MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Blind Multi-Spectral Image Pan-Sharpening

Yu, Lantao; Liu, Dehong; Mansour, Hassan; Boufounos, Petros T.; Ma, Yanting

TR2020-047 April 11, 2020

Abstract

We address the problem of sharpening low spatial-resolution multi-spectral (MS) images with their associated misaligned high spatial-resolution panchromatic (PAN) image, based on priors on the spatial blur kernel and on the cross-channel relationship. In particular, we formulate the blind pan-sharpening problem within a multi-convex optimization framework using total generalized variation for the blur kernel and local Laplacian prior for the crosschannel relationship. The problem is solved by the alternating direction method of multipliers (ADMM), which alternately updates the blur kernel and sharpens intermediate MS images. Numerical experiments demonstrate that our approach is more robust to large misalignment errors and yields better super resolved MS images compared to state-of-the-art optimization-based and deep-learning-based algorithms.

IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)

BLIND MULTI-SPECTRAL IMAGE PAN-SHARPENING

Lantao Yu[1]*, Dehong Liu*[2]*, Hassan Mansour*[2]*, Petros T. Boufounos*[2]*, and Yanting Ma*[2]

[1] Rice University, Houston, TX, USA
[2] Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA

ABSTRACT

We address the problem of sharpening low spatial-resolution multi-spectral (MS) images with their associated misaligned high spatial-resolution panchromatic (PAN) image, based on priors on the spatial blur kernel and on the cross-channel relationship. In particular, we formulate the blind pan-sharpening problem within a multi-convex optimization framework using total generalized variation for the blur kernel and local Laplacian prior for the cross-channel relationship. The problem is solved by the alternating direction method of multipliers (ADMM), which alternately updates the blur kernel and sharpens intermediate MS images. Numerical experiments demonstrate that our approach is more robust to large misalignment errors and yields better super resolved MS images compared to state-of-the-art optimization-based and deep-learning-based algorithms.

Index Terms— blind image fusion, pan-sharpening, local Laplacian prior, total generalized variation

1. INTRODUCTION

1.1. Background

Blind multi-spectral (MS) image pan-sharpening aims to enhance the spatial resolution of a set of spatially low-resolution MS channels, covering a wide spectral range, using their corresponding misaligned spatially high-resolution panchromatic (PAN) image. Since the original MS and PAN images are typically captured by different sensors, from different view angles, or at different times, they are not guaranteed to be well-aligned with each other or to share the same blur kernel. Further, the parametric relationship between MS and PAN images is unclear since the spectrum of PAN image only covers a fraction of the entire spectra of MS image. In this paper, we aim to fuse MS and PAN images without knowledge of the misalignment, the blur kernel, or any parametric models of cross-channel relationship, and to obtain images with the spatial resolution of the PAN image and the spectral resolution of the MS images.

1.2. Related Work and Motivation

Image alignment is necessary in image pan-sharpening. Most model-based [1–3] and learning-based [4, 5] pan-sharpening methods typically assume that the input images are pre-registered and well-aligned. More recently introduced, blind pan-sharpening [6–8] aims to jointly perform both registration and pan-sharpening. Most blind pan-sharpening methods model low-resolution MS images as a blurred and downsampled version of a target high-resolution MS image with an unknown blur kernel, which incorporates the unknown misalignment, and assume that the target high-resolution MS image and the PAN image bear certain spatial relationship. The blind pan-sharpening approach typically solves an optimization problem which enforces certain constraints on the blur kernel and on the cross-channel relationship. However, these methods exhibit limited success in MS image pan-sharpening tasks.

In this paper, we exploit newly developed regularization functions in order to improve the performance of MS image pan-sharpening. First, recent advances in pan-sharpening using local gradient constraints (LGC) to regularize the cross-channel relationship has led to significant improvement when the blur kernel is known [9]. Thus, it is useful to explore related cross-channel priors for blind image pan-sharpening. Second, the commonly used total variation (TV)-based regularizer applied to the blur kernel often forces small gradients to be 0, resulting in non-trivial errors when the ground-truth blur kernel is smooth. Meanwhile, recent work demonstrates that Second-Order Total Generalized Variation (TGV^2) provides more flexible features than total variation [10]. Third, due to the non-convexity of the problem, some methods can be trapped in bad local minima when misaligned displacements are large, causing poor fusion performance [8]. Our paper exploits these developments to contribute the following:

- We propose a novel local Laplacian prior (LLP) to regularize the relationship between MS and PAN images, which offers better performance than LGC.
- We use TGV^2 to regularize the blur kernel, which offers more robust and accurate estimation of the blur kernel than existing TV-based priors.
- We adopt an initialization strategy for the the blur kernel that helps avoid undesirable local minima in the optimization.

2. PROBLEM FORMULATION

We use $\mathbf{X} \in \mathbb{R}^{hw \times N}$ to denote the measured low-resolution MS image with N spectral bands, where h, w is the height and width of each band, respectively. We denote the measured high-resolution PAN image as $\mathbf{Y} \in \mathbb{R}^{HW \times 1}$, where H, W are its height and width, respectively. The target, well-aligned and high-resolution, MS image of consistent sharpness with $\mathbf{Y}$ is denoted as $\mathbf{Z} \in \mathbb{R}^{HW \times N}$.

To reconstruct the target image $\mathbf{Z}$, we solve the following regularized inverse problem:

$$\min_{\mathbf{Z},\mathbf{u}} \frac{1}{2}\|\mathbf{X} - \mathbf{D}\mathbf{B}(\mathbf{u})\mathbf{Z}\|_{\mathrm{F}}^2 + \mathbf{R}_1(\mathbf{Z},\mathbf{Y}) + \mathbf{R}_2(\mathbf{u}), \qquad (1)$$

in which the first component is the data fidelity term, $\mathbf{u} \in \mathbb{R}^{n^2 \times 1}$ is the blur kernel, incorporating the misalignment and relative blur between $\mathbf{Z}$ and $\mathbf{X}$ prior to downsampling, $\mathbf{B}(\mathbf{u}) \in \mathbb{R}^{HW \times HW}$ is the Toeplitz matrix implementing the convolution due to the blur kernel $\mathbf{u}$, and $\mathbf{D} \in \mathbb{R}^{hw \times HW}$ is the downsampling operator. The second

Lantao Yu performed this work as an intern at MERL.

term $\mathbf{R}_1$ characterizes the relationship between $\mathbf{Z}$ and $\mathbf{Y}$ and the third term $\mathbf{R}_2$ regularizes the blur kernel.

2.1. Cross-Channel Image Prior, $\mathbf{R}_1$

The data-fidelity term in (1) constrains only the low-frequency components of $\mathbf{Z}$, to match those of $\mathbf{X}$. To recover the high frequency components, we incorporate information in $\mathbf{Y}$ using the regularizer $\mathbf{R}_1(\mathbf{Z}, \mathbf{Y})$. Specifically, motivated by recent advances in variational pan-sharpening [9] and guided image filtering [11], we use the local Laplacian prior (LLP) as a penalty term.

$$\mathbf{R}_1(\mathbf{Z}, \mathbf{Y}) = \frac{\lambda}{2} \sum_{i,j} \sum_{k \in \omega_j} \left([\mathcal{L}(\mathbf{Z}_i)]_{j,k} - a_{i,j} [\mathcal{L}(\mathbf{Y})]_{j,k} - c_{i,j} \right)^2, \tag{2}$$

where parameters are defined as follows: λ is a scalar factor; ω_j is the j^{th} square window of size $(2r+1) \times (2r+1)$ in a $H \times W$ image, with r an integer; k refers to the k^{th} element within the window, $k = 1, 2, \ldots, (2r+1)^2$; $a_{i,j}$ and $c_{i,j}$ are both constant coefficients of the linear affine transform in window ω_j, corresponding to the i^{th} band; $\mathbf{Z}_i$ is the i^{th} band of $\mathbf{Z}$, $\mathcal{L}(\cdot)$ is a function that computes the Laplacian of the input image, i.e., $\mathcal{L}(\mathbf{Z}) = \mathbf{Z} \circledast \mathbf{S}$, with

$$\mathbf{S} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}. \tag{3}$$

2.2. Blur Kernel Prior, $\mathbf{R}_2$

The estimation quality of the blur kernel significantly affects the quality of the reconstructed high-resolution MS. We assume this kernel, constraining the target high-resolution MS and the measured low-resolution MS, is smooth, band-limited, and has compact spatial support with non-vanishing tail. We regularize the blur kernel using the Second-Order Total Generalized Variation (TGV^2), given by

$$\mathbf{R}_2(\mathbf{u}) = \min_{\mathbf{p}} \{ \alpha_1 \|\nabla \mathbf{u} - \mathbf{p}\|_{2,1} + \alpha_2 \|\mathcal{E}(\mathbf{p})\|_{2,1} \} + \mathbf{I}_{\mathbb{S}}(\mathbf{u}), \tag{4}$$

where $\nabla \mathbf{u} = [\nabla_h \mathbf{u} \; \nabla_v \mathbf{u}] \in \mathbb{R}^{n^2 \times 2}$ are the gradients of $\mathbf{u}$, $\mathbf{p} = [\mathbf{p}_1 \; \mathbf{p}_2]$ is an auxiliary variable of the same size as $\nabla \mathbf{u}$,

$$\mathcal{E}(\mathbf{p}) = \begin{bmatrix} \nabla_h \mathbf{p}_1 & \dfrac{\nabla_v \mathbf{p}_1 + \nabla_h \mathbf{p}_2}{2} & \dfrac{\nabla_v \mathbf{p}_1 + \nabla_h \mathbf{p}_2}{2} & \nabla_v \mathbf{p}_2 \end{bmatrix}, \tag{5}$$

$\|\mathbf{X}\|_{2,1} = \sum_{i=1}^{n} \sqrt{\sum_{j=1}^{m} x_{i,j}^2}$, and α_1, α_2 are both scalars that control the regularization strength of $\mathbf{p}$'s approximation to $\nabla \mathbf{u}$ and of the partial derivatives of $\mathbf{p}$. $\mathbb{S} = \left\{ \mathbf{S} \in \mathbb{R}^{n^2 \times 1} | s_i \geqslant 0, \sum_i s_i = 1 \right\}$ is the a Simplex, and $\mathbf{I}_{\mathbb{S}}(\cdot)$ is its indicator function, which ensures that the computed blur kernel is non-negative and preserves the energy of the image, i.e., has sum equal to 1. We name our approach as Blind pan-sharpening with local Laplacian prior and Total generalized variation prior, or BLT in short.

3. PAN-SHARPENING ALGORITHM

3.1. Formulation of Lagrangian

Combining the optimization (1), with the regularizers in (2) and (4), we introduce constraints $\mathbf{x} = \nabla \mathbf{u} - \mathbf{p}$, $\mathbf{y} = \mathcal{E}(\mathbf{p})$, and $\mathbf{z} = \mathbf{u}$, and apply the classical augmented Lagrangian method by minimizing

$$\begin{aligned} &\Phi(\mathbf{Z}, \mathbf{u}, \mathbf{p}, \mathbf{A}, \mathbf{C}, \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{\Lambda_1}, \mathbf{\Lambda_2}, \mathbf{\Lambda_3}) \\ =& \sum_{i=1}^{N} [\frac{1}{2} \|\mathbf{X}_i - \mathbf{D}\mathbf{B}(\mathbf{z})\mathbf{Z}_i\|_2^2 \\ &+ \frac{\lambda}{2} \sum_j \sum_{k \in \omega_j} \left([\mathcal{L}(\mathbf{Z}_i)]_{j,k} - a_{i,j} [\mathcal{L}(\mathbf{Y})]_{j,k} - c_{i,j} \right)^2] \\ &+ \alpha_1 \|\mathbf{x}\|_{2,1} + \frac{\alpha_1 \mu_1}{2} \|\mathbf{x} - (\nabla \mathbf{u} - \mathbf{p}) - \mathbf{\Lambda}_1\|_{\mathrm{F}}^2 \\ &+ \alpha_2 \|\mathbf{y}\|_{2,1} + \frac{\alpha_2 \mu_2}{2} \|\mathbf{y} - \mathcal{E}(\mathbf{p}) - \mathbf{\Lambda}_2\|_{\mathrm{F}}^2 \\ &+ \mathbf{I}_{\mathbb{S}}(\mathbf{z}) + \frac{\mu_3}{2} \|\mathbf{z} - \mathbf{u} - \mathbf{\Lambda}_3\|_2^2 \end{aligned} \tag{6}$$

with $\mu_1, \mu_2, \mu_3 > 0$. We solve the problem using the alternating direction method of multipliers(ADMM) [12] by alternating between a succession of minimization steps and update steps.

3.2. Minimization and Update Solutions

The minimization subproblems of $\mathbf{x}$ and $\mathbf{y}$ are similar to each other and the solutions are given by component-wise soft-thresholding. The l^{th} row of $\mathbf{x}^{t+1}$ and $\mathbf{y}^{t+1}$ are update using

$$\mathbf{x}^{t+1}(l) = \mathrm{shrink}_2(\nabla \mathbf{u}(l) - \mathbf{p}(l) + \mathbf{\Lambda}_1(l), \frac{1}{\mu_1}), \tag{7}$$

$$\mathbf{y}^{t+1}(l) = \mathrm{shrink}_2(\mathcal{E}(\mathbf{p})(l) + \mathbf{\Lambda}_2(l), \frac{1}{\mu_2}), \tag{8}$$

where $\mathrm{shrink}_2(\mathbf{e}, t) = \max(\|\mathbf{e}\|_2 - t, 0) \frac{\mathbf{e}}{\|\mathbf{e}\|_2}$.

We use $\mathbf{B}(\mathbf{u})\mathbf{Z}$ to denote the operator implementing the convolution between $\mathbf{u}$ and $\mathbf{Z}$. Therefore, we can rewrite $\mathbf{B}(\mathbf{u})\mathbf{Z}$ as $\mathbf{B}(\mathbf{u})\mathbf{Z} = \mathbf{u} \circledast \mathbf{Z} = \mathbf{Z} \circledast \mathbf{u} = \mathscr{C}(\mathbf{Z})\mathbf{u}$, where $\mathscr{C}$ is a Toeplitz matrix corresponding to the convolution operation. Using this notation, the $\mathbf{z}$-subproblem first solves

$$\min_{\mathbf{z}} \sum_{i=1}^{N} \frac{1}{2} ||\mathbf{D}\mathscr{C}(\mathbf{Z}_i)\mathbf{z} - \mathbf{X}_i||_2^2 + \frac{\mu_3}{2} ||\mathbf{z} - \mathbf{u} - \mathbf{\Lambda}_3||_2^2 \tag{9}$$

using conjugate gradient descent [13] and then projects the solution onto the simplex $\mathbb{S}$ [14].

The $\{\mathbf{u}, \mathbf{p}\}$-subproblem minimizes

$$\min_{\mathbf{u},\mathbf{p}} \frac{\alpha_1 \mu_1}{2} \|\mathbf{x} - (\nabla \mathbf{u} - \mathbf{p}) - \mathbf{\Lambda}_1\|_{\mathrm{F}}^2 + \frac{\alpha_2 \mu_2}{2} \|\mathbf{y} - \mathcal{E}(\mathbf{p}) - \mathbf{\Lambda}_2\|_{\mathrm{F}}^2. \tag{10}$$

To solve the problem efficiently, we define $\mathbf{q} = [\mathbf{u}^\top \; \mathbf{p}_1^\top \; \mathbf{p}_2^\top]^\top$. By enforcing the first-order necessary conditions for optimality, we obtain the following linear equation:

$$\mathbf{\Sigma}\mathbf{q} = \mathbf{b}, \tag{11}$$

where, in the interest of space, we omit $\mathbf{\Sigma}$ and $\mathbf{b}$, other than noting that $\mathbf{\Sigma}$ is a diagonal block-Toeplitz matrix, diagonalized using the Fourier transform. Thus (11) can be efficiently solved [10].

The $\mathbf{A}, \mathbf{C}$-subproblem can be approximated as

$$\min_{a_{i,j}, c_{i,j}} \sum_j \sum_{k \in \omega_j} \left([\mathcal{L}(\mathbf{Z}_i)]_{j,k} - a_{i,j} [\mathcal{L}(\mathbf{Y})]_{j,k} - c_{i,j} \right)^2. \tag{12}$$

Similar to guided image filtering, $a_{i,j}$ and $c_{i,j}$ can be stably computed using $\mathcal{L}(\mathbf{Z}_i)$'s local window as the input image and $\mathcal{L}(\mathbf{Y})$'s local window as the guide image.

The $\mathbf{Z}$-subproblem in each individual channel is reformulated as

$$\min_{\mathbf{Z}_i} \frac{1}{2}\|\mathbf{DBZ}_i - \mathbf{X}_i\|_2^2 + \frac{\lambda}{2}\|\mathbf{LZ}_i - \hat{\mathbf{L}}_i^{\mathbf{z}}\|_2^2, \tag{13}$$

where $\hat{\mathbf{L}}_i^{\mathbf{z}}$ is the output of guided image filtering with input image $\mathcal{L}(\mathbf{Z}_i)$ and guide image $\mathcal{L}(\mathbf{Y})$, and $\mathbf{L}$ is Toeplitz matrix of the Laplacian. Equation (13) has a closed-form solution:

$$\mathbf{Z}_i = (\mathbf{B}^\top\mathbf{D}^\top\mathbf{DB} + \lambda\mathbf{L}^\top\mathbf{L})^{-1}(\mathbf{B}^\top\mathbf{D}^\top\mathbf{X}_i + \lambda\hat{\mathbf{L}}_i^{\mathbf{z}}). \tag{14}$$

Similar to the solution of (11), we use Fast Fourier Transform to accelerate the computation based on [15].

Finally, the update steps are given by:

$$\begin{cases} \mathbf{\Lambda_1}^{t+1} = \mathbf{\Lambda_1}^t + \mu(\nabla\mathbf{u}^{t+1} - \mathbf{p}^{t+1} - \mathbf{x}^{t+1}) \\ \mathbf{\Lambda}_2{}^{t+1} = \mathbf{\Lambda_2}^t + \mu(\mathcal{E}(\mathbf{p}^{\mathbf{t+1}}) - \mathbf{y}^{t+1}) \\ \mathbf{\Lambda}_3{}^{t+1} = \mathbf{\Lambda_3}^t + \mu(\mathbf{u}^{t+1} - \mathbf{z}^{t+1}). \end{cases} \tag{15}$$

3.3. Initialization of u

Due to the non-convexity of our problem, the initialization of $\mathbf{u}$ plays a crucial role in avoiding bad local minima, especially when the misalignment is large. To overcome this problem, we propose to treat the stacked PAN as the ground-truth MS in the data fidelity term constrained by N_0 low-resolution MS bands whose spectra overlap with PAN, and initialize $\mathbf{u}$ by solving the optimization problem:

$$\min_{\mathbf{u},\mathbf{p}} \sum_i^{N_0} \frac{1}{2}\|\mathbf{D}\mathscr{C}(\mathbf{Y})\mathbf{u} - \mathbf{X}_i\|_2^2 + \alpha_1\|\nabla\mathbf{u} - \mathbf{p}\|_{2,1} + \alpha_2\|\mathcal{E}(\mathbf{p})\|_{2,1} + \mathbf{I}_{\mathbb{S}}(\mathbf{u}). \tag{16}$$

4. NUMERICAL EXPERIMENTS

In our numerical experiments, we first verify the improvements due to each of the LLP and TGV^2 separately, then we demonstrate the performance of our method in a blind pan-sharpening experiment. Our dataset includes a 4-channel MS image (blue, green, red and infra-red) and a PAN image, both of spatial resolution 610×338. These images are synthesized from the *Pavia University* dataset: each MS channel or PAN is generated by a weighted linear combination of bands of hyperspectral imagery. We use the mean of Peak Signal-to-Noise Ratio of all reconstructed MS channels (Average PSNR) as the metric.

4.1. Verification of Priors

To verify the effect of the LLP, we first conduct an experiment on guided image upsampling, using the PAN image to guide the upsampling of the MS image. We assume the blur kernel is a δ-function and generate the input MS image by downsampling the ground-truth MS image by a factor of 2, both horizontally and vertically. Then, we treat $\mathbf{u}$ fixed and $\mathbf{R}_2(\mathbf{u})$ as 0 to solve for $\mathbf{Z}$. We compare the LLP with the LGC [9] in an Average PSNR sense. In this experiment, the LLP-reconstructed MS image has 37.57 dB PSNR vs 37.33 dB for the LGC-reconstructed one.

To verify TGV^2, we choose the 2D Gaussian kernel $\mathbf{K}(i,j) = e^{-[(i-x)^2+(j-y)^2]/(2\sigma^2)}$, where $-r \le i \le r$, $-r \le j \le r$, $n = 19$, $x = 1.33$, $y = 0.42$ and $\sigma = 2$ to blur the PAN image *West of Sichuan* from IKONOS, and downsample it by a factor of 4, both horizontally and vertically. To test the robustness to noise, we add additive white Gaussian, generating observed images at 10 dB, 20 dB, 30 dB, and 40 dB PSNR. We recover a kernel estimate by solving the optimization in (17) and (18), and estimate performance using the relative error $\epsilon_r = \|\mathbf{K} - \hat{\mathbf{K}}\|_{\mathrm{F}}/\|\mathbf{K}\|_{\mathrm{F}}$ as the metric. We compare our regularizer (TGV^2) with the widely used isotropic total variation ($\mathrm{TV}_{\mathrm{iso}}$) that solves

$$\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{Eu} - \mathbf{f}\|_2^2 + \alpha\|\nabla\mathbf{u}\|_{2,1} + \mathbf{I}_{\mathbb{S}}(\mathbf{u}). \tag{17}$$

Similarly, our regularizer can be written as:

$$\min_{\mathbf{u},\mathbf{p}} \frac{1}{2}\|\mathbf{Eu}-\mathbf{f}\|_2^2+\alpha_1\|\nabla\mathbf{u}-\mathbf{p}\|_{2,1}+\alpha_2\|\mathcal{E}(\mathbf{p})\|_{2,1}+\mathbf{I}_{\mathbb{S}}(\mathbf{u}). \tag{18}$$

In both of the above, $\mathbf{E}$ refers to the measurement matrix of $\mathbf{u}$, i.e. the vectorized 2D blur kernel. Each row of $\mathbf{E}$ stores the pixels from the input image, which convolve with the blur kernel to obtain a noise-free pixel. $\mathbf{f}$ is the vectorized form of the observed image. Table 1 lists the relative errors for both regularizers in all noise levels, with the smaller error in each case highlighted in bold. As evident, TGV^2 recovers a better estimate of the ground truth kernel and provides more robustness to noise.

PSNR(dB)	10	20	30	40
TV_{iso}	0.2904	0.1818	0.1008	0.0502
TGV^2	**0.1607**	**0.0940**	**0.0520**	**0.0288**

Table 1: Relative error of TV_{iso} and TGV^2 at different noise levels.

4.2. Blind Pan-sharpening

In this experiment, we simulate the input MS image via first low-pass filtering and then downsampling the ground-truth MS image by a factor of 2, both horizontally and vertically. Given the low-resolution input MS image and the high-resolution PAN image, we use our pan-sharpening algorithm to output the fused high-resolution MS image. To demonstrate our algorithm's effectiveness in fusing misaligned images, we conduct two experiments, setting the kernel's center to be $(0.87, 0.11)$ in Experiment 1 and $(5.87, 4.11)$ in Experiment 2, both with the same standard deviation (1.2), to simulate small and large misalignment. We compare our results with two baseline algorithms: HySure [6] and BHMIFGLR [8].

Figure 1 provides a qualitative evaluation of Experiment 1. The input PAN and low resolution MS are shown in parts (a) and (b) respectively, with high resolution ground-truth MS shown in (c). The fused results are presented in parts (d), (e), and (f), for each method, respectively. We observe that BHMIFGLR is prone to generating spurious textures and ignoring details. For example, in Fig. 1(d), the reconstructed texture on the right of the parallel white lines, along the diagonal direction, is not present in the ground truth image. Also, the left of the three parallel white lines along the diagonal direction was not reconstructed. In comparison, HySure managed to fuse images with large misalignment, but failed to preserve the details of edges and textures. In Fig. 1(e) we observe that the three parallel white lines in the ground-truth image are blurred, and the details on the yellow roof are not identifiable. Instead, our approach, shown in Fig. 1(f) is visually much sharper and preserves more detail compared to the baseline methods.

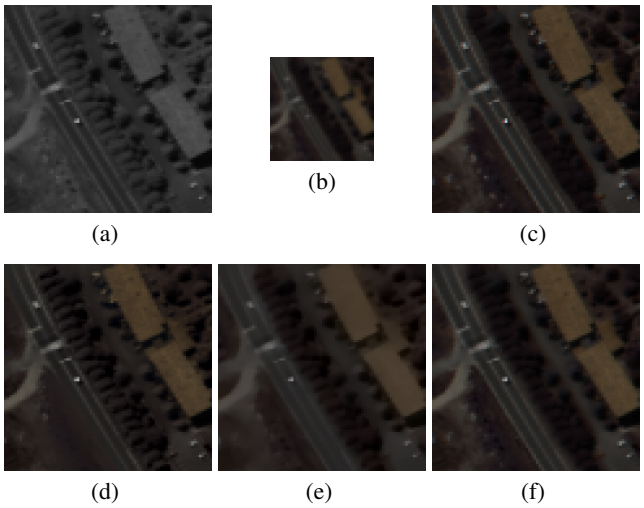

Fig. 1: Pan-sharpening results using different methods in Experiment 1: (a) the PAN image, (b) the input MS image (only RGB channels are shown), (c) Ground truth, and fused results using (d) BHMIFGLR, (e) HySure, and (f) BLT, respectively. BLT successfully preserves three sharp parallel lines and avoids the fake details.

Approach	BHMIFGLR	HySure	BLT
Exp. 1/Exp. 2	31.72/21.38	30.71/30.70	**37.40/37.40**

Table 2: Quantitative analysis of blind pan-sharpening results.

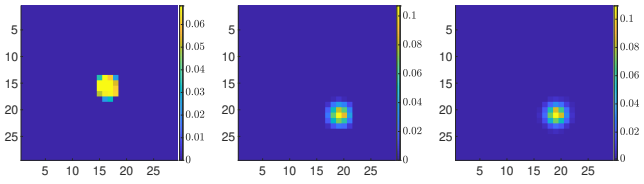

Fig. 2: Comparison of estimated kernels in Experiment 2 using (a) BHMIFGLR and (b) BLT with (c) Ground truth.

A quantitative performance comparison of the three algorithms is shown in Table 2. Our results are consistent at small/large misalignments and outperform the baseline algorithms by nearly 6 dB. In comparison, BHMIFGLR failed to fuse MS images with consistent performance; the blur kernel estimated in Experiment 2, shown in Fig. 2(a), was trapped in a local minimum or saddle point which is far away from the ground-truth, shown in Fig. 2(c). This is due to the large misalignment and the poor initial estimate of the blur kernel. Since the target MS image is aligned to the PAN, our approach treats the stacked PAN as the target MS in (16), thereby generating a reasonable initialization of the blur kernel that aligns well to the ground-truth and resulting to a good estimation of the blur kernel.

4.3. Comparison with a Deep Learning based Approach

We also compared our algorithm with deep learning based approaches, noting that our approach requires no training data. We choose the most recent work [5] as the benchmark and use the same dataset for comparison. Table 3 demonstrate the results of both algorithms. Overall, our algorithm yields much higher average PSNR, with a 1.07 dB average improvement over Unrolled PGD (UPGD) in four test images: *Moffett*, *Cuprite*, *L.A.*, and *C.F.*. Figure 3 also demonstrates qualitatively the performance of the two algorithms in recovering the RGB channels of the *L.A.* image, as well as the residual images in the green channel. As evident from these residuals, our approach outperforms UPGD, especially in smooth areas. The main reason for this discrepancy seems to be that it is more difficult for a learning algorithm to be trained in a large variety of conditions, including blur size, misalignment level, noise level, etc., requiring data to be provisioned for a wide variety of cases to provide robustness. In contrast, our approach adaptively determines a kernel for each test image set, providing more robustness, especially to the level of blurring under different blur conditions.

Test Images	Moffett	Cuprite	L.A.	C.F.	Mean
BLT	**39.94**	**41.17**	**38.53**	38.91	**39.64**
UPGD	38.17	39.02	37.77	**39.33**	38.57

Table 3: Quantitative analysis of blind pan-sharpening results using our proposed method and a learning-based method.

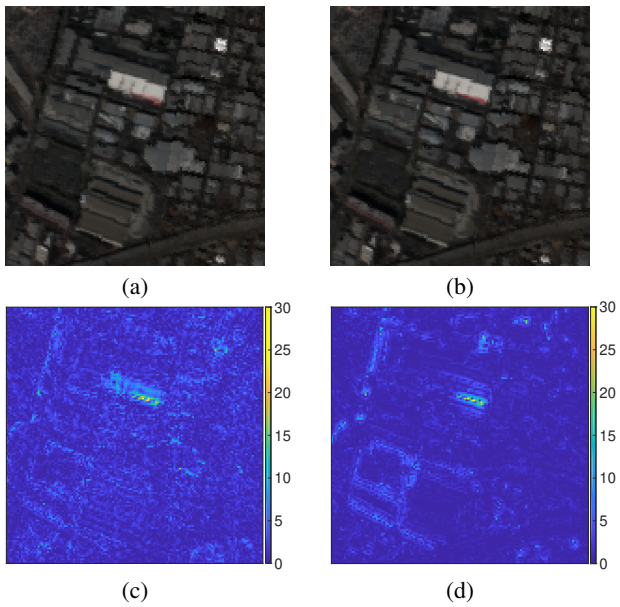

Fig. 3: Comparison of fused MS images in RGB channels using (a) UPGD and (b) BLT. (c) and (d) are the green channel residual images of (a) and (b) compared to the ground truth.

5. CONCLUSION

In this paper, we develop a novel method for misaligned multispectral (MS) image pan-sharpening based on the local Laplacian prior (LLP) and the Second-Order Total Generalized Variation (TGV^2). Numerical experiments show that our approach significantly outperforms state-of-the-art optimization-based and deep learning-based baselines. Moreover, our model has a better generalization ability than deep learning based methods, without external training data, providing flexibility and adaptability to deal with multi-spectral imagery from a large variety of imaging platforms.

6. REFERENCES

[1] Coloma Ballester, Vicent Caselles, Laura Igual, Joan Verdera, and Bernard Rougé, "A variational model for p+xs image fusion," *International Journal of Computer Vision*, vol. 69, no. 1, pp. 43–58, 2006.

[2] Hongyan Zhang, Wei He, Liangpei Zhang, Huanfeng Shen, and Qiangqiang Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, 2013.

[3] Zhiyuan Zha, Xin Liu, Xiaohua Huang, Henglin Shi, Yingyue Xu, Qiong Wang, Lan Tang, and Xinggan Zhang, "Analyzing the group sparsity based on the rank minimization methods," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2017, pp. 883–888.

[4] Dehong Liu and Petros T Boufounos, "Dictionary learning based pan-sharpening," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 2397–2400.

[5] Suhas Lohit, Dehong Liu, Hassan Mansour, and Petros T Boufounos, "Unrolled projected gradient descent for multispectral image fusion," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 7725–7729.

[6] Miguel Simões, José Bioucas-Dias, Luis B Almeida, and Jocelyn Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2014.

[7] Leon Bungert, David A Coomes, Matthias J Ehrhardt, Jennifer Rasch, Rafael Reisenhofer, and Carola-Bibiane Schönlieb, "Blind image fusion for hyperspectral imaging with the directional total variation," *Inverse Problems*, vol. 34, no. 4, pp. 044003, 2018.

[8] Chandrajit Bajaj and Tianming Wang, "Blind hyperspectral-multispectral image fusion via graph laplacian regularization," *arXiv preprint arXiv:1902.08224*, 2019.

[9] Xueyang Fu, Zihuang Lin, Yue Huang, and Xinghao Ding, "A variational pan-sharpening with local gradient constraints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10265–10274.

[10] Weihong Guo, Jing Qin, and Wotao Yin, "A new detail-preserving regularization scheme," *SIAM journal on imaging sciences*, vol. 7, no. 2, pp. 1309–1334, 2014.

[11] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397–1409, 2012.

[12] M. Tao and J. Yang, "Alternating direction algorithms for total variation deconvolution in image reconstruction," *TR0918,Department of Mathematics, Nanjing University*, 2009.

[13] Jonathan Richard Shewchuk et al., "An introduction to the conjugate gradient method without the agonizing pain," 1994.

[14] Weiran Wang and Miguel A Carreira-Perpinán, "Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application," *arXiv preprint arXiv:1309.1541*, 2013.

[15] Ningning Zhao, Qi Wei, Adrian Basarab, Nicolas Dobigeon, Denis Kouamé, and Jean-Yves Tourneret, "Fast single image super-resolution using a new analytical solution for $\ell_2 - \ell_2$ problems," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3683–3697, 2016.