

PAN-SHARPENING WITH MULTI-SCALE WAVELET DICTIONARY

Dehong Liu, Petros T. Boufounos

Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139
{*liudh,petrosb*}@merl.com

ABSTRACT

In satellite image processing, pan-sharpening is the fusion process in which a low resolution (LR) multi-spectral (MS) image is sharpened using the corresponding high resolution (HR) panchromatic (Pan) image to obtain a HR MS image. In this paper we propose a novel pan-sharpening method which combines the ideas of classical wavelet-based pan-sharpening with recently developed dictionary learning (DL) methods. The HR MS image is generated using wavelet-based pan-sharpening, regulated by promoting sparsity with respect to a dictionary. The dictionary is obtained using DL on the multi-scale wavelet tree vectors of the image to be pan-sharpened. A significant advantage of our approach compared to most DL-based approaches is that it does not require a large database of images on which to train the dictionary. Experiments on degraded satellite images demonstrate that our method significantly reduces color distortions and wavelet artifacts compared to the state of the art.

1. INTRODUCTION

A large number of optical satellite imaging systems produce two kinds of imagery: panchromatic (Pan) and multi-spectral (MS). Typically Pan imagery provides high spatial resolution but no color information, whereas MS imagery provides color spectrum information but reduced spatial resolution. However, for a large variety of applications, high resolution (HR) MS images are preferred. These HR MS images—not readily available from the satellite—can only be obtained by fusing the satellite Pan and MS images. This fusion process is referred to as pan-sharpening.

In the past decades, a large number of pan-sharpening methods have been developed. These methods can be divided into four categories: Intensity-Hue-Saturation (IHS) transform based ones, Principal Component Analysis (PCA) based ones, arithmetic combination based ones and wavelet-based ones [1, 2]. Generally, pan-sharpened images generated by methods in the first three categories have good spatial resolution, but distorted color spectrum [3]. On the other hand, images generated using wavelet-based methods exhibit relatively better color spectrum, but produce wavelet-induced artifacts. A variety of techniques have been developed to improve spatial and spectral accuracy, each specific to a particular fusion technique or image set [4].

Recent developments in compressive sensing, sparse representations and dictionary learning (DL) have provided new tools to address this problem [5, 6]. These approaches assume that the HR MS image is sparse in some basis or dictionary. The HR MS image is recovered using sparsity-promoting optimization, subject to data fidelity constraints derived from the available LR MS and HR Pan images. The choice of basis or dictionary is often critical in the performance of such techniques; it has been shown that a large number of sparsity-based approaches can benefit significantly from an appropriate sparsity-inducing dictionary learned from available data [7].

In this paper, we propose a novel pan-sharpening method which combines the ideas of classical wavelet-based pan-sharpening and dictionary learning. Specifically, our approach learns a dictionary in the wavelet domain and uses that dictionary in a sparsity promoting optimization problem to produce the HR MS image. In contrast to typical DL approaches, we learn the dictionary from a single HR Pan and LR MS image—the one to be pan-sharpened. Thus we do not require prior training on expensive and scarce large image datasets, as typically required by a number of classical DL approaches. Our work is motivated by recent results on multi-scale DL [8], as well as our prior work on pan-sharpening using image-domain DL [9]. Our earlier work has already demonstrated the potential of DL with significant improvements over the state of the art; the present paper improves these results.

Our approach improves standard wavelet-based pan-sharpening methods. These methods perform a discrete wavelet transform on the MS and Pan images to extract the low frequency coefficients from the MS image and the high frequency coefficients from the Pan image. These coefficients are then combined to create a fused wavelet representation. The final pan-sharpened image is generated by performing the inverse wavelet transform on the fused representation.

Our method re-visits how the wavelet coefficients are fused to form the pan-sharpened image. Instead of simply combining the coefficients in a naive way, we regulate the fusion by assuming that each wavelet-tree vector of the fused map can be sparsely represented by dictionary learned from the wavelet-tree vectors of the Pan image. The main principle in this assumption is that a dictionary learned from the wavelet transform of HR Pan image captures the structure of the image much better than the wavelet transform itself. The final coefficients are both consistent with the data and sparse in the learned dictionary, yielding a HR MS image with fewer color distortions and fewer artifacts.

The following sections provide more details on our method as well as experimental results validating our approach.

2. DICTIONARY LEARNING BASED PAN-SHARPENING

2.1. Sparse representations and dictionary learning

Natural images are compressible in the DCT or a wavelet domain, i.e., they can be sparsely represented with little distortion using the DCT or a wavelet basis. This structural property is the foundation for a large number of image acquisition, processing and compression methods. A significant advantage in using these transformations to representing a signal is their computationally efficient implementation. However, they do not always provide the best representation of the signal structure.

Recent work on dictionary learning (DL) has shown that a dictionary learned from example training data may result in even sparser representations for signals in a particular class or domain, compared

to a dictionary based on a pre-determined basis [7]. This domain-specific dictionary is able to capture in the learning process the similarities in the structure of the training data and exploit this structure in sparsely representing the signal.

Typical sparse approximation problems consider a signal $\mathbf{y} \in \mathbf{R}^{n \times 1}$ and attempt to represent it using $\mathbf{y} = \mathbf{D}\alpha$, where $\mathbf{D} \in \mathbf{R}^{n \times k}$ is the dictionary that contains the atoms as its columns and $\alpha \in \mathbf{R}^{k \times 1}$ is a sparse representation vector. The sparsity of α captures the structure of the signal with respect to the dictionary: the linear combination of a few atoms from \mathbf{D} is able to represent the signal \mathbf{y} with small error. One approach in determining the sparse approximation of a signal is to solve the optimization

$$\hat{\alpha} = \arg \min_{\alpha} \|\mathbf{y} - \mathbf{D}\alpha\|_2 \text{ s.t. } \|\alpha\|_0 \leq T_0, \quad (1)$$

where the ℓ_0 norm counts the number of non-zero coefficients of the representation and T_0 is the desired sparsity level. This optimization can be solved under certain conditions on the dictionary using a variety of methods such as Iterative Hard Thresholding (IHT) or CoSaMP [10, 11].

Dictionary learning uses a training set of signals to learn a dictionary \mathbf{D} that sparsifies the data. The underlying assumption is that other signals (testing data) in the same domain exhibit similar structure and will also produce good sparse representations in the learned dictionary. All training signals are assumed to be sparsely representable under \mathbf{D} , each with a different linear combination of few atoms. Using the same notation, DL attempts to optimize

$$\arg \min_{\mathbf{D}, \alpha} \|\mathbf{Y} - \mathbf{D}\alpha\|_F^2 \text{ s.t. } \forall l, \|\alpha_l\|_0 \leq T_0, \quad (2)$$

where \mathbf{Y} is a matrix of the training data, α is a sparse matrix with α_l representing the l^{th} data point, and $\|\cdot\|_F$ denotes the Frobenius norm. This is not a convex problem; most DL algorithms alternate between sparsely representing the training data using a given dictionary, and then updating that dictionary given the sparse representation of the data, until convergence. The best known such algorithm is the K-SVD [7].

2.2. Multi-scale wavelet dictionary

We aim to perform pan-sharpening using only one LR MS image and the corresponding HR Pan image by learning an appropriate dictionary. Our earlier work operates in the image-domain, learning a dictionary using small image patches [9]. Instead, in this work we learn a wavelet-domain dictionary using multi-scale wavelet tree vectors.

To build the dictionary atoms we consider a quad-tree structure with different sized blocks at each wavelet decomposition level. All levels are used to make up one joint global dictionary in a manner similar to [8]. We assume the Pan image has N times the spatial resolution and size of the LR MS image, and the LR MS image is composed of M spectral bands. Furthermore, we treat each color-band of the MS image independently.

We first interpolate the LR MS image to the size of the Pan image and then register the MS image with the Pan image using a sub-pixel registration method [12]. A K -level wavelet decomposition is performed on both the Pan and the interpolated MS images to form the wavelet-tree representation of the image. From this decomposition we obtain the detail coefficients at every scale to build the dictionary atoms. Each atom is composed of the coefficients in a small patch of the detail coefficients at the coarsest scale, together with the detail coefficients in the corresponding larger patches from the finer scales: $\mathbf{x}_{Pan} = [(\mathbf{x}_{Pan}^L)^T (\mathbf{x}_{Pan}^H)^T]^T$

and $\mathbf{x}_{MS} = [(\mathbf{x}_{MS}^L)^T (\mathbf{x}_{MS}^H)^T]^T$, where the subscripts ‘Pan’ and ‘MS’ denote Pan and MS images respectively, the superscripts ‘L’ and ‘H’ denote the coarsest scale and all the remaining finer scales respectively, and the superscript ‘T’ denotes the matrix transpose. We select the patches such that they tile the whole wavelet decomposition and use $\mathbf{X}_{Pan} = [(\mathbf{X}_{Pan}^L)^T (\mathbf{X}_{Pan}^H)^T]^T$ and $\mathbf{X}_{MS} = [(\mathbf{X}_{MS}^L)^T (\mathbf{X}_{MS}^H)^T]^T$ to denote the matrices containing all of them. Before fusion process, a scale factor of $w = \|\mathbf{X}_{Pan}^L\|_F / \|\mathbf{X}_{MS}^L\|_F$ is used for \mathbf{X}_{MS} to ensure the Frobenius norms of \mathbf{X}_{Pan}^L and \mathbf{X}_{MS}^L are matched.

Classical wavelet-based pan-sharpening methods fuse the coefficients using $[(\mathbf{X}_{MS}^L)^T (\mathbf{X}_{Pan}^H)^T]^T$ to estimate the transform of the HR MS image $\hat{\mathbf{X}}_{MS}$. Instead, our approach uses a dictionary to regulate the fusion, assuming that the HR MS image, estimated in $\hat{\mathbf{X}}_{MS}$, is sparse in the same dictionary.

To obtain a data-dependent dictionary \mathbf{D} , one may learn from the atoms of Pan image using the K-SVD method [7]

$$\mathbf{D} = \operatorname{argmin}_{\mathbf{D}, \alpha} \|\mathbf{X}_{Pan} - \mathbf{D}\alpha\|_F^2 \text{ s.t. } \forall l, \|\alpha_l\|_0 \leq T_0, \quad (3)$$

where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_L]$ is the set of sparse coefficient vectors corresponding to the L training vectors. Although a learned dictionary using this method performs well, the learning process is computationally expensive.

To improve the efficiency of DL-based pan-sharpening, an alternative way to obtain dictionary is to use localized tree vectors as dictionary atoms rather than learning from the whole tree vectors. This is motivated by the fact that for natural images, a patch of particular structure tends to recur often in its closest vicinity, and its frequency of recurrence decays rapidly as the distance from the patch increases [13]. Therefore, a localized dictionary may outperform a dictionary learned from the whole image and reduce training time. To exploit this property, for each tree vector of MS image, we define a localized dictionary to represent it. Let the spatial location of each tree vector be the root coordinate index in the coarsest level. The localized dictionary is composed of a set of tree vectors centered at spatial location (i_0, j_0) within radius r_0 , as

$$\mathbf{D}_{i_0, j_0} = \{ \{ \mathbf{x}_{Pan} \}_{i, j} | \sqrt{(i - i_0)^2 + (j - j_0)^2} \leq r_0 \}, \quad (4)$$

by which the wavelet-tree vector at the center location (i_0, j_0) of the MS image is sparsely represented. The localized dictionary is a related to full dictionary learning since it can be considered as a special case of the learned dictionary when setting the sparsity level $T_0 = 1$ and using a localized training set in (3).

2.3. Pan-sharpening using wavelet dictionary

Given learned dictionary \mathbf{D} , the reconstruction attempts to optimize the objective

$$\begin{aligned} \hat{\mathbf{X}}_{MS} = \operatorname{argmin}_{\mathbf{X}, \alpha} & \|\mathbf{X} - [(\mathbf{X}_{MS}^L)^T (\mathbf{X}_{Pan}^H)^T]^T\|_F^2 \\ & + \lambda \|\mathbf{X} - \mathbf{D}\alpha\|_F^2 \text{ s.t. } \forall l, \|\alpha_l\|_0 \leq T_1. \end{aligned} \quad (5)$$

The objective (5) enforces that (a) the coarse scale coefficients of the HR MS image are close to those of the LR MS image, (b) the finer scale coefficients are close to those of the HR Pan image and (c) the quad-tree vectors are sparse with respect to the learned dictionary. The weight λ controls the contribution of the sparse representation component. When $\lambda = 0$, the solution of (5) is identical to classical wavelet-based pan-sharpening. To compute (5) we use an iterative algorithm such as in [9], which iteratively updates α and \mathbf{X}

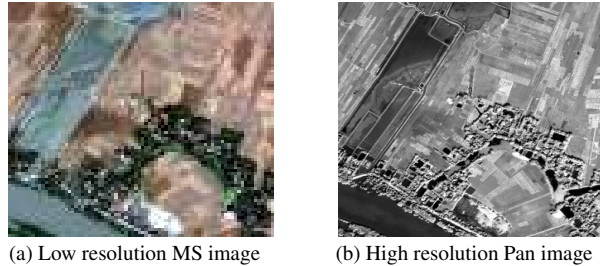


Fig. 1. Degraded QuickBird MS and Pan image

until convergence. Finally, we obtain the HR MS using the inverse wavelet transform on $\hat{\mathbf{X}}_{MS}$.

3. EXPERIMENTAL RESULTS

To demonstrate the advantages of our approach, we pan-sharpen degraded satellite images and compare the fused MS image with the original MS image before degradation (which we consider the ground truth). For our experiments we use Quickbird [14] and IKONOS [15] satellite images.

3.1. QuickBird Data

The QuickBird satellite system provides Pan images at 0.7m/pixel spatial resolution and MS images at 2.8m/pixel spatial resolution with four bands of near infra-red (NIR), red, green and blue. To generate test data we manually degrade the Pan and MS images using a low-pass filter (Gaussian kernel with unit variance) and down-sample by four to obtain one 2.8m/pixel Pan image and one 11.2m/pixel MS image. To evaluate the algorithm we pan-sharpen the degraded images with 3-level wavelet decomposition and compare the results with the original 2.8m/pixel MS images. Fig. 1 shows an example of the degraded QuickBird MS and Pan images of Sundarbans. The ground truth and pan-sharpened images for a number of methods are shown in Fig. 2. For comparison we demonstrate the classical wavelet-based fusion approach, our earlier efforts using image-domain DL [9], and wavelet-domain DL as described above with weight $\lambda = 0.1$, localized dictionary using $r_0 = 5$ and sparsity level $T_1 = 8$.

Classical wavelet-based pan-sharpening result exhibits many artifacts. Our earlier approach using image-domain DL has very good spatial resolution but noticeable color distortions, especially in the dark green area. Using wavelet domain DL we preserve the color spectrum with minor artifacts. For a quantitative comparison, Table 1 assesses the quality of the reconstructed images using a variety of metrics: the spectral angle mapper (SAM), the correlation coefficient (CC), the signal-to-noise ratio (SNR) and the peak signal-to-noise ratio (PSNR) [16]. As evident in the table, the wavelet-domain DL based Pan-sharpening outperforms the other two methods. Our method implemented in Matlab requires roughly 20 minutes of processing on a 3.16GHz Dual-core PC to pan-sharpen a 320×320 -pixel image with the K-SVD learning process in (3). The time is reduced to less than 1 minute using a localized dictionary in (4).

3.2. IKONOS Data

The IKONOS satellite captures 1m/pixel Pan image and 4m/pixel MS image covering NIR, red, green, and blue bands as well [15]. As with the QuickBird data, we perform Pan-sharpening on degraded

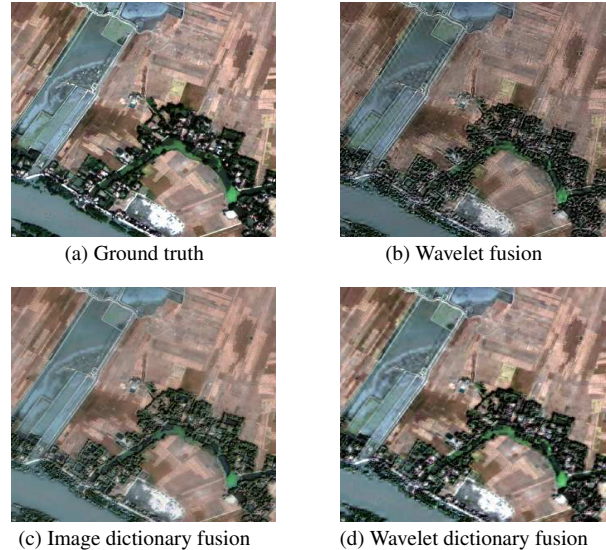


Fig. 2. (a) Ground truth and Pan-sharpened QuickBird images using (b) Wavelet fusion, (c) Image dictionary fusion, (d) Wavelet dictionary fusion.

Table 1. Comparison of fusion results with degraded QuickBird data

		Wavelet	Image DL	Wavelet DL
SAM		0.1445	0.1066	0.0941
	nir	0.9156	0.9625	0.9744
CC	r	0.8857	0.9338	0.9508
	g	0.8559	0.9190	0.9345
	b	0.8368	0.9097	0.9227
	all			
SNR	nir	14.79	17.97	19.92
	r	13.68	15.98	17.49
	g	13.60	15.84	17.21
	b	12.41	14.96	15.86
	all	13.51	16.00	17.33
PSNR	nir	21.60	24.78	26.74
	r	18.93	21.23	22.74
	g	19.39	21.63	23.00
	b	18.59	21.15	22.05
	all	19.49	21.97	23.30

images and compare with the original MS image. To evaluate the algorithm we pan-sharpen the degraded images with 3-level wavelet decomposition and compare the results with the original 4m/pixel MS images. We show in Fig. 3 an example of the degraded IKONOS MS and Pan images of Sichuan, China, and in Fig. 4 the corresponding ground truth and pan-sharpened results using the same approaches and parameters as with the QuickBird images. Again, using wavelet domain DL we preserve better color spectrum than the image-domain DL based Pan-sharpening results. For quantitative comparison, Table 2 lists the SAM, the CC, the SNR and the PSNR [16]. It is evident that the wavelet-domain DL based Pan-sharpening outperforms the other two methods.

4. CONCLUSION

We propose a novel Pan-sharpening process which obtains the image structure using multi-scale wavelet DL, and exploits that structure to produce the pan-sharpened high-resolution images. In contrast to a large number of DL-based applications, our approach learns

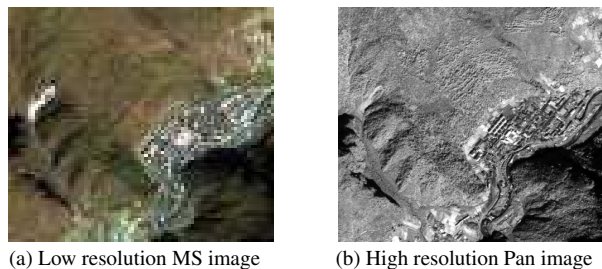


Fig. 3. Degraded IKONOS MS and Pan image

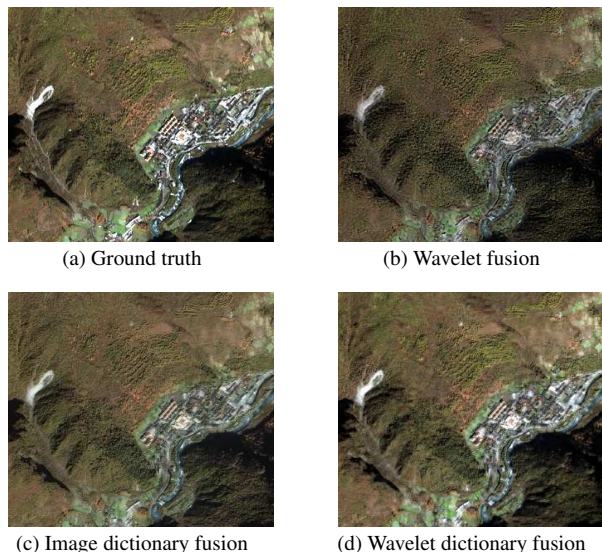


Fig. 4. (a) Ground truth and Pan-sharpened IKONOS images using (b) Wavelet fusion, (c) Image dictionary fusion, (d) Wavelet dictionary fusion.

the dictionary from the image to be pan-sharpened and does not require a large image database for training. We also proposed a localized DL modification which significantly reduces training complexity and improves the results.

Our experiments demonstrate the improvements due to our approach compared to classical wavelet-based methods as well as our earlier improvements using image-domain DL. The improvements are evident in a number of figures of merit including the spectral angle mapper (SAM), the correlation coefficient(CC), the signal-to-noise ratio (SNR), and the peak signal-to-noise ratio (PSNR).

5. REFERENCES

- [1] Yun Zhang, "Understanding image fusion," *Photogrammetric Engineering and Remote Sensing*, pp. 657–661, June 2004.
- [2] Qian Du, Oguz Gungor, and Jie Shan, "Performance evaluation for pan-sharpening techniques," in *IEEE Geosci. Remote Sens. Symposium*, 2005, pp. 4264–4266.
- [3] Claire Thomas, Thierry Ranchin, Lucien Wald, and Jocelyn 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(5), pp. 1301–1312, May 2008.
- [4] Luciano Alparone, Lucien Wald, Jocelyn Chanussot, Claire Thomas, Paolo Gamba, and Lori Mann Bruce, "Comparison of

Table 2. Comparison of fusion results with degraded IKONOS data

	Wavelet	Image domain DL	Wavelet domain DL
SAM	.1617	.1203	.1161
CC	nir	.9195	.9543
	r	.8756	.9134
	g	.8838	.9215
	b	.8260	.8761
SNR	nir	15.59	17.47
	r	13.05	14.54
	g	13.00	14.65
	b	9.87	11.76
	all	13.16	14.88
PSNR	nir	21.17	23.05
	r	20.61	22.09
	g	21.10	22.75
	b	20.06	21.95
	all	20.71	22.43

pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. remote sens.*, vol. 45(10), pp. 3012–3021, October 2007.

- [5] Atul Divekar and Okan Ersoy, "Image fusion by compressive sensing," in *Geoinformatics, 2009 17th International Conference on*, 2009.
- [6] Shutao Li and Bin Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. remote sens.*, vol. 49(2), pp. 738–746, February 2011.
- [7] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Sig. Proc.*, vol. 54(11), pp. 4311–4322, 2006.
- [8] Boaz Ophir, Michael Lustig, and Michael Elad, "Multi-scale dictionary learning using wavelets," *IEEE Journal of Selected Topics in Sig. Proc.*, vol. 5(5), pp. 1014–1024, 2011.
- [9] Dehong Liu and Petros T. Boufounos, "Dictionary learning based pan-sharpening," in *IEEE International Conference on Acoustic, Speech, and Signal processing (ICASSP)*, 2012.
- [10] T. Blumensath and M.E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.
- [11] D. Needell and J.A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [12] Manuel Guizar-Sicairos, Samuel T. Thurman, and James R. Fienup, "Efficient subpixel image registration algorithms," *Optics Letters*, vol. 33(2), pp. 156–158, 2008.
- [13] Maria Zontak and Michal Irani, "Internal statistics of a single natural image," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2011, pp. 977–984.
- [14] DigitalGlobe, "QuickBird data repository," available at <http://www.glcf.umiacs.umd.edu/data/quickbird/>.
- [15] DigitalGlobe, "IKONOS data repository," available at <http://www.glcf.umiacs.umd.edu/data/ikonos/>.
- [16] Melissa Strait, Sheida Rahmani, and Daria Markurjev, "Evaluation for pan-sharpening methods," Tech. Rep., 2008.