SIRF: Simultaneous Satellite Image Registration and Fusion in a Unified Framework

Chen Chen, *Student Member, IEEE*, Yeqing Li, Wei Liu, and Junzhou Huang, *Member, IEEE*

Abstract—In this paper, we propose a novel method for image fusion with a high-resolution panchromatic image and a low-resolution multispectral (Ms) image at the same geographical location. The fusion is formulated as a convex optimization problem which minimizes a linear combination of a least-squares fitting term and a dynamic gradient sparsity regularizer. The former is to preserve accurate spectral information of the Ms image, while the latter is to keep sharp edges of the high-resolution panchromatic image. We further propose to simultaneously register the two images during the fusing process, which is naturally achieved by virtue of the dynamic gradient sparsity property. An efficient algorithm is then devised to solve the optimization problem, accomplishing a linear computational complexity in the size of the output image in each iteration. We compare our method against six state-of-the-art image fusion methods on Ms image data sets from four satellites. Extensive experimental results demonstrate that the proposed method substantially outperforms the others in terms of both spatial and spectral qualities. We also show that our method can provide high-quality products from coarsely registered real-world IKONOS data sets. Finally, a MATLAB implementation is provided to facilitate future research.

Index Terms—Image fusion, pan-sharpening, image registration, dynamic gradient sparsity, group sparsity, joint fusion.

I. INTRODUCTION

MULTISPECTRAL (Ms) images are widely used in many fields of remote sensing such as environmental monitoring, agriculture, mineral exploration, *etc*. However, the design of Ms sensors with high resolution is confined by infrastructure limits in onboard storage and bandwidth transmission [1]. In contrast, panchromatic (Pan) gray-scaled images with high spatial resolution can be obtained more conveniently because they are composed of a highly reduced numbers of pixels. The combinations of Pan images in high spatial resolution and Ms images in high spectral resolution can be acquired simultaneously from most existing satellites. Therefore, we expect to obtain images with both high spatial
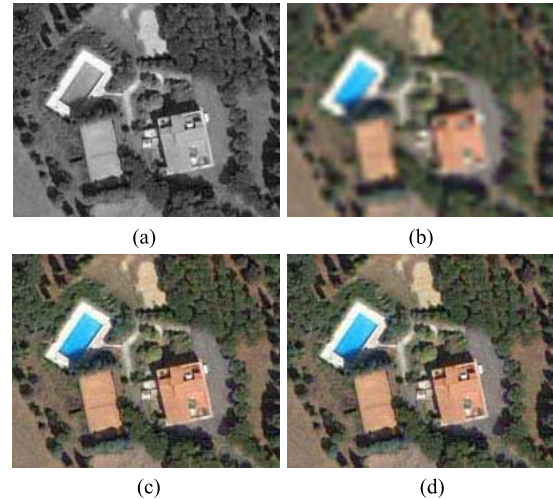
Fig. 1. (a) A high-resolution panchromatic image. (b) The corresponding low-resolution multispectral image. (c) Our fusion result. (d) The ground-truth.

and high spectral resolutions via image fusion, which is also called pan-sharpening in the literature. A fusion example taken from Quickbird satellite images is shown in Figure 1.

Image fusion is a typical inverse problem and generally difficult to handle. The first question is how to preserve accurate information from both the Pan and Ms images. A number of conventional methods use projection and substitution, which includes principal component analysis (PCA) [2], intensity hue saturation (IHS) [3], wavelet [4], and their combinations. These methods perform fusion in the following scheme: upsampling, forward transform, intensity matching, component substitution, and reverse transform [5]. Other methods such as Brovey [6] assume that the Pan image is a linear combination of all bands of the fused image. A detailed survey of the existing methods can be found in [5]. While previous classical methods have provided some good visual results, they often contain spectral distortion. The reason is that they all made strong assumptions that are not realistic from the viewpoint of remote sensing physics [1].

In order to overcome the issue caused by spectral distortion, a suite of variational approaches have emerged recently, all of which formulate an energy function based on somewhat weak assumptions, and then optimize such a function to obtain the optimum. These methods are also called model-based fusion [7]. The earliest variational method P+XS (Pan and Ms images) [8] is based on the linear combination assumption used by Brovey [6], and additionally assumes that

the upsampled Ms image is the fused result after blurring. As an accurate blur kernel is difficult to pre-estimate, AVWP (alternate variational wavelet pansharpening) [9] replaces this term with a spectral ratio constraint to preserve spectral information, and meanwhile forces the fused image close to the wavelet fused image [4]. Another variational model is engaged in estimating the fused image in conjunction with the blurring model parameters iteratively [10]. Owing to the success of compressive sensing [11], some methods are proposed to employ sparsity regularization and dictionary learning to tackle image fusion [12]–[16]. In a recent study [17], a highpass filter was introduced to compensate for the lowpass filter used for spectral observation. Promising results achieved by these variational methods have shown that they can reduce spectral distortion. However, due to the lack of an effective model to preserve spatial information, visible artifacts or blurriness may appear in the fused results. Moreover, all these methods often involve high computational complexities which prevent these methods from being scalable to massive datasets.

The second question in fusion is how to reduce the effect of misalignments. Almost all the above methods require a precise registration before fusion can take place. However, pre-registration is quite challenging due to the significant resolution difference between input images [18]–[20]. After pre-registration, 0.5-pixel's misalignment on the multi-spectral image corresponds to 2-pixel misalignment on the Pan image when the resolution difference is four times. It has been shown that a geometrical distortion of only 0.1 pixel in standard deviation yields a significant impact on pixel-to-pixel fusion [21]. The existing fusion methods have to "tolerate" such misalignments. Therefore, fusion accuracy is inevitably compromised for real-world datasets.

In this paper, we propose a novel method for *Simultaneous Image Registration and Fusion*, SIRF. We assume that the fused image after downsampling should be close to the input Ms image, which is formulated into a least-squares fitting term to keep spectral information. Motivated by the geographical relationship between the fused image and the input Pan image, a dynamic gradient sparsity property is discovered, defined, and then exploited to improve spatial quality. Importantly, we find that the combined model does not violate remote sensing physics, while the "linear combination" in many previous methods do. Therefore, the dynamic gradient sparsity naturally induces more accurate registration even under severe intensity distortions. Moreover, our method incorporates the inherent correlation of different bands, which has been rarely considered before. To optimize the entire energy function, a new algorithm is designed based on the proximal gradient technique. In specific, we solve the subproblems efficiently by applying the fast iterative shrinkage-thresholding algorithm (FISTA) [22] and the gradient descent method with backtracking, respectively. The overall algorithm remains a linear computational complexity in each iteration, and is thus scalable to massive datasets. The algorithm can directly be applied to real-world datasets without prefiltering, feature extraction, training, *etc*. Finally, there is only one non-sensitive parameter in SIRF, which is another advantage compared to existing variational methods. Extensive experimental results

demonstrate that our method can significantly reduce spectral distortions and improve the overall image quality in terms of different image quality metrics. In particular, our method is shown to be much more powerful and can preserve sharp object boundaries on real-world datasets with pre-registration errors.

The rest of this paper is organized as follows. We define the dynamic gradient sparsity and give the formulation of SIRF in Section II. In Section III, an efficient algorithm is presented to solve the optimization problem. The experimental results are shown in Section IV. Finally, we conclude the paper in Section V. Some preliminary results of this work have been presented in [23].

II. MODELING

A. Notations

Scalers are denoted by lowercase letters. Bold letters denote matrices. Specially, $\mathbf{P} \in \mathbb{R}^{m \times n}$ denotes the Pan image and $\mathbf{M} \in \mathbb{R}^{\frac{m}{c} \times \frac{n}{c} \times s}$ denotes the low-resolution Ms image, where m and n denote the numbers of rows and columns, s denotes the number of bands. c is a constant. For example $c = 4$ when the resolution of the Pan image is 0.6 m and that of the Ms image is 2.4 m in Quickbird acquisition. The image to be fused is denoted by $\mathbf{X} \in \mathbb{R}^{m \times n \times s}$. For simpleness, $\mathbf{X}_{i,j,d}$ denotes the element in i-th row, j-th column and d-th band in $\mathbf{X}$. And $\mathbf{X}_d$ denotes the whole d-th band, which is therefore a matrix. $|| \cdot ||_F$ denotes the Frobenius norm for matrices and $||\mathbf{X}||_F^2 = \sum_{d=1}^{s} ||\mathbf{X}_d||_F^2$ is defined for multi-channel data.

B. Local Spectral Consistency

Many existing methods upsample the Ms image and extract spectral information from this upsampled Ms image. However, the upsampled image is blurry and not accurate. Therefore, we only assume the fused image after downsampling is close to the original Ms image. Least squares fitting is used to model this relationship:

$$E_1 = \frac{1}{2}||\psi \mathbf{X} - \mathbf{M}||_F^2, \qquad (1)$$

where ψ denotes a downsampling operator. Local spectral information is forced to be consistent with each Ms pixel. This function can significantly reduce the spectral distortion in the result.

Minimizing E_1 would be a severely ill-posed problem, due to the very low undersampling rate (*e.g.*, 1/16 when $c = 4$). Without strong prior information, $\mathbf{X}$ is almost impossible to be estimated accurately.

C. Dynamic Gradient Sparsity

Fortunately, the Pan image provides such prior information. Due to the strong geographical correlation with the fused image $\mathbf{X}$, the Pan image has already provided us with clear edge information of land objects. Many researchers attempt to build this relationship mathematically. From recent reviews [1], [5], however, few methods can effectively characterize this relationship.

As remotely sensed images are often piece-wise smooth, their gradients tend to be sparse and the non-zeros corresponds to the edges. In addition, the positions of such edges should be the same as those on the Pan image when the images have been well aligned. It demonstrates that the sparsity property is not fixed but dynamic according to a reference image. This property has not been studied in sparsity theories yet. We call the data with such a property a dynamic gradient sparse signal/image.

Definition: Let $x \in \mathbb{R}^N$ and $r \in \mathbb{R}^N$ denote the signal and the reference signal. Ω_x and Ω_r denote the support sets[1] of their gradients, respectively. In particular, r only has K non-zero gradients with $K \ll N$. The set of dynamic gradient sparse signals $\mathcal{S}_x$ is defined as:

$$\mathcal{S}_x = \{x \in \mathbb{R}^N : |\Omega_x| = K, \Omega_x = \Omega_r, \text{ with } K \ll N\}.$$

Similar logic can be extended to multi-channel/spectral signals and images. The first term in P+XS [8] and AVWP [9] does not induce sparseness and tends to over-smooth the image by penalizing large values. In [10], the first term is derived from the linear combination assumption in P+XS; it does not promote sparsity for each band. In [15], total variation is used to encourage sparseness of the gradients. However, no reference information from the Pan image is integrated in this regularization term. Different from previous work, dynamic gradient sparsity is encouraged in our method. In addition to the prior information that previous methods attempt to use, we also notice the intra- correlations across different bands. As all bands are representations of the same land objects, their gradients (which correspond to the edges of the land objects) tend to be at the same spatial positions. Therefore, they should not only be sparse but also group sparse in the gradient domain. The pixels with the same spatial position across different bands are assigned into one group. It is widely known that the ℓ_1 norm encourages sparsity and the $\ell_{2,1}$ norm encourages group sparsity [24]. Thus we propose a new energy function to encourage dynamic gradient sparsity and group sparsity simultaneously:

$$E_2 = ||\nabla\mathbf{X} - \nabla\bar{\mathbf{P}}||_{2,1} \qquad (2)$$

$$= \sum_i \sum_j \sqrt{\sum_d \sum_q (\nabla_q \mathbf{X}_{i,j,d} - \nabla_q \mathbf{P}_{i,j})^2}, \qquad (3)$$

where $q = 1, 2$ and $\bar{\mathbf{P}}$ means duplicating $\mathbf{P}$ to s bands. Interestingly, when there is no reference image, *i.e.*, $\mathbf{P} = \mathbf{0}$, the result is identical to that of vectorial total variation (VTV) [25]–[27], which is widely used in multi-channel image denoising, deblurring, and reconstruction.

To demonstrate why E_2 encourages dynamic gradient sparsity, we show a simple example on a 1D multi-channel signal in Figure 2. We could observe that, if the solution has a different support set from the reference, the total sparsity of the gradients will be increased. Cases (a)-(d) have group sparsity number 8, 4, 4, 2 respectively. Therefore, (a)-(c) will be penalized because they are not the sparsest solution in our method.

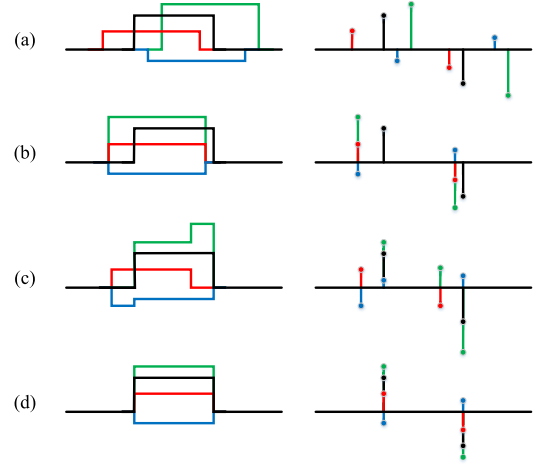[1] Here we mean the indices of the non-zero components.

Fig. 2. Illustration of possible solutions for different gradient-based penalties. The black denotes a reference signal. RGB color lines denote the solutions of different models. Left: 1D signals. Right: the corresponding gradients. (a) A possible solution of TV [15]: the gradients of RGB channels are sparse but may not be correlated. (b) A possible solution of VTV: the gradients of R, G, B channels are group sparse, but may not be correlated to the reference signal. (c) A possible solution of [10], which does not encourage sparseness for each channel individually. (d) A possible solution of dynamic gradient sparsity regularization: The gradients can only be group sparse following the reference.

Combining the two energy functions, the image fusion problem can be formulated as:

$$\min_{\mathbf{X}}\{E(\mathbf{X}) = E_1 + \lambda E_2$$
$$= \frac{1}{2}||\psi\mathbf{X} - \mathbf{M}||_F^2 + \lambda||\nabla\mathbf{X} - \nabla\bar{\mathbf{P}}||_{2,1}\}, \qquad (4)$$

where λ is a positive parameter.

Comparing our method with existing methods, the first benefit of our method comes from the local spectral constraint. It does not rely on the upsampled Ms image and linear-combination assumption. Therefore, only accurate spectral information is kept. Second, the proposed dynamic gradient sparsity only forces the support sets to be the same, while the signs of the gradients as well as the magnitudes of the signal are not required to be the same. These properties make it invariant under contrast inversion [1] and they are not sensitive to illumination conditions. It is possible to be applied to image fusion from different sources or at different acquisition times. Last but not least, our method can jointly fuse multiple bands simultaneously, while most existing method can only fuse the image band-by-band, e.g., [8]–[10]. which provides robustness to noise [33], [34]. These advantages make our method very effective on different data.

D. Simultaneous Image Registration

The image fusion process (4) requires accurate registration between the Pan image and the multispectral image, while the misalignment is very difficult to eliminate during preprocessing. To overcome this difficulty, we propose to register the images simultaneously during the fusion process. On one hand, the multispectral image is sharpened to higher resolution. This allows us to register the images more accurately. On the other hand, the misalignment is gradually
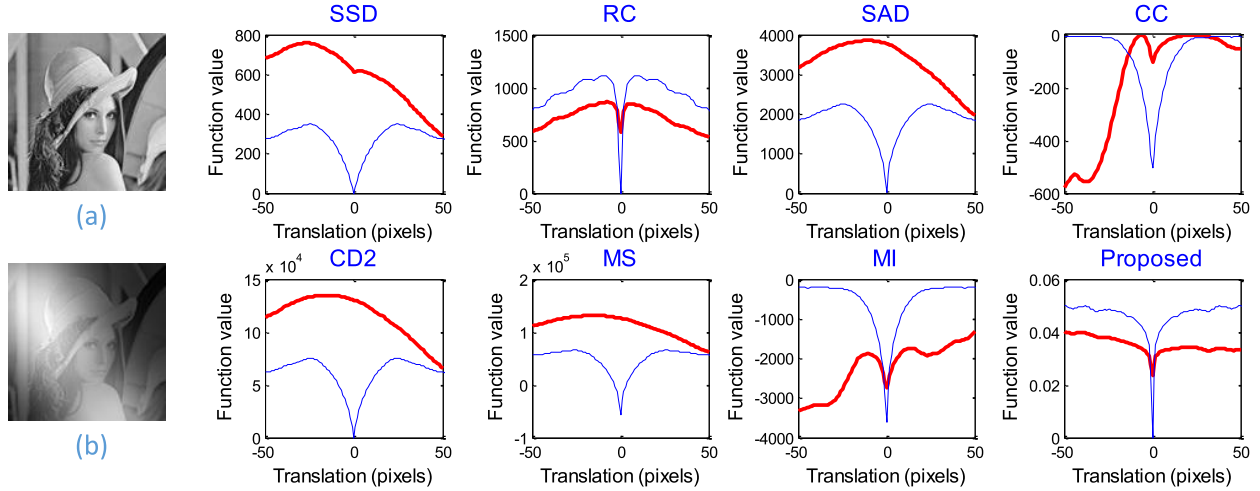
Fig. 3. A toy registration example with respect to 1D translation using different similarity measures (SSD [28], RC [29], SAD, CC, CD2 [30], MS [31], MI [32] and the proposed method based on normalized dynamic gradient sparsity (DGS)). (a) The Lena image (128×128). (b) A toy Lena image under a severe intensity distortion. Blue curves: registration between (a) and (a); red curves: registration between (b) and (a). The red curves of CC and MI are amplified by eight times for better visualization.

eliminated and the image can be fused more precisely. We iteratively run these two processes until convergence.

In the literature, existing image registration methods can be classified into feature-based registration [35]–[37] and pixel-based (or intensity-based) registration [29], [38], based on the features used in registration. Feature-based methods rely on the landmarks extracted from the images. However, locating reliable features is still an open problem and an active topic of research [39]. Here, we are more interested in intensity-based registration, which can be combined with fusion in a unified optimization scheme.

One of the most essential components of image registration is the energy function to measure (dis)similarity. The optimized similarity should correspond to the correct spatial alignment. A few similarity measures have been used for registration such as sum-of-squared-difference (SSD) [28], residual complexity (RC) [29], sum-of-absolute value (SAD), correlation coefficient (CC), CD2 [30], MS [31] and mutual information (MI) [32], [40]. When the input images have similar intensity fields, all existing similarity measures are able to find the true transformation based on the pixel values. However, due to the physics of remote sensing [1], the intensity fields of remotely sensed images may vary significantly (*e.g.*, aquired under different illumination conditions or by different types of sensors). Lots of existing intensity-based similarity measures are not robust to such intensity distortions, *e.g.*, the widely used SSD. Although some methods are proposed for simultaneous registration and intensity correction [41], [42], they often involve much higher computation complexity and multiple local minima. Considering the large image size in remote sensing, we need a stable similarity measure without introducing high computational complexity.

Fortunately, we already have one. We use the dynamic gradient sparsity to preserve spatial information (Fig. 2). Any misalignment will increase the sparsity of the gradients. Thus, the dynamic gradient sparsity can be naturally used as

a similarity measure. We can revise the energy function for simultaneous image registration and fusion:

$$E(\mathbf{X}, \mathcal{T}) = \frac{1}{2}||\psi\mathbf{X} - \mathbf{M}||_F^2 + \lambda||\nabla\mathbf{X} - \nabla\bar{\mathbf{P}}(\mathcal{T})||_{2,1}, \quad (5)$$

where $\mathcal{T}$ is the transformation to be estimated. For example, $\mathcal{T}$ is a 2-by-1 vector for translations. As the above function value is the summation of the cost of each pixel, a trivia solution would yield a large enough transformation that results in no overlapping of two images. To avoid such a trivial solution, we use the gradient descent method to decrease the normalized energy function value instead. We find this can effectively avoid iteratively transforming the image to the wrong direction. More details can be found in Algorithm 3.

We compare the proposed similarity measure with existing approaches with respect to 1D translations in Fig. 3, with SSD [28], RC [29], SAD, CC, CD2 [30], MS [31] and MI [32]. The function value of each measure is computed only from the overlapping image regions. The blue curves in Fig. 3 show the responses of different measures, all of which can find the optimal alignment at the zero translation. After adding intensity distortions and rescaling, the appearance of source image shown in Fig. 3(b) is not consistent with that of the original Lena image. The results denoted by the red curves show that only the global minimum of the proposed measure corresponds to the correct translation. If a local optimization algorithm is used to search the transformation parameters, our method is more robust for large transformations. In RC, the discrete cosine transform (DCT) is used to sparisify the residual image, which has $\mathcal{O}(N \log N)$ complexity. For this registration problem, our method only has linear complexity in each iteration. More importantly, it is unknown how to combine RC in the image fusion process and simultaneously achieve image registration and fusion in the unified model (5).

III. ALGORITHM

Now our goal is to minimize the energy function (5). We first solve the problem with respect to $\mathbf{X}$ by fixing $\mathcal{T}$,

---

Algorithm 1 SIRF

Input: L, λ, $t^1 = 1$, $\mathbf{Y}^0$
for $k = 1$ to $Maxiteration$ **do**
    $\mathbf{Y} = \mathbf{Y}^k - \psi^{-1}(\psi\mathbf{X} - \mathbf{M})/L$
    $\mathbf{X}^k = \arg\min_{\mathbf{X}}\{\frac{L}{2}\|\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda\|\nabla\mathbf{X} - \nabla\bar{\mathbf{P}}(\mathcal{T})\|_{2,1}\}$
                                       (Algorithm 2)
    $\mathcal{T} = \arg\min_{\mathcal{T}}\{E(\mathcal{T}) = \|\nabla\mathbf{X}^k - \nabla\bar{\mathbf{P}}(\mathcal{T})\|_{2,1}\}$
                                         (Algorithm 3)
    $t^{k+1} = [1 + \sqrt{1 + 4(t^k)^2}]/2$
    $\mathbf{Y}^{k+1} = \mathbf{X}^k + \frac{t^k - 1}{t^{k+1}}(\mathbf{X}^k - \mathbf{X}^{k-1})$
end for

Algorithm 2 VTV-Denoising

Input: λ, $\mathbf{Y}$, $\mathbf{P}$, $\mathcal{T}$, $(\mathbf{U}, \mathbf{V}) = (\mathbf{R}, \mathbf{S}) = (\mathbf{0}, \mathbf{0})$, $t^1 = 1$
$\mathbf{B} = \mathbf{Y} - \bar{\mathbf{P}}(\mathcal{T})$
for $k = 1$ to $Maxiteration$ **do**
    $(\mathbf{R}^k, \mathbf{S}^k) = \mathbb{P}[(\mathbf{U}^k, \mathbf{V}^k) + \frac{1}{8\lambda}\mathcal{L}^T(\mathbf{B} - \lambda\mathcal{L}(\mathbf{U}^k, \mathbf{V}^k))]$
    $t^{k+1} = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}$
    $(\mathbf{U}^{k+1}, \mathbf{V}^{k+1}) = (\mathbf{R}^k, \mathbf{S}^k) + \frac{t^k - 1}{t^{k+1}}(\mathbf{R}^k - \mathbf{R}^{k-1}, \mathbf{S}^k - \mathbf{S}^{k-1})$
end for
$\mathbf{Z} = \mathbf{B} - \lambda\mathcal{L}(\mathbf{R}^k, \mathbf{S}^k)$
$\mathbf{X} = \mathbf{Z} + \bar{\mathbf{P}}(\mathcal{T})$

Algorithm 3 Gradient Descent With Backtracking

Input: $\mathbf{X}, D(\mathbf{P}), t^0, \eta < 1, \mathcal{T}^0, k = 0$.
repeat
    1)   compute $\mathcal{T}^{k+1} = \mathcal{T}^k + t^k\Delta\mathcal{T}^k$
    2)   if $E(\mathcal{T}^{k+1})/M_{k+1} > E(\mathcal{T}^k)/M_k$, set $t^k = \eta t^k$
    and go back to (1)
    3) $t^{k+1} = t^k$
    4) $k = k + 1$
until Stop criterion is met

---

and then solve the problem with respect to $\mathcal{T}$ by fixing $\mathbf{X}$. For the $\mathbf{X}$ subproblem:

$$E(\mathbf{X}) = \frac{1}{2}\|\psi\mathbf{X} - \mathbf{M}\|_F^2 + \lambda\|\nabla\mathbf{X} - \nabla\bar{\mathbf{P}}(\mathcal{T})\|_{2,1}, \quad (6)$$

it is an obvious convex function. The first term is smooth while the second term is non-smooth. This motivates us to solve this subproblem in the FISTA framework [22], [43]. It has been proven that FISTA can achieve the optimal convergence rate for first order methods. That is, $E(\mathbf{X}^k) - E(\mathbf{X}^*) \sim \mathcal{O}(1/k^2)$, where $\mathbf{X}^*$ is the optimal solution and k is the iteration counter.

The second subproblem with respect to $\mathcal{T}$ can be written as:

$$\min E(\mathcal{T}) = \|\nabla\mathbf{X} - \nabla\bar{\mathbf{P}}(\mathcal{T})\|_{2,1}. \quad (7)$$

We solve this subproblem during each iteration of FISTA, and the steps will be discussed soon. We summarize the proposed simultaneous image registration and fusion (SIRF) for pan-sharpening in Algorithm 1.

Here ψ^{-1} denotes the inverse operator of ψ. L is the Lipschitz constant for $\psi^{-1}(\psi\mathbf{X} - \mathbf{M})$. We could observe that the solution is updated based on both $\mathbf{X}^k$ and $\mathbf{X}^{k-1}$, while the Bregman method used in previous methods [8], [10] updates $\mathbf{X}$ only based on $\mathbf{X}^k$. This is a reason why our method converges faster.

There are two subproblems in SIRF. For the first subproblem, $L = 1$ and

$$\mathbf{X}^k = \arg\min_{\mathbf{X}}\{\frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda\|\nabla\mathbf{X} - \nabla\bar{\mathbf{P}}(\mathcal{T})\|_{2,1}\}. \quad (8)$$

Let $\mathbf{Z} = \mathbf{X} - \bar{\mathbf{P}}(\mathcal{T})$ and we can rewrite the problem:

$$\mathbf{Z}^k = \arg\min_{\mathbf{Z}}\{\frac{1}{2}\|\mathbf{Z} - (\mathbf{Y} - \bar{\mathbf{P}}(\mathcal{T}))\|_F^2 + \lambda\|\nabla\mathbf{Z}\|_{2,1}\}. \quad (9)$$

This alternative problem is now a VTV denoising problem [25]–[27] and $\mathbf{X}^k$ can be updated by $\mathbf{Z}^k + \bar{\mathbf{P}}(\mathcal{T})$. The slow version of the VTV denoising algorithm [25] is accelerated based on the FISTA framework to solve (8), which is summarized in Algorithm 2.

Algorithm 2 solves the dual problem of (9). To make the algorithm easier to understand, we use the same notations as those in [26] and [27]. The linear operator is defined as: $\mathcal{L}(\mathbf{R}, \mathbf{S})_{i,j,d} = \mathbf{R}_{i,j,d} - \mathbf{R}_{i-1,j,d} + \mathbf{S}_{i,j,d} - \mathbf{S}_{i,j-1,d}$. The corresponding inverse operator is defined as $\mathcal{L}^T(\mathbf{X}) = (\mathbf{R}, \mathbf{S})$ with $\mathbf{R}_{i,j,d} = \mathbf{X}_{i,j,d} - \mathbf{X}_{i+1,j,d}$ and $\mathbf{S}_{i,j,d} = \mathbf{X}_{i,j,d} - \mathbf{X}_{i,j+1,d}$.

$\mathbb{P}$ is a projection operator used to ensure that $\sum_{d=1}^s(\mathbf{R}_{i,j,d}^2 + \mathbf{S}_{i,j,d}^2) \leq 1$, $|\mathbf{R}_{i,n,d}| \leq 1$, and $|\mathbf{S}_{m,j,d}| \leq 1$.

We solve the second subproblem (7) with the gradient descent method by backtracking. One of the advantages is that no extra parameters are introduced during the minimization.

The $\ell_{2,1}$ norm is not smooth. Let $\mathbf{r} = \nabla\mathbf{X} - \nabla\bar{\mathbf{P}}(\mathcal{T})$ be the residual and we can have a tight approximation for the function:

$$E(\mathcal{T}) \approx \sum_i \sum_j \sqrt{\sum_d (\nabla_1\mathbf{r}_{i,j,d})^2 + (\nabla_2\mathbf{r}_{i,j,d})^2 + \epsilon}, \quad (10)$$

where ϵ is a small constant (*e.g.*, 10^{-10}).

The image transformtion is nonlinear. Here we use the first order Taylor approximation that $\nabla\bar{\mathbf{P}}(\mathcal{T} + \Delta\mathcal{T}) \approx \nabla\bar{\mathbf{P}}(\mathcal{T}) + \mathbf{J}\Delta\mathcal{T}$, where $\mathbf{J} \in \mathbb{R}^{mns \times p}$ is the image Jacob of $\nabla\bar{\mathbf{P}}(\mathcal{T})$, and p is the number of parameters for the transformation. Now, it is not hard to obtain the gradient of the energy function with respect to $\Delta\mathcal{T}$ by the chain rule. We can iteratively update $\mathcal{T}$ with $\Delta\mathcal{T}$ and warp the image with $\mathcal{T}$ to obtain a new Jacob matrix until convergence.

$$\Delta\mathcal{T} = \mathbf{J}^T vec(\frac{\partial E(\mathcal{T})}{\partial\mathbf{r}}) \quad (11)$$

where $vec()$ denotes vectorization. Although our method can be applied to more complex transformations, the transformation is assumed to be affine or translational in the image fusion application, with 6 and 2 motion parameters, respectively.

Gradient descent with backtracking is used to minimize the energy function (10), which is summarized in Algorithm 3. We set the initial step size $t^0 = 1$ and $\eta = 0.8$. It can be easily observed that all operations are linear. Hierarchical estimation is applied for the registration [44]. The function value of (10)
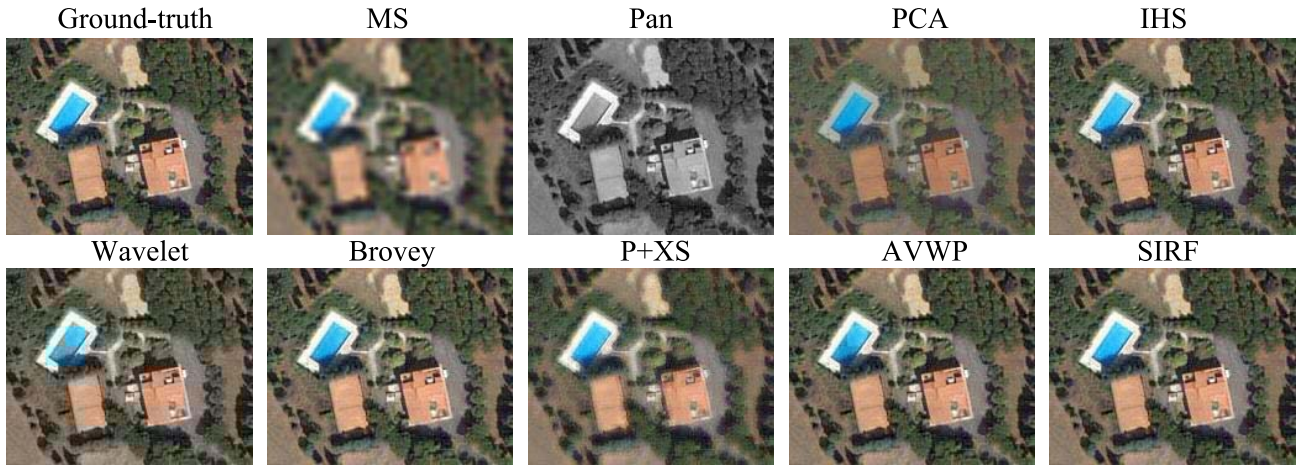
Fig. 4.   Fusion Results comparison (source: Quickbird). The Pan image has 200×160 pixels.

is calculated on the overlapping area of two images. To avoid trivial solutions such as zooming in on a dark region, we use the normalized function value here (divided by the overlapping pixels in the k-th iteration M_k). When there is no overlapping, the function value will be infinity. We found this approach could effectively rule out the trivial solutions.

Due to the tradeoff between accuracy and computational cost for solving the subproblem, the inner loops of Algorithms 2 and 3 only run three iterations in all experiments. It has been shown that such inexact solution in FISTA does not change the overall convergence rate [45].

IV. EXPERIMENTS

A. Simulation

The proposed method is validated on multispectral datasets from Quickbird, Geoeye, SPOT and IKONOS satellites. The resolution of Pan images ranges from 0.41 m to 1.5 m. All the corresponding Ms images have lower resolutions with $c = 4$ and contain blue, green, red and near-infrared bands. For convenience, only the RGB bands are presented. Due to the lack of multi-resolution images of the same scene, the original images are viewed as ground-truth and low-resolution images are downsampled from the ground-truth images. This strategy is common for comparing fusion algorithms (*e.g.*, [8]–[10], [46], [47]). In these simulations, there is no misalignment for the input images. We do not run the registration steps in SIRF except in the case with artificial translations.

We compare our method with classical methods PCA [2], IHS [3], wavelet [4], Brovey [6] and variation methods P+XS [8], and AVWP [9]. The recent model-based fusion method (MBF) [17] requires multiple parameters to be tuned based on the individual satellite. The current version can only support the datasets from the IKONOS satellite. Therefore, it is not compared in these simulations. The effectiveness of our method was validated via extensive experiments with visual and quantitative analysis. Comparisons with P+XS [8] and AVWP [9] demonstrated its efficiency. The parameters for each method were tuned individually according to the authors'

suggestions, and the best set was selected for each method, respectively. All experiments were conducted using MATLAB on a desktop computer with 3.4GHz Intel core i7 3770 CPU.[2] The simulation results are shown in Subsections IV-B to IV-E.

B. Visual Comparison

First, we compare the fusion results by our method with those of previous works [2]–[4], [6], [8], [9]. Figure 4 shows the fusion results as well as the ground-truth Quickbird images. All the methods can produce images with higher resolutions than the original Ms image. Obviously, PCA [2] performed the worst as the overall intensities of the image changed. No obvious artifacts can be found on the images produced by IHS [3] and Brovey [6]. However, a closer look shows that the color on these images tends to change, especially on the trees and grass. This is a sign of spectral distortion [1]. Wavelet fusion [4] suffers from both spectral distortion and blocky artifacts (*e.g.*, on the swimming pool). Blurred edges is a general issue in the image fused by P+XS [8]. AVWP [9] performs much better than all of them but it inherits the blocky artifacts of the wavelet fusion. The results of another experiment on an IKONOS image are shown in Figure 5, with similar performance by each algorithm. Some visible bright pixels can be found at the top-left corner of Brovey.

For better visualization, the error images compared with the ground-truth are presented in Figure 6 and Figure 7 at the same scale. From these error images, the spectral distortion, blocky artifacts, and blurriness can be clearly observed. These results are consistent with those presented in previous work [9]. Due to the spectral distortion, the conventional methods are not adapted to vegetation study [1]. Previous variational methods [8], [9] try to break such hard assumptions by combining a few weak assumptions. However, such combinations involve more parameters that must be tuned. Moreover, the fusion from the upsampled Ms image often results in inaccuracy. In contrast, we only constrain the

[2]The demo code of SIRF can be downloaded from: https://dl.dropboxusercontent.com/u/58080329/codeSIRF.zip
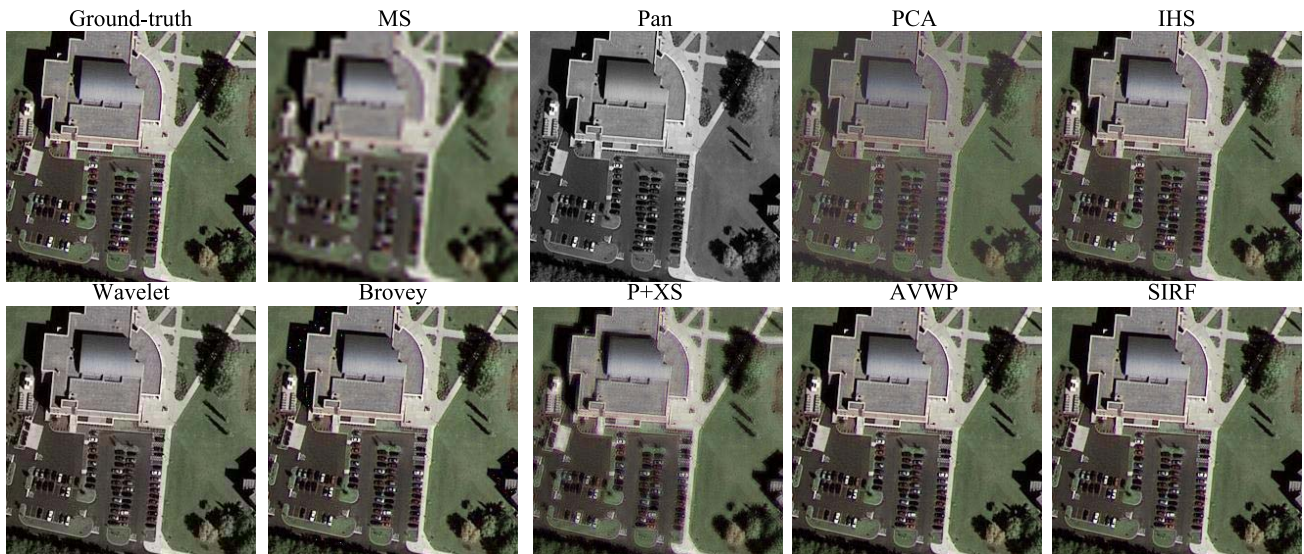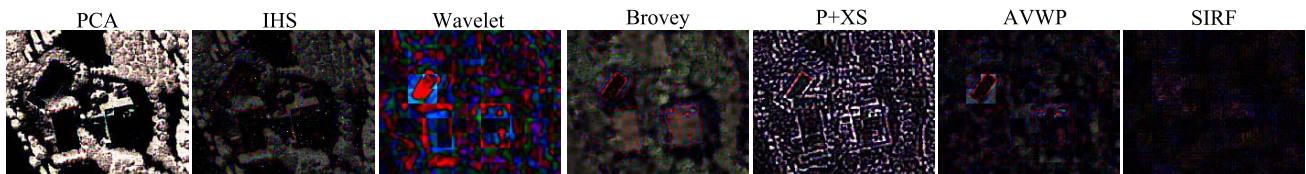
Fig. 5. Fusion Results comparison (source: IKONOS). The Pan image has 256×256 pixels.

Fig. 6. The corresponding error images to those in Figure 4. Brighter pixels represent larger errors.
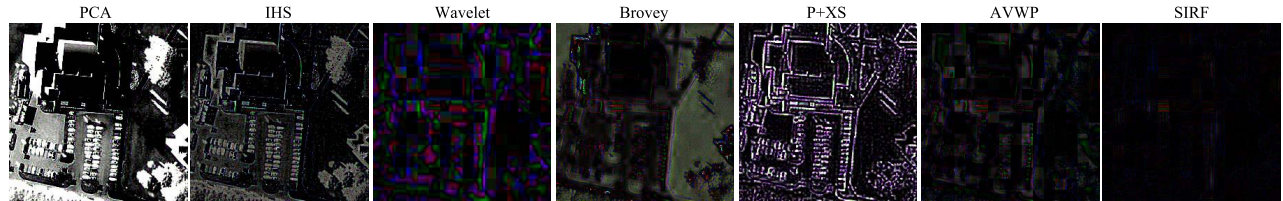

Fig. 7. The corresponding error images to those in Figure 5. Brighter pixels represent larger errors.

spectral information of the fused image to be locally consistent with the original Ms image. The fusion results are impressively good on these two images.

C. Quantitative Analysis

In addition to the two images used previously, 156 test images of different sizes (from 128×128 to 512×512) are cropped from Quickbird, Geoeye, IKONOS and SPOT datasets, which contain vegetation (*e.g.*, forest, farmland), bodies of water (*e.g.*, river, lake) and urban scenes (*e.g.*, building, road). This test set is much larger than the size of all datasets considered in previous variational methods (31 images in [8], 7 images in [9] and 4 images in [10]). Example images are shown in Figure 8.

To evaluate the fusion quality of different methods, we used four metrics that measure spectral quality and one metric that measures spatial quality. The spectral metrics include the relative dimensionless global error in synthesis (ERGAS) [48], spectral angle mapper (SAM) [48],

Fig. 8. Example images used in our experiments.

universal image quality index (Q-average) [49] and relative average spectral error (RASE) [50]. The filtered correlation coefficients (FCC) [4] were used as a spatial quality metric. In addition, peak signal-to-noise ratio (PSNR), and root mean squared error (RMSE) and mean structural

TABLE I

PERFORMANCE COMPARISON ON THE 158 REMOTELY SENSED IMAGES

| Method	ERGAS	QAVE	RASE	SAM	FCC	PSNR	MSSIM	RMSE
PCA [2]	5.67±1.77	0.664±0.055	22.3±6.8	2.11±1.35	0.972±0.014	20.7±2.7	0.799±0.067	24.1±6.7
IHS [3]	1.68±0.86	0.734±0.011	6.63±3.4	0.79±0.54	0.989±0.006	31.2±4.6	0.960±0.035	8.1±4.2
Wavelet [4]	1.18±0.45	0.598±0.113	4.50±1.6	2.45±1.18	**0.997±0.002**	36.1±3.6	0.983±0.009	4.5±1.9
Brovey [6]	1.22±1.08	0.733±0.011	5.18±4.6	0.61±0.58	0.940±0.170	38.2±5.6	0.989±0.008	9.1±19.7
P+XS [8]	0.89±0.33	0.720±0.036	3.47±1.3	0.66±0.36	0.898±0.024	25.9±3.5	0.854±0.051	14.7±5.4
AVWP [9]	0.46±0.17	0.733±0.013	1.81±0.6	0.69±0.70	0.996±0.002	40.0±3.5	0.991±0.006	2.9±1.0
SIRF	**0.07±0.03**	**0.746±0.004**	**0.3±0.1**	**0.18± 0.11**	**0.997±0.002**	**47.5±3.6**	**0.998±0.001**	**1.1±0.5**
Desired Value	0	1	0	0	1	$+\infty$	1	0
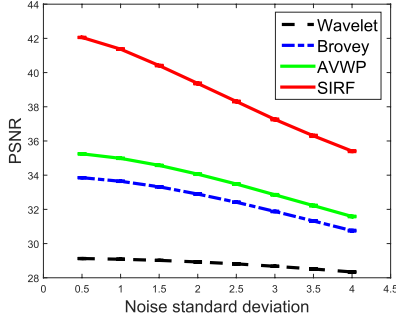
Fig. 9. The PSNR values of Wavelet, Brovey, AVWP and SIRF with different levels of noise. The test image is the one shown in Fig. 4.
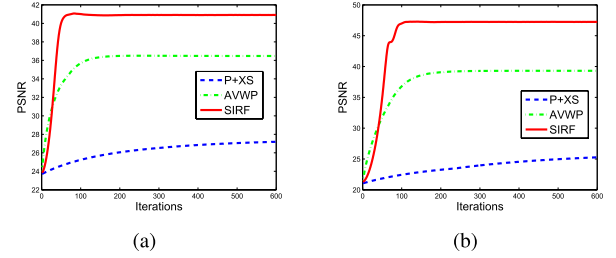

Fig. 10. Convergence rate comparison among P+XS, AVWP and the proposed method. (a) Result corresponds to Figure 4. (b) Result corresponds to Figure 5.

similarity (MSSIM) [49] were used to evaluate the fusion accuracy when compared with the ground-truth.

The average results and the variance on this test set are listed in Table I. The ideal value for each metric is shown in the last row. The results of variational methods [8], [9] have much lower values in ERGAS and RASE than those of conventional methods [2]–[4], [6]. From QAVE and SAM, the results are comparable to conventional methods. We can conclude that these variational methods can preserve more spectral information. Due to the blurriness, P+XS has the worse spatial resolution in terms of FCC. In terms of error and similarity metrics (PSNR, MSSIM, RMSE), AVWP and P+XS are always the second best and second worst, respectively. Except for the same FCC as the wavelet fusion, our method is consistently better than all previous methods in terms of all metrics. These results are enough to demonstrate the success of our method, where the dynamic gradient sparsity can preserve sharp edges and the spectral constraint keeps accurate spectral information. In terms of PSNR, it can outperform the second best AVWP by more than 7 dB. We added zero-mean Gaussian noise with different variances on the input images. The results in Fig. 9 demonstrate that the robustness of our method with respect to the noise and always achieve the highest PSNR value.

If we consider the prior information that has been used, the performance of each algorithm is easy to explain. Conventional projection-substitution methods only treat the input images as vectorial information (*i.e.*, 1D). The difference is the substitution performed on various projection spaces. However, 2D information such as edges is not utilized. The edge information has been considered in both variational methods P+XS [8] and AVWP [9], although their models can not

effectively exploit this prior information. Promising results, especially by AVWP, have already outperformed conventional methods. By using the proposed dynamic gradient sparsity, our method has successfully learned more prior knowledge provided by the Pan image. Due to the group sparsity across different bands, our method is less sensitive to noise. This is why our method consistently outperforms the others.

D. Efficiency Comparison

To evaluate the efficiency of the proposed method, we compared the proposed method with previous variational methods, P+XS [8] and AVWP [9], in terms of both accuracy and computational cost. PSNR was used to measure fusion accuracy. Figure 10 demonstrates the convergence rate comparison of these algorithms corresponding to the images in Figures 4 and 5. Inheriting the benefit of the FISTA [22] framework, our method often converged in 100 to 150 outer iterations. AVWP often converges in 200 to 400 iterations. P+XS, which uses the classic gradient decent method, has not converged even with 600 iterations. Due to the very slow convergence speed of P+XS, we terminated it after 600 iterations in all experiments. After each algorithm converged, our method outperformed AVWP by more than 5 dB and 8 dB on these two images in terms of PSNR. Note that AVWP is the second best method from previous analysis.

The average computational costs of these three methods are listed in Table II for different sizes of test images. Both the proposed method and AVWP terminate when a fixed tolerance is reached (*e.g.*, 10^{-3} of the relative change on **X**). The computational cost of our method tends to be linear from these results. Even the second fastest method AVWP takes about 50% more time than ours on an image of 512 by 512 pixels.

TABLE II
COMPUTATIONAL TIME (SECOND) COMPARISON

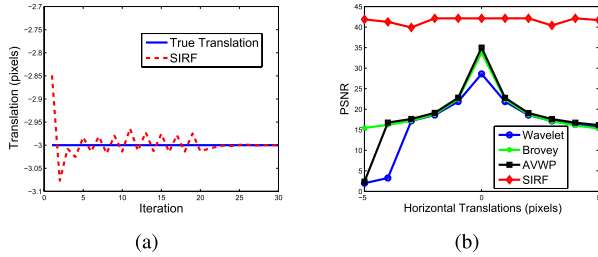|	128×128	256×256	384×384	512×512
P+XS	6.7	16.0	48.3	87.4
AVWP	1.7	8.3	28.2	54.7
SIRF	**1.4**	**5.0**	**19.3**	**36.8**

(a)   (b)

Fig. 11. (a) The convergence property of the registration in SIRF. (b) The fusion performance of Wavelet [4], Brovey [6], AVWP [9] and SIRF with respect to horizontal translations on the Quickbird image (Fig. 1). The estimated horizontal translation by our method is -5.02, -3.99, -2.99, -1.98, -0.97, 0.02, 1.03, 2.03, 3.01, 4.02, 4.97 pixels, respectively.
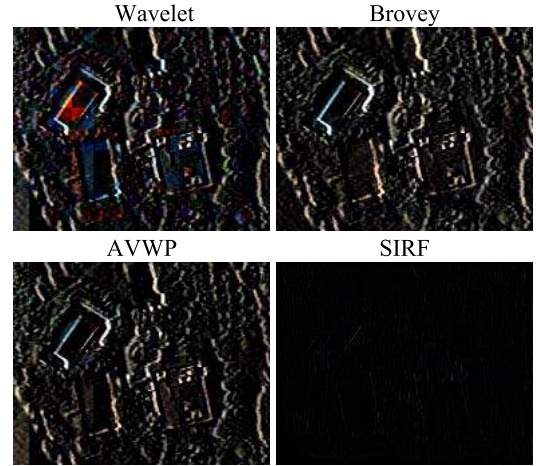

Fig. 12. The fusion errors of Wavelet [4], Brovey [6], AVWP [9] and SIRF when there is a horizontal misalignment of 3 pixels on the Pan image. The errors are shown at the same scale.

These comparisons are sufficient to demonstrate the efficiency and effectiveness of our method.

E. Translation

Since the registration error is almost unavoidable by existing pre-registration approaches, we validated the performance of different methods under various amount of translations. We added an artificial horizontal translation of 3 pixels in the Pan image shown in Fig. 1. The estimated translation of SIRF is shown in Fig. 11 (a). With around 30 iterations, our method was able to recover the true translation very accurately. The registration in each iteration costs about 3.7 seconds on this data, which is much slower than the fusion process. Comparing against existing intensity-based registration methods with higher complexity [29], our registration method only has linear complexity to the size of images. Therefore, such speed is quite acceptable. Considering this registration cost, we only ran the registration in the first 3 iterations in all the later experiments. With the inexact registration, we find that the error is often below or around 0.03 pixels (*i.e.*, 1% error), and this precision is sufficient for accurate fusion.

The best four algorithms in Table I (based on PSNR), Wavelet [4], Brovey [6], AVWP [9] and SIRF were selected for comparisons under horizontal translations. The results in Fig. 11 (b) shows that Wavelet [4], Brovey [6] and AVWP [9] cannot "tolerate" large misalignments, while the performance of SIRF is quite stable. In this experiment, the largest registration error is always within 0.03 pixel by our method. The fusion errors are presented in Fig. 12 when there is a horizontal misalignment of 3 pixels on the Pan image. Due to the misalignment, visible errors can be observed on the boundaries of land objects by Wavelet, Brovey and AVWP. Our method can simultaneously estimate the translation and achieve substantial improvement over existing methods.

F. Real-World Datasets

Finally, we evaluated the fusion results of different methods on real-world datasets (non-simulated). The imagery was acquired by IKONOS multispectral imaging satellite [51], which contains preregistered Pan and Ms images at their capture resolutions. The registration error was within a subpixel on the Ms images (up to 4 pixels on the Pan image), and we did not add artificial transformation on the datasets. The multispectral image has four bands: blue, green, red and near-infrared, with 4 meter resolution. The Pan image has 1 meter resolution. We rescaled these images to 0-255 before processing. The bicubic interpolation was used for both upsampling and downsampling.

In addition to the methods we have compared, the latest method MBF [17] was also compared for these datasets. The parameters of MBF have already been optimized on these datasets and we use their default setting for experiments. Fig. 13 shows the fusion results on a portion of the datasets.[3] The images obtained by PCA, IHS, and P+XS have blurry edges due to misalignment. Blocky artifacts can be found in the results by Wavelet and AVWP, which is consistent with the observation in [17]. Brovey, MBF and the proposed SIRF provide high quality images with sharp object boundaries. Such results are reasonable because the pixel-to-pixel based methods are often sensitive to the misalignment. In Brovey, the spectral information is mapped to the Pan image, and the high spatial resolution can always be preserved. Based on the designed high-pass and low-pass filters in MBF, it can somewhat tolerate misalignment. In SIRF, the transformation is estimated simultaneously during the fusion process. Therefore, accurate fusion can be achieved from the well-aligned images.

We then took a close look at the fused images by Brovey, MBF and SIRF, which are shown in Fig. 14. From the upsampled Ms image, it shows the roofs of the buildings should be in white and dark orange colors. The spectral distortion can be clearly observed in the fused image by Brovey, where most of the roof is covered by dark orange. Both MBF and SIRF provided results with high quality. The zoomed-in areas show that the result by our method has sharper edges than

---

[3]The results are shown on the 0-255 scale without histogram stretch.

Pan   MS

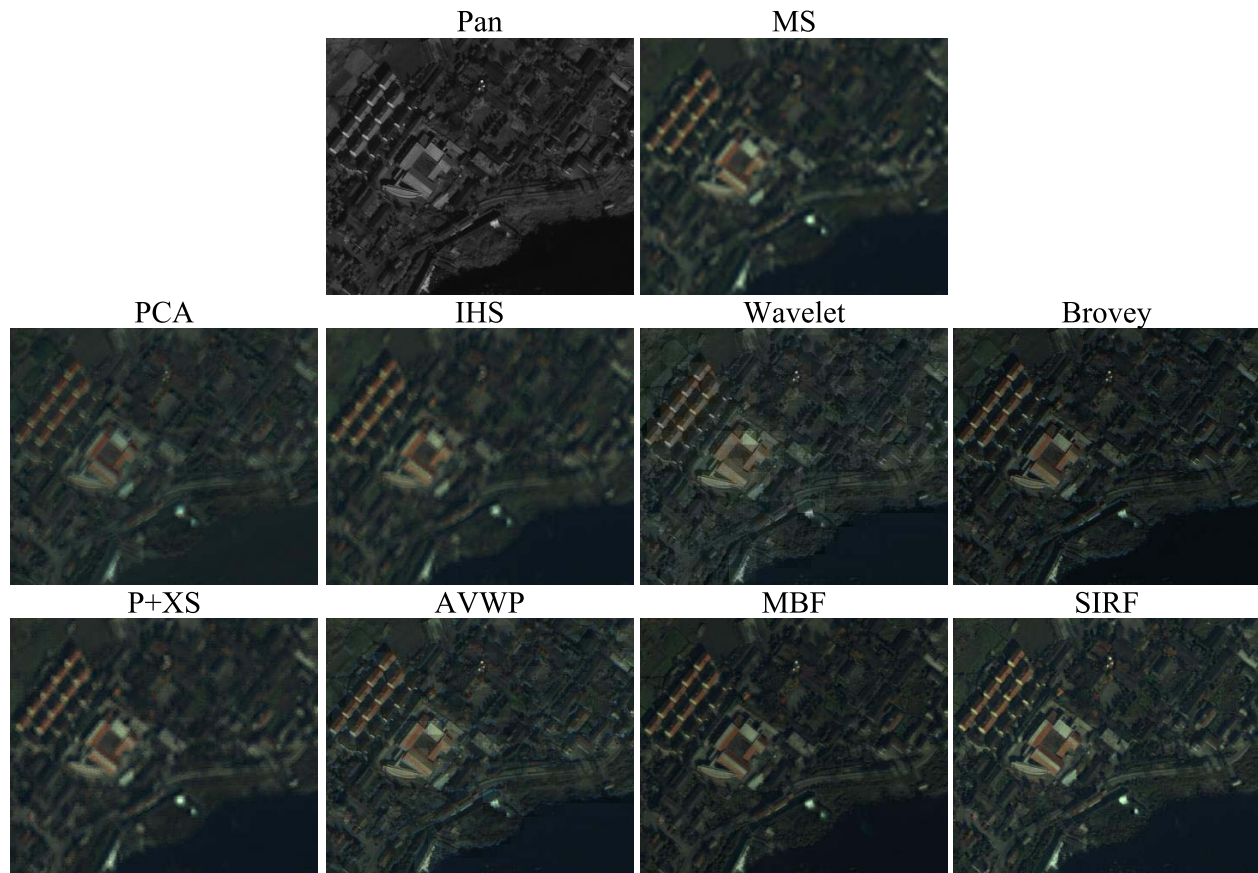PCA   IHS   Wavelet   Brovey

P+XS   AVWP   MBF   SIRF

Fig. 13.   The fusion result for a portion of the IKONOS China-Sichuan 58208_0000000.20001108 dataset. The RGB channels are presented. This figure is better viewed on screen with 200% size.
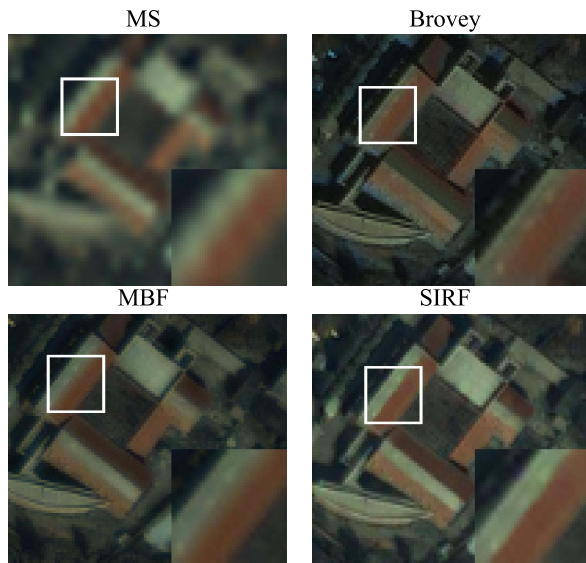
MS   Brovey

MBF   SIRF

Fig. 14.   The zoomed-in areas fused by Brovey, MBF and SIRF in Fig. 13.

TABLE III
QUANTITATIVE PERFORMANCE COMPARISON BETWEEN MBF AND
SIRF ON THE IKONOS CHINA-SICHUAN DATASETS

Method	ERGAS	QAVE	RASE	SAM	FCC
Desired Value	0	1	0	0	1
China-Sichuan 58204_0000000.20001116					
MBF	1.80	0.973	7.46	2.22	0.887
SIRF	**0.06**	**1.000**	**0.25**	**0.07**	**0.919**
China-Sichuan 58205_0000000.20001003					
MBF	1.73	0.980	7.49	1.68	0.861
SIRF	**0.06**	**1.000**	**0.23**	**0.06**	**0.920**
China-Sichuan 58207_0000000.20000831					
MBF	2.78	0.947	11.49	2.39	0.793
SIRF	**0.04**	**1.000**	**0.17**	**0.04**	**0.869**
China-Sichuan 58208_0000000.20001108					
MBF	2.38	0.954	9.89	2.34	0.915
SIRF	**0.08**	**0.999**	**0.33**	**0.11**	**0.927**

that by MBF. The two colors are mixed together in the result by MBF, while this mixed color does not actually exist in the original Ms image. These results are sufficient to demonstrate that the our method outperformed existing methods on this dataset due to the inherent registration scheme.

From the previous visual results, MFB and SIRF provided the most accurate results. We quantitatively compared the results by MFB and SIRF on four portions of the China-Sichuan IKONOS datasets,[4] such as that shown in Fig. 13. As there is no ground-truth for these real datasets, we only used the ERGAS, QAVE, RASE, SAM and FCC for comparisons. The results are shown in Table III. For both spectral and spatial information, our method is consistently better than MBF. Due to the misalignment of preregistration, the fusion error in MBF was expected. We considered the correlations across different bands and performed the fusion jointly, while MBF could only fuse the different

[4]We use the same portions as in [17].

bands individually. Both the fusion scheme and simultaneous registration contributed to the promising results of SIRF.

V. CONCLUSION AND DISCUSSION

We have proposed a novel and powerful variational model for simultaneous image registration and fusion in a unified framework, based on the property of dynamic gradient sparsity. The model naturally incorporates the gradient prior information from a high-resolution Pan image and the spectral information from a low-resolution Ms image. Our method leads to several advantages over existing methods. First, the proposed dynamic gradient sparsity can directly exploit the sharp edges from the Pan image, which has been shown very effective. Second, we jointly sharpen the images by incorporating the intra-correlations across different bands, while most existing methods are based on band-by-band fusion. The last and most importantly, although the registration is quite challenging between the Pan and Ms images due to their different spatial resolutions, our method can simultaneously register the two input images during the fusing process, acting directly on the input images without any pre-filtering and feature extraction.

An efficient optimization algorithm has been devised to solve the problem, with sufficient implementation details. Extensive experiments were conducted on 158 simulated images stemming from a variety of sources. Due to the proposed unique techniques, our method is corroborated to consistently outperform the state-of-the-arts in terms of both spatial and spectral qualities. We further evaluated our method on real IKONOS datasets with comparison to the latest methods. The results show that our method can effectively eliminate the misalignment during preregistration, and provide higher quality products than the existing methods. Due to the high accuracy and simultaneous registration property, our method may benefit more applications in remote sensing, *e.g.*, classification, change detection, *etc*. The gradient prior information and joint structure were separately utilized in other image enhancement tasks [52], [53]. Such a success further demonstrates the effectiveness of our modeling.

Parallel programming can further accelerate the running speed of our method. No information from another patch is required for fusion. Since our method does not require a strong correlation between two input images, it may be used in fusing images from different capture times (shadows should be taken into account) or different sources (*e.g.*, images from different satellites). Currently, the land objects in the input images are assumed to be consistent. If there exist moving objects (*e.g.*, vehicles), blurriness may occur at the corresponding locations. Our method may fail in some extreme cases. For example, the input images were acquired by different platforms, particularly, one image was occluded by clouds while the other one was not. In the future, we will work on such cases.

REFERENCES

[1] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.

[2] P. S. Chavez, S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogram. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, 1991.

[3] R. Haydn, G. W. Dalke, J. Henkel, and J. E. Bare, "Application of the IHS color transform to the processing of multisensor data and image enhancement," in *Proc. Int. Symp. Remote Sens. Environ.*, 1982, pp. 599–616.

[4] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge Landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, 1998.

[5] I. Amro, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "A survey of classical methods and new trends in pansharpening of multispectral images," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 79, pp. 1–22, 2011.

[6] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color enhancement of highly correlated images. I. Decorrelation and HSI contrast stretches," *Remote Sens. Environ.*, vol. 20, no. 3, pp. 209–235, 1986.

[7] H. Aanæs, J. R. Sveinsson, A. A. Nielsen, T. Bovith, and J. A. Benediktsson, "Model-based satellite image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1336–1346, May 2008.

[8] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for P+XS image fusion," *Int. J. Comput. Vis.*, vol. 69, no. 1, pp. 43–58, 2006.

[9] M. Möller, T. Wittman, A. L. Bertozzi, and M. Burger, "A variational approach for sharpening high dimensional images," *SIAM J. Imag. Sci.*, vol. 5, no. 1, pp. 150–178, 2012.

[10] F. Fang, F. Li, C. Shen, and G. Zhang, "A variational approach for pan-sharpening," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2822–2834, Jul. 2013.

[11] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[12] X. Ding, Y. Jiang, Y. Huang, and J. Paisley, "Pan-sharpening with a Bayesian nonparametric dictionary learning model," in *Proc. 17th Int. Conf. Artif. Intell. Statist.*, 2014, pp. 176–184.

[13] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.

[14] C. Jiang, H. Zhang, H. Shen, and L. Zhang, "A practical compressed sensing-based pan-sharpening method," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 629–633, Jul. 2012.

[15] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "A new pansharpening algorithm based on total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 318–322, Jan. 2014.

[16] Z. Li and H. Leung, "Fusion of multispectral and panchromatic images using a restoration-based method," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1482–1491, May 2009.

[17] H. A. Aly and G. Sharma, "A regularized model-based optimization framework for pan-sharpening," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2596–2608, Jun. 2014.

[18] J. Inglada, V. Muron, D. Pichard, and T. Feuvrier, "Analysis of artifacts in subpixel remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 1, pp. 254–264, Jan. 2007.

[19] S. Leprince, S. Barbot, F. Ayoub, and J.-P. Avouac, "Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1529–1558, Jun. 2007.

[20] M. L. Uss, B. Vozel, V. A. Dushepa, V. A. Komjak, and K. Chehdi, "A precise lower bound on image subpixel registration accuracy," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3333–3345, Jun. 2014.

[21] P. Blanc, L. Wald, and T. Ranchin, "Importance and effect of co-registration quality in an example of 'pixel to pixel' fusion process," in *Proc. 2nd Conf. Fusion Earth Data, Merging Point Meas., Raster Maps Remotely Sensed Images*, 1998, pp. 67–74.

[22] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[23] C. Chen, Y. Li, W. Liu, and J. Huang, "Image fusion with local spectral consistency and dynamic gradient sparsity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2760–2765.

[24] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc., B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[25] X. Bresson and T. F. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Problems Imag.*, vol. 2, no. 4, pp. 455–484, 2008.

[26] J. Huang, C. Chen, and L. Axel, "Fast multi-contrast MRI reconstruction," in *Proc. Med. Image Comput. Comput.-Assist. Intervent*, 2012, pp. 281–288.

[27] C. Chen, Y. Li, and J. Huang, "Calibrationless parallel MRI with joint total variation regularization," in *Proc. Med. Image Comput. Comput.-Assist. Intervent.*, 2013, pp. 106–114.

[28] P. Thevenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Trans. Image Process.*, vol. 7, no. 1, pp. 27–41, Jan. 1998.

[29] A. Myronenko and X. Song, "Intensity-based image registration by minimizing residual complexity," *IEEE Trans. Med. Imag.*, vol. 29, no. 11, pp. 1882–1891, Nov. 2010.

[30] B. Cohen and I. Dinstein, "New maximum likelihood motion estimation schemes for noisy ultrasound images," *Pattern Recognit.*, vol. 35, no. 2, pp. 455–463, 2002.

[31] A. Myronenko, X. Song, and D. J. Sahn, "Maximum likelihood motion estimation in 3D echocardiography through non-rigid registration in spherical coordinates," in *Functional Imaging and Modeling of the Heart*. Berlin, Germany: Springer-Verlag, 2009, pp. 427–436.

[32] P. Viola and W. M. Wells, III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, 1997.

[33] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *J. Mach. Learn. Res.*, vol. 12, pp. 3371–3412, Nov. 2011.

[34] J. Huang, X. Huang, and D. Metaxas, "Learning with dynamic group sparsity," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 64–71.

[35] Y. Zheng *et al.*, "Landmark matching based retinal image alignment by enforcing sparsity in correspondence matrix," *Med. Image Anal.*, vol. 18, no. 6, pp. 903–913, 2014.

[36] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.

[37] J. Ma, J. Zhao, Y. Ma, and J. Tian, "Non-rigid visible and infrared face registration via regularized Gaussian fields criterion," *Pattern Recognit.*, vol. 48, no. 3, pp. 772–784, 2015.

[38] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2233–2246, Nov. 2012.

[39] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1153–1190, Jul. 2013.

[40] S. Suri and P. Reinartz, "Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 939–949, Feb. 2010.

[41] K. M. Pohl *et al.*, "A unifying approach to registration, segmentation, and intensity correction," in *Proc. Med. Image Comput. Comput.-Assist. Intervent.*, 2005, pp. 310–318.

[42] J. Modersitzki and S. Wirtz, "Combining homogenization and registration," in *Biomedical Image Registration*. Berlin, Germany: Springer-Verlag, 2006, pp. 257–263.

[43] J. Huang, S. Zhang, H. Li, and D. Metaxas, "Composite splitting algorithms for convex optimization," *Comput. Vis. Image Understand.*, vol. 115, no. 12, pp. 1610–1622, 2011.

[44] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Proc. Eur. Conf. Comput. Vis.*, 1992, pp. 237–252.

[45] M. W. Schmidt, N. LeRoux, and F. Bach, "Convergence rates of inexact proximal-gradient methods for convex optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 1458–1466.

[46] V. Vijayaraj, C. G. O'Hara, and N. H. Younan, "Quality analysis of pansharpened images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2004, pp. 85–88.

[47] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogram. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.

[48] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.

[49] Z. Wang and A. C. Bovik, "Modern image quality assessment," *Synth. Lect. Image, Video, Multimedia Process.*, vol. 2, no. 1, pp. 1–156, 2006.

[50] M. Choi, "A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1672–1682, Jun. 2006.

[51] Space Imaging. (2000). *IKONOS Scene*, GeoEye, Dulles, VA, USA. [Online]. Available: http://glcf.umd.edu/data/ikonos/

[52] D. Krishnan and R. Fergus, "Dark flash photography," in *Proc. SIGGRAPH*, 2009, Art. ID 96.

[53] M.-Y. Liu, O. Tuzel, and Y. Taguchi, "Joint geodesic upsampling of depth images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 169–176.

Chen Chen (S'13) received the B.E. and M.S. degrees from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2011, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering with the University of Texas at Arlington. His major research interests include image processing, medical imaging, computer vision, and machine learning.

Yeqing Li received the B.E. degree in computer science and technology from Shantou University, China, in 2006, and the M.E. degree from Nanjing University, Nanjing, China, in 2009. He is currently pursuing the Ph.D. degree with the Department of Computer Science, University of Texas at Arlington. His major research interests include machine learning, pattern recognition, medical image analysis, and computer vision.

Wei Liu received the M.Phil. and Ph.D. degrees in electrical engineering from Columbia University, New York, NY, USA, in 2012. He is currently a Research Staff Member with the IBM T. J. Watson Research Center, Yorktown Heights, NY, USA, and holds an adjunct faculty position with the Rensselaer Polytechnic Institute, Troy, NY, USA. He has been the Josef Raviv Memorial Post-Doctoral Fellow with the IBM T. J. Watson Research Center for one year since 2012. His research interests include machine learning, data mining, computer vision, pattern recognition, and information retrieval. He was a recipient of the 2011–2012 Facebook Fellowship and the 2013 Jury Award for best thesis of the Department of Electrical Engineering, Columbia University.

Junzhou Huang (M'12) received the B.E. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1996, the M.S. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2003, and the Ph.D. degree from Rutgers University, New Brunswick, New Jersey, in 2011. He is currently an Assistant Professor with the Computer Science and Engineering Department, University of Texas at Arlington His research interests include biomedical imaging, machine learning, and computer vision, with a focus on the development of sparse modeling, imaging, and learning for large scale inverse problems.