], MOD [16], and K-SVD [14], considered having the potential to outperform commonly the fixed dictionaries. The K-SVD algorithm has been widely recognized in [17]–[19], so in this paper we propose to learn a dictionary with K-SVD. Since the K-SVD dictionary learning process has in it a noise rejection capability (see experiments reported in [14]), one can get a clear dictionary from the noisy images.

Let $y_i \in R^{M_1 \times M_2}$ represent the corrupted image. By sliding window technique, each image is divided into $\sqrt{N} \times \sqrt{N}$ patches $y_i(t) (t \in [1, L], L = (M_1 - \sqrt{N} + 1)(M_2 - \sqrt{N} + 1))$. Each patch is ordered lexicographically as vector $v_i(t)$. Assume that the vectors responding to all the patches in image

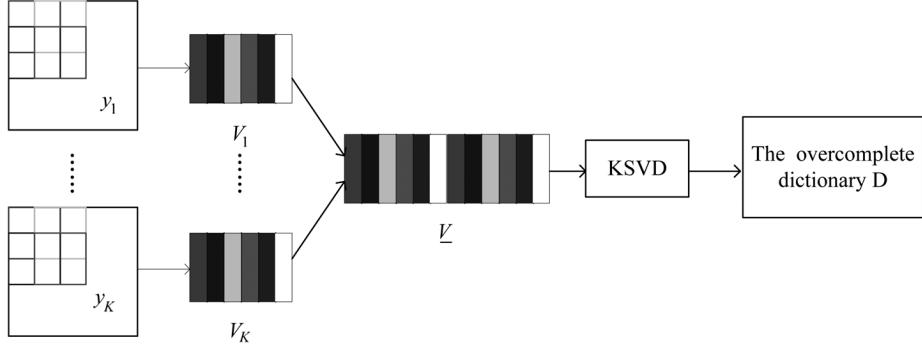


Fig. 1. Dictionary training based on K-SVD.

\$y_i\$ are constituted into one matrix \$V_i\$ and \$\underline{V} = [V_1 V_2 \cdots V_K]\$. The overcomplete dictionary \$D \in R^{N \times M}\$ is trained from the joint-matrix \$\underline{V}\$ (shown in Fig. 1), by approximating the solution of (6) with the K-SVD algorithm:

$$\min_{D, \underline{\theta}} \left\{ \| \underline{V} - D\underline{\theta} \|_F^2 \right\} \quad \text{subject to } \forall t, \| \underline{\theta}(t) \|_0 \leq T_0 \quad (6)$$

where \$\underline{\theta}\$ is the sparse coefficient matrix, \$\underline{\theta}(t)\$ is the \$t\$th column vector of \$\underline{\theta}\$, and \$T_0\$ stands for the count of the nonzero entries in \$\underline{\theta}(t)\$.

B. Features Extraction by JSR

In JSM1 model, all signals share a common component while each individual signal contains an innovation component; that is,

$$V_i = V^C + V_i^U \quad (7)$$

where \$V^C\$ and \$V_i^U\$ are the common component and innovation component of \$V_i\$. We can represent \$V_i\$ with common component \$\theta^C\$ and innovation component \$\theta_i^U\$ of sparse coefficient matrixes, and noise \$n_i\$:

$$V_i = V^C + V_i^U = D\theta^C + D\theta_i^U + n_i. \quad (8)$$

The concatenated source images matrix can be represented sparsely by the concatenated coefficient matrix

$$\begin{bmatrix} V_1 \\ \vdots \\ V_K \end{bmatrix} = \begin{bmatrix} D & D & 0 & \cdots & 0 \\ D & 0 & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D & 0 & 0 & \cdots & D \end{bmatrix} \begin{bmatrix} \theta^C \\ \theta_1^U \\ \vdots \\ \theta_K^U \end{bmatrix} + \begin{bmatrix} n_1 \\ \vdots \\ n_K \end{bmatrix}. \quad (9)$$

Let

$$V = \begin{bmatrix} V_1 \\ \vdots \\ V_K \end{bmatrix}, \quad \underline{D} = \begin{bmatrix} D & D & 0 & \cdots & 0 \\ D & 0 & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D & 0 & 0 & \cdots & D \end{bmatrix}$$

$$\theta = \begin{bmatrix} \theta^C \\ \theta_1^U \\ \vdots \\ \theta_K^U \end{bmatrix}, \quad \text{and } n = \begin{bmatrix} n_1 \\ \vdots \\ n_K \end{bmatrix}.$$

Then, (9) can be simplified, as follows:

$$V = \underline{D}\theta + n \quad (10)$$

The estimator for denoising the sparse coefficients is built by solving

$$\begin{aligned} \hat{\theta}(t) &= \arg \min_{\theta(t)} \| \theta(t) \|_0 \\ \text{subject to } & \| \underline{D}\theta(t) - V(t) \|_2^2 \leq T \end{aligned} \quad (11)$$

where \$V(t)\$ and \$\theta(t)\$ are the \$t\$th column of \$V\$ and \$\theta\$, and \$T\$ is dictated by \$\sigma_i\$ [17]. The approximate solutions can be calculated with basis pursuit (BP) [20], which suggests a convexification of the problem by replacing the \$l_0\$-norm with an \$l_1\$-norm. In this paper, we use the orthonormal matching pursuit (OMP) [21] to get an approximation solution of (11) because of its simplicity and fast execution. The common and innovation components of the source images can be reconstructed by \$\hat{\theta}^C\$ and \$\hat{\theta}_i^U\$. An example of separating the common and innovation components of the source images is given in Fig. 2. It can be seen that the common and innovation features are sufficiently well extracted and separated. For example, the white square and the trees on the mountain are extracted as common features, while the trees in the distance, the flowers in the lawn, the pedestrian, and the roof of the house are extracted as innovation features.

Then, we compare the computational complexity of sparse coding by sparse representation with by JSR. If all \$\theta_i, i \in [1, K]\$ are calculated separately as the fusion method based on sparse representation in [7], the aggregate computational complexity with OMP is \$T_{SR} \approx KT_0LMN\$. In this paper, all \$\theta_i, i \in [1, K]\$ are calculated using JSR with OMP, the computational complexity is \$T_{JSR} \approx T'LMN\$. It is verified in [10] that the joint sparsity is often smaller than the aggregate over individual image, i.e., \$T' \leq KT_0\$ (\$T' = KT_0\$, while there is no common component at all in the ensemble of source images), so the JSR offers a reduction in the computational complexity.

C. Reconstruction

Image fusion is the process of detecting salient features in the source images and fusing these details to a synthetic image. In order to account for the contributions from all possible image to

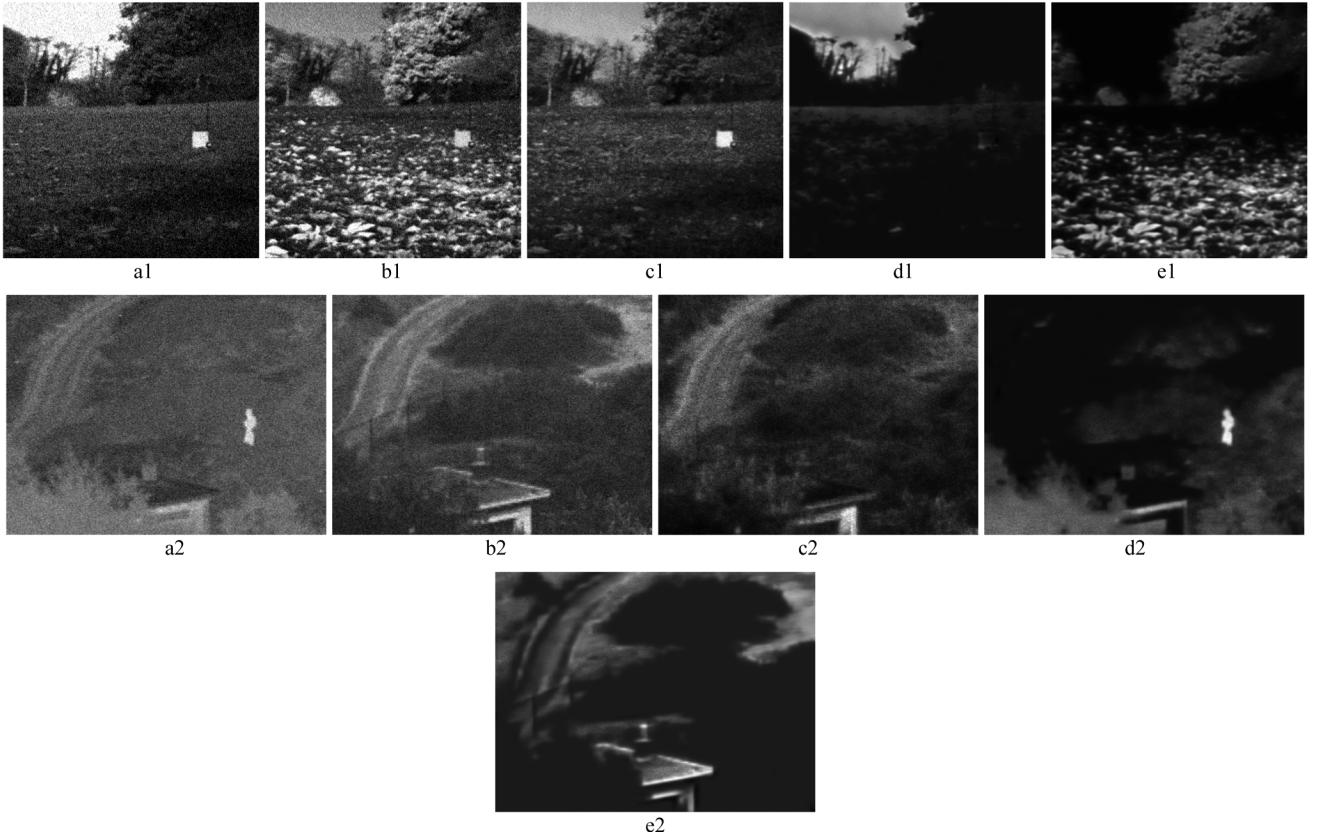


Fig. 2. Separation of the common and innovation features using multisensor images 051 and UnCamp corrupted with $\sigma = [20, 15]$. (a1), (b1), (a2), (b2) the source images, (c1) and (c2) the images with common features, (d1) and (d2) the images with innovation features of (a1), and (a2), (e1) and (e2) the images with innovation features of (b1) and (b2).

be fused, the final denoised estimated of the sparse coefficient vector of the fused image patch is written as

$$\theta_f(t) = \frac{1}{w^C(t)} \left(w^C(t)\theta^C(t) + \sum_{i=1}^K w_i^U(t)\theta_i^U(t) \right) \quad (12)$$

where $w^C(t)$ and $w_i^U(t)$ ($w_i^U(t) > 0, \sum w_i^U(t) = 1$) are the weight factors. Assume $m = \arg \max_{1 \leq i \leq K} (w_i^U(t))$, then $\theta_m^U(t)$ is the primary feature and others are the subsidiary features. The ratio of the weights between the primary and common features should be 1, so $w^C(t)$ is set as $w^C(t) = w_m^U(t)$. It seems that the projection that forms the result of fusion will have some scale deformation. However, because of sparsity of the coefficient vector, we can prove our method will not result in scale deformation. A detailed proof is given in the Appendix.

In [7] and [8], the mean absolute value (l_1 -norm) of each patch (arranged in a vector) in the transform domain has been calculated as an activity indicator. Following the method in [7] and [8], we can use the l_1 -norm of the innovation coefficient vector as the activity indicator in each patch:

$$n_i(t) = \|\theta_i^U(t)\|_1. \quad (13)$$

The weights $w_i^U(t)$ should emphasize sources that feature more intense activity, as represented by $n_i(t)$:

$$w_i^U(t) = n_i(t) / \sum_{i=1}^K n_i(t). \quad (14)$$

Then, (12) can be simplified as follows:

$$\begin{aligned} \theta_f(t) &= \theta^C(t) + \frac{1}{n_{\max}(t)} \sum_{i=1}^K n_i(t)\theta_i^U(t) \\ \text{subject to } n_{\max}(t) &= \max_{1 \leq i \leq K} (n_i(t)). \end{aligned} \quad (15)$$

The fused image matrix can be reconstructed by

$$V_f = D\theta_f. \quad (16)$$

Finally, we can transform the matrix V_f to the image patches $x_f(t)$ ($t \in [1, L]$) and synthesize the fused image x_f by averaging the image patches in the same order they were selected during the analysis step.

IV. EXPERIMENTS

In this section, the proposed fusion method is compared with three state-of-the-art methods, including discrete wavelet transform (DWT) [4], the ICA [8], and sparse representation (SR)-based [7] algorithms. In the DWT method, a five-level decomposition is used and the fusion is performed by selecting the coefficients with maximum absolute values. The ICA method are trained using 10 000 training patches selected randomly from a set of images with similar content, and 40 out of the most significant bases obtained by training are selected using the PCA algorithm. The SR applies the fixed dictionary—an overcomplete

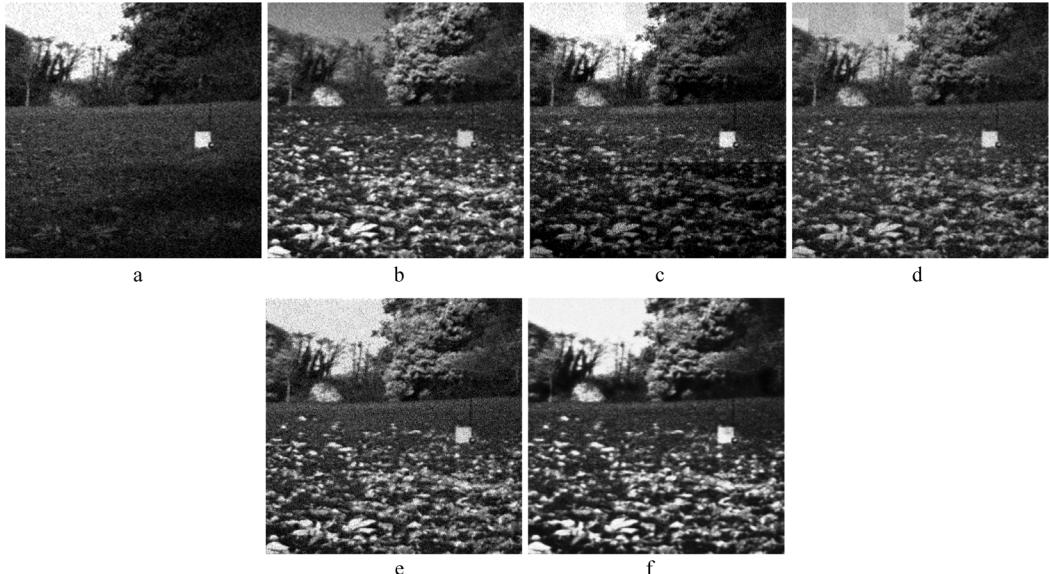


Fig. 3. Visual comparison of the performance of the fusion methods using multisensor images 051 corrupted with $\sigma = [20, 15]$. (a) and (b) The source images. (c) Fused image, DWT fusion method. (d) Fused image, ICA Fusion method. (e) Fused image, SR fusion method. (f) Fused image, the proposed method.

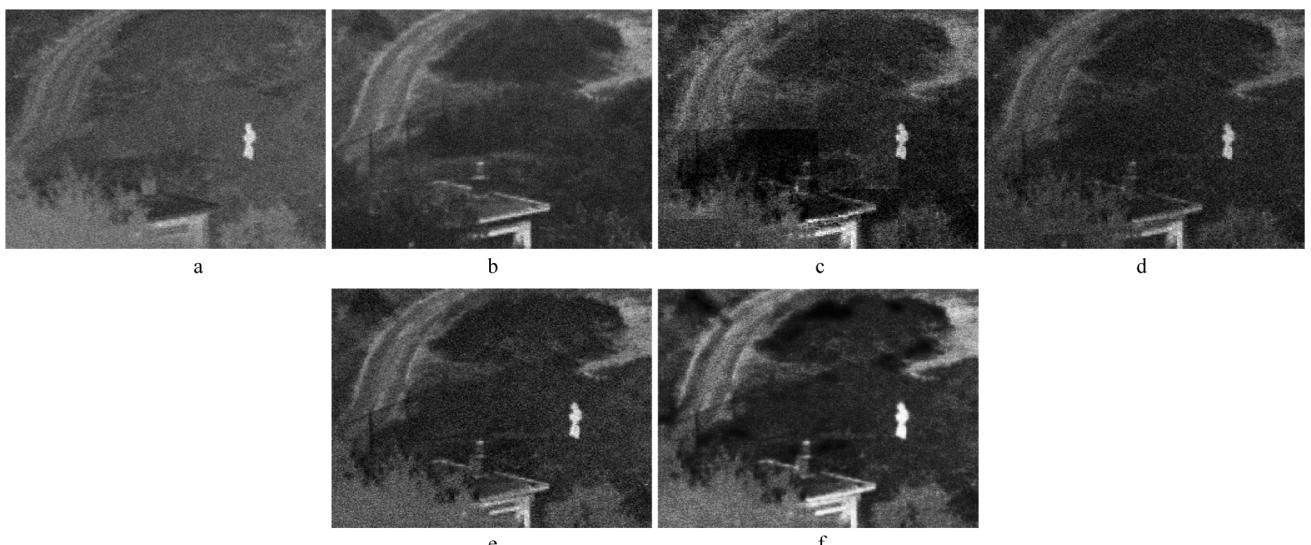


Fig. 4. Visual comparison of the performance of the fusion methods using IR and visible images UnCamp corrupted with $\sigma = [20, 15]$. (a) input IR image. (b) Input visible image. (c) Fused image, DWT fusion method. (d) Fused image, ICA fusion method. (e) Fused image, SR fusion method. (f) Fused image, the proposed method.

separable version of the DCT dictionary and the sparse coefficients are estimated by OMP. Our method trains an overcomplete dictionary of size 64×500 by K-SVD, and the maximum number of iterations is set to 100. For a fair comparison, the source images are divided into small patches of size 8×8 using sliding window technique in the ICA, SR, and JSR methods. The Piella metric [22] and Petrovic metric [23] are used to evaluate the performance of these fusion methods. These metrics are calculated using the fused images and the corresponding noise-free source images. The closer to 1 the Piella metric and Petrovic metric are, the better the fusion result will be. All the experiments are implemented in Matlab 7.9.0 and on a Pentium (R) 2.5-GHz PC with 2-GB RAM.

First, we give results concerning the experiments that have been conducted using three sets of representative images from the Image Fusion Server [24]. The noisy versions of the images

of all the test sets are obtained by adding zero mean Gaussian noise sequences to the gray images. Fig. 3 shows the multisensor images 051 of size 256×256 , and their fused outputs using the DWT, ICA, SR and the proposed one. As to the performance of image fusing, visual (subjective) comparison between methods indicates that our method is superior to the DWT, ICA, and SR fusion methods and at same time, the noise has been removed significantly. In Fig. 3, it is clear that the fine details from the source images (such as lawn borders and the flowers) are far better transferred into the fused image in the proposed method than in other methods. In addition, the trees in the top left corner of the image is visually more pleasing in the fused image obtained by the proposed method whereas all the other algorithms have transferred the blurring caused by noise.

Fig. 4 depicts the two input images from the TNO UN Camp image sequence, and their fused outputs using the DWT, ICA,

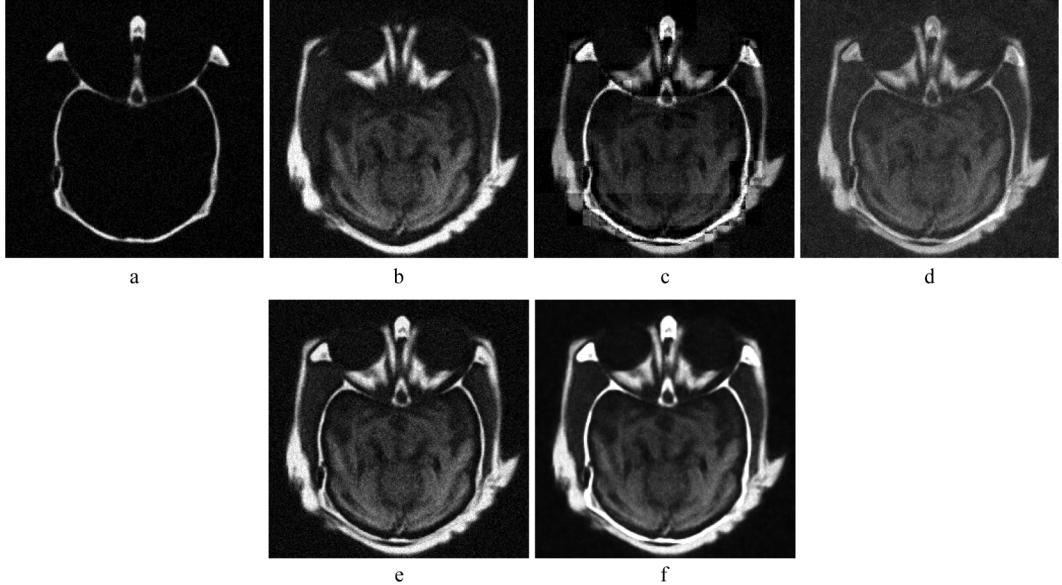


Fig. 5. Visual comparison of the performance of the fusion methods using medical images corrupted with $\sigma = [20, 15]$. (a) Input computed tomography (CT) image. (b) Input magnetic resonance (MR) image. (c) Fused image, DWT fusion method. (d) Fused image, ICA fusion method. (e) Fused image, SR fusion method. (f) Fused image, the proposed method.

TABLE II
PERFORMANCE OF IMAGE FUSION METHODS BY THE STANDARD FUSION METRICS

Image	Multisensor images 051		UnCamp		CT and MR	
	Metric	Piella	Petrovic	Piella	Petrovic	Piella
Noise standard deviation	$\sigma = [10, 5]$		$\sigma = [10, 5]$		$\sigma = [10, 5]$	
DWT	0.8894	0.8721	0.8734	0.8620	0.7025	0.7412
ICA	0.9050	0.8817	0.8950	0.8605	0.7148	0.8304
SR	0.9187	0.8874	0.9058	0.8678	0.7044	0.9025
JSR	0.9263	0.8969	0.9263	0.8721	0.7557	0.9217
Noise standard deviation	$\sigma = [20, 15]$		$\sigma = [20, 15]$		$\sigma = [20, 15]$	
DWT	0.8443	0.8633	0.8286	0.8535	0.6577	0.7326
ICA	0.8504	0.8634	0.8402	0.8428	0.6595	0.8122
SR	0.8995	0.8809	0.8863	0.8603	0.6859	0.8957
JSR	0.9077	0.8899	0.9078	0.8658	0.7407	0.9144

SR and the proposed one. The size of images is 320×240 . It is evident that the proposed method maintains the image features while reducing the noise significantly when compared with other methods. For example, the walking person in the IR image is better transferred into the fused image and the roof of the house is clearer in the proposed method compared to other state-of-the-art methods. Fig. 5 shows two medical images CT and MR, and their fused images obtained using the DWT, ICA, SR methods, along with the proposed method. The size of images is 256×256 . It can be seen the proposed method is less influenced by noise compared to DWT, ICA, and SR methods. The structure of bones of fused images by ICA and DWT is clear, but the contrast is reduced to some extent and their edges are not easily distinguished. For DWT, their edges are less smooth. The method SR has the better result than ICA and DWT, but the structure of bones of fused images is less

clear than that by our method. The fused images by our method can distinguish the soft and hard tissues easily, and reserve the details and edges completely. The proposed method provides the result with the best visual appearance.

The corresponding objective measures the Piella metric and Petrovic metric between the source and fused images of four methods are listed in Table II. The highest quality measures obtained over all methods are indicated by the values in bold. It can be seen from this table that the fusion performance degrades for all the methods with increasing strength of the noise. From Table II, we also see that, our method provides higher values of the Piella metric and Petrovic metric compared with other methods. The Petrovic metric evaluates the amount of edge information that is transferred from input images to the fused image and the Piella metric evaluates the correlation, luminance distortion and contrast distortion simultaneously.



Fig. 6. Twenty pairs of test images from Image Fusion Server.

TABLE III
MEAN VALUE OF THE METRIC OVER 20 PAIRS OF IMAGES

Metric	Method	$\sigma = [10, 5]$	$\sigma = [20, 15]$
Piella	ICA	0.9045	0.8514
	SR	0.9199	0.9003
	JSR	0.9278	0.9048
Petrovic	ICA	0.8819	0.8615
	SR	0.8834	0.8817
	JSR	0.8999	0.8907

TABLE IV
MEAN VALUE OF CPU TIME OF THREE METHODS

Method	ICA	SR	JSR
CPU time (seconds)	65.02	107.88	121.13

Thus our method not only acquires the fused image of high quality but also guarantees the correlation between the source and fused images.

To further confirm the effectiveness of the proposed method, 20 pairs of multisensor images 001–020 from Image Fusion Server are fused by ICA, SR, and the proposed method, as shown in Fig. 6. In addition, the images are added Gaussian noise $\sigma = [20, 15]$ and $\sigma = [10, 5]$. The fused images are evaluated by the Piella metric and Petrovic metric, listed in Table III. The values demonstrate that the proposed method is effective and superior to other methods.

Table IV reports the average computation (CPU) time of the 20 pairs of images required by the three methods. The average CPU time of JSR is longer than that of ICA and SR. Although sparse coding need less time in the JSR method than in the SR method, the trained dictionary spends more time than the fixed dictionary. So the JSR is slightly slower than the SR for image fusion. With ever growing computational capabilities, computational cost may become secondary in importance to the improved performance.

V. CONCLUSION

In this paper, we have presented a novel image fusion method based on joint sparse representation. The proposed method can overcome the drawback of the “weighted average” fusion rule. We extract common and innovation features of source images simultaneously by joint sparse representation, making use of the joint-sparsity of all source images. The common and innovation features are fused separately, and the sparse coefficients are consequently weighted by the mean absolute values of the innovation coefficients. The experimental results show that the proposed scheme has better fusion performance than the state-of-the-art algorithms.

APPENDIX

In this Appendix, we want to prove our algorithm will not result in scale deformation. As discussed in Section III, the sparse coefficients of the fused image are obtained by

$$\theta_f(t) = \frac{1}{w^C(t)} \left(w^C(t) \theta^C(t) + \sum_{i=1}^K w_i^U(t) \theta_i^U(t) \right)$$

TABLE V
VALUES RANGE OF $\theta_{fp}(t)$ WITH DIFFERENT VALUES OF $\theta_{1p}(t)$ AND $\theta_{2p}(t)$

$\theta_{1p}(t), \theta_{2p}(t)$	$\{\theta^{Cp}(t), \theta_1^{Up}(t), \theta_2^{Up}(t)\}$	$\theta_{fp}(t)$	The values range of $\theta_{fp}(t)$
$\theta_{1p}(t) = \theta_{2p}(t) = 0$	$\{0, 0, 0\}$	0	
$\theta_{1p}(t) \neq 0$ and $\theta_{2p}(t) = 0$	$\{0, \theta_{1p}(t), 0\}$	$\theta_{fp}(t) = g(t)\theta_{1p}(t)$ or $\theta_{fp}(t) = \theta_{1p}(t)$	
$\theta_{1p}(t) > \theta_{2p}(t) > 0$	$\{\theta_{2p}(t), \theta_{1p}(t) - \theta_{2p}(t), 0\}$	$\theta_{fp}(t) = \theta_{2p}(t) + g(t)$ $(\theta_{1p}(t) - \theta_{2p}(t))$ or $\theta_{fp}(t) = \theta_{1p}(t)$	$\theta_{fp}(t) \in [\theta_{1p}(t), \theta_{2p}(t)]$ or $\theta_{fp}(t) \in [\theta_{2p}(t), \theta_{1p}(t)]$
$\theta_{2p}(t) < \theta_{1p}(t) < 0$	$\{\theta_{1p}(t), 0, \theta_{2p}(t) - \theta_{1p}(t)\}$	$\theta_{fp}(t) = \theta_{2p}(t)$ or $\theta_{fp}(t) = \theta_{1p}(t) + g(t)$ $(\theta_{2p}(t) - \theta_{1p}(t))$	
$\theta_{2p}(t) < 0 < \theta_{1p}(t)$	$\{0, \theta_{1p}(t), \theta_{2p}(t)\}$	$\theta_{fp}(t) = g(t)\theta_{1p}(t) + \theta_{2p}(t)$ or $\theta_{fp}(t) = \theta_{1p}(t) + g(t)\theta_{2p}(t)$	

or

$$\theta_f(t) = \theta^C(t) + \frac{1}{n_{\max}(t)} \sum_{i=1}^K n_i(t) \theta_i^U(t)$$

subject to $n_{\max}(t) = \max_{1 \leq i \leq K} (n_i(t))$

where $\theta^C(t)$ and $\theta_i^U(t)$ can be calculated by (9). For purposes of analysis, we consider the situation of fusion of two images and replace l_0 -norm with l_1 -norm. Then the (9) is approximated by

$$\begin{aligned} & \left\{ \hat{\theta}^C(t), \hat{\theta}_1^U(t), \hat{\theta}_2^U(t) \right\} \\ &= \arg \min \left(\|\theta^C(t)\|_1 + \|\theta_1^U(t)\|_1 + \|\theta_2^U(t)\|_1 \right) \\ & \text{subject to } \begin{cases} \theta_1(t) = \theta^C(t) + \theta_1^U(t) \\ \theta_2(t) = \theta^C(t) + \theta_2^U(t) \end{cases}. \end{aligned} \quad (17)$$

Assume $\theta^{Cp}(t), \theta_1^{Up}(t), \theta_2^{Up}(t), \theta_{1p}(t), \theta_{2p}(t)$, and $\theta_{fp}(t)$ represent the p th ($p \in [1, M]$) elements of $\theta^C(t), \theta_1^U(t), \theta_2^U(t), \theta_1(t), \theta_2(t)$ and $\theta_f(t)$. According to (12), (15), and (17) $\theta_{fp}(t)$ can be calculated by

$$\theta_{fp}(t) = \theta^{Cp}(t) + g(t)\theta_1^{Up}(t) + \theta_2^{Up}(t)$$

$\left(\text{if } n_1(t) \leq n_2(t), \text{ then } g(t) = \frac{n_1(t)}{n_2(t)} \right) \quad (18)$

or

$$\theta_{fp}(t) = \theta^{Cp}(t) + \theta_1^{Up}(t) + g(t)\theta_2^{Up}(t)$$

$\left(\text{if } n_1(t) \geq n_2(t), \text{ then } g(t) = \frac{n_2(t)}{n_1(t)} \right) \quad (19)$

where

$$\begin{aligned} & \left\{ \theta^{Cp}(t), \theta_1^{Up}(t), \theta_2^{Up}(t) \right\} \\ &= \arg \min_{\theta^{Cp}(t), \theta_1^{Up}(t), \theta_2^{Up}(t)} \left(|\theta^{Cp}(t)| + |\theta_1^{Up}(t)| + |\theta_2^{Up}(t)| \right) \\ & \text{subject to } \begin{cases} \theta_{1p}(t) = \theta^{Cp}(t) + \theta_1^{Up}(t) \\ \theta_{2p}(t) = \theta^{Cp}(t) + \theta_2^{Up}(t) \end{cases} \end{aligned} \quad (20)$$

From (18), (19), and (20), $\theta_{fp}(t)$ is determined by $\theta_{1p}(t)$ and $\theta_{2p}(t)$. Table V shows five kinds of representative conditions of $\theta_{1p}(t)$ and $\theta_{2p}(t)$. It can be seen from this table that the value range of $\theta_{fp}(t)$ is from $\theta_{1p}(t)$ (or $\theta_{2p}(t)$) to $\theta_{2p}(t)$ (or $\theta_{1p}(t)$), and hence, our method will not result in scale deformation.

REFERENCES

- [1] N. Cvejic, T. Seppänen, and S. J. Godsill, "A nonreference image fusion metric based on the regional importance measure," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 212–220, Apr. 2009.
- [2] S. M. Mahbubur Rahman, M. Omair Ahmad, and M. N. S. Swamy, "Contrast-based fusion of noisy images using discrete wavelet transform," *IET Image Process.*, vol. 4, no. 5, pp. 374–384, 2010.
- [3] S. Nikolov, D. Bull, and N. Canagarajah, "Wavelets for image fusion," in *Wavelets in Signal and Image Analysis*. Norwell, MA: Kluwer, 2001.
- [4] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 57, no. 3, pp. 235–245, 1995.
- [5] N. Mitianoudis and T. Stathaki, "Optimal contrast correction for ICA-based fusion of multimodal images," *IEEE Sens. Journal*, vol. 8, no. 12, pp. 2016–2026, Dec. 2008.
- [6] N. Cvejic, D. Bull, and N. Canagarajah, "Region-based multimodal image fusion using ICA bases," *IEEE Sens. J.*, vol. 7, no. 5, pp. 743–751, 2007.
- [7] B. Yang and S. Li, "Multifocus image fusion and restoration with sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 4, pp. 884–891, Apr. 2010.
- [8] N. Mitianoudis and T. Stathaki, "Pixel-based and region-based image fusion schemes using ICA bases," *Inf. Fusion*, vol. 8, no. 2, pp. 131–142, 2007.
- [9] E. Candès, X. Li, Y. Ma, and J. Wright, Robust Principal Component Analysis? preprint, 2009 [Online]. Available: <http://watt.cs.illinois.edu/~perceive/matrix-rank/references.html>
- [10] D. Baron, M. Duarte, M. Wakin, S. Sarvotham, and R. Baraniuk, "Distributed compressive sensing," in *Proc. Sens., Signal, Inf. Process. Workshop*, 2008 [Online]. Available: <http://arxiv.org/abs/0901.3403>.
- [11] M. Wakin, M. Duarte, S. Sarvotham, D. Baron, and R. Baraniuk, "Recovery of jointly sparse signals from few random projections," in *Proc. Workshop Neural Inf. Process. Syst. (NIPS)*, Vancouver, BC, Canada, Dec. 2005, pp. 1435–1442.
- [12] Y. Deng, D. Li, X. Xie, and K. Lam, "Partially occluded face completion and recognition," in *Proc. Int. Conf. Image Process. (ICIP)*, Cairo, Egypt, Nov. 2009, pp. 4145–4148.
- [13] M. Duarte, S. Sarvotham, D. Baron, M. Wakin, and R. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2005, pp. 1537–1541.

- [14] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- [15] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer, 2002.
- [16] K. Engan, S. O. Aase, and J. H. Husøy, "Multi-frame compression: Theory and design," in *Proc. EURASIP Signal Process.*, 80, 2000, no. 10, pp. 2121–2140.
- [17] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.
- [18] M. Protter and M. Elad, "Image sequence denoising via sparse and redundant representations," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 27–35, Jan. 2009.
- [19] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognition*, San Francisco, CA, 2010, pp. 2691–2698.
- [20] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
- [21] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Conf. Rec. 27th Asilomar Conf. Signals, Syst. Comput.*, 1993, vol. 1, pp. 40–44.
- [22] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. Int. Conf. Image Process. (ICIP)*, Barcelona, Spain, 2003, pp. 173–176.
- [23] C. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, Feb. 2000.
- [24] The Image Fusion Server [Online]. Available: <http://www.imagefusion.org/>



Nannan Yu received the B.S. and M.S. degrees from the Harbin Institute of Technology, Harbin, China, in 2003 and 2006, respectively. She is currently working towards the Ph.D. degree in signal and information processing from Dalian University of Technology, Dalian, China.

She is currently an Assistant Professor at Northeast Petroleum University, Daqing, China. Her research interests are in image fusion, biomedical signal processing, and sparse representation.



Tianshuang Qiu received the B.S. degree from Tianjin University, Tianjin, China, in 1983, the M.S. degree from Dalian University of Technology, Dalian, China, in 1993, and the Ph.D. degree from Southeastern University, Nanjing, China, in 1996, all in electrical engineering.

He was a Research Scientist at Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China, from 1983 to 1996. He served on the Faculty of Electrical Engineering, Dalian Railway Institute, in 1996. After that, he conducted his

post-doctoral research in the Department of Electrical Engineering, Northern Illinois University, DeKalb. He is currently a Professor in the Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology. His research interests include non-Gaussian and non-stationary signal processing, radio frequency signal processing, and biomedical signal processing.



Feng Bi received the B.S. degree from Liaoning University, Shenyang, China, in 1995 and the M.S. degree from Liaoning Normal University, Dalian, China, in 2005. He is currently working towards the Ph.D. degree in signal and information processing at Dalian University of Technology, Dalian, China.

He is currently an Assistant Professor in Eastern Liaoning University, Shenyang, China. His research interests are in non-Gaussian and non-stationary signal processing, biomedical signal processing, and blind signal processing.



Aiqi Wang received the B.S. degree from the Department of Mathematics, Liaoning Normal University, Dalian, China, in 2002 and the M.S. degree from the Department of Mathematics, Dalian University of Technology, in 2005. He is currently pursuing the Ph.D. degree at the School of Electronic and Information Engineering, Dalian University of Technology.

His research interests include pattern recognition, medical image processing and computer vision.