Fusion of Multi-band Images Using Bayesian Approaches: Beyond Pansharpening

Jean-Yves Tourneret[1]

[1] University of Toulouse, IRIT/INP-ENSEEIHT & TéSA, Toulouse, France

Cambridge, July 12, 2018

Panchromatic Image (50cm)

Thanks to Mathias Ortner from Airbus Defence and Space

Multispectral Image (2m)

Thanks to Mathias Ortner from Airbus Defence and Space

Pansharpened Image (50cm)

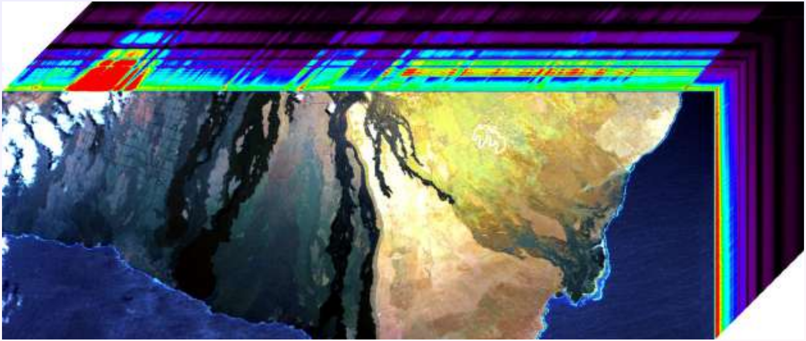
Thanks to Mathias Ortner from Airbus Defence and Space

Hyperspectral Imagery

Hyperspectral Images

- Spectral: same scene observed at different wavelengths
- Spatial: pixel represented by a vector of hundreds of measurements.

Hyperspectral Cube



Fusion of Multiband Images

(a) (b) (c)

(a) Hyperspectral Image (size: 99 $\times$ 46 $\times$ 224, res.: 80m $\times$ 80m) (b) Panchromatic Image (size: 396 $\times$ 184 $\times$ 1 res.: 20m $\times$ 20m) (c) Target (size: 396 $\times$ 184 $\times$ 224 res.: 20m $\times$ 20m)

Name	AVIRIS (HS)[1]	SPOT-5 (MS)	Pleiades (MS)	WorldView-3 (MS)
Res. (m)	20	10	2	1.24
# bands	224	4	4	8

[1] R. O. Green *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. of Environment*, 1998.

Summary

State-of-the-Art
- Component Substitution
- MultiResolution Analysis
- Bayesian Methods
- Matrix Factorization

Revisited Bayesian Fusion

Sparse Prior Based on Dictionary Learning

Fast fUsion Based on a Sylvester Equation (FUSE)

Conclusions

References


A Critical Comparison Among Pansharpening Algorithms

Review papers

- L. Alparone *et als.*, “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, Sep. 2007.
- G. Vivone *et als.*, “A Critical Comparison Among Pansharpening Algorithms,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565-2586, May 2015.
- L. Loncan *et als.*, “Hyperspectral Pansharpening: A Review,” *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27-46, Sept. 2015.

Classes of Existing Methods

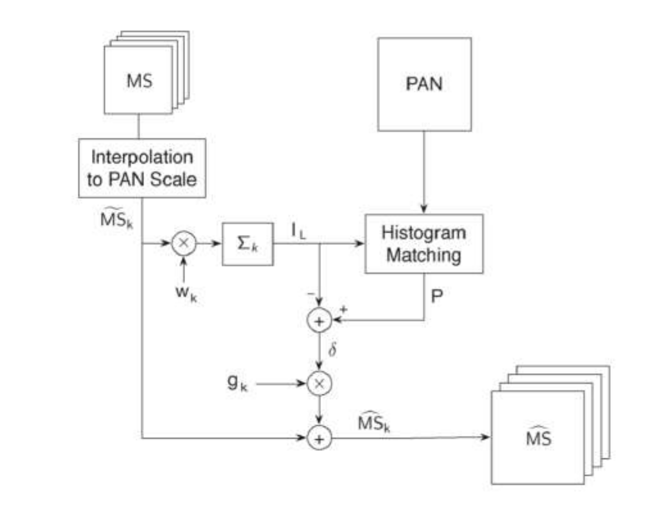

A Critical Comparison Among Pansharpening Algorithms


Hyperspectral Pansharpening: A Review

- Component Substitution
- Multiresolution Analysis
- Bayesian Inference
- Matrix Factorization

Outline

Component Substitution



Flowchart of CS Methods [Vivone2015]

Component Substitution

Principle

$$\widehat{\mathrm{MS}}_k = \widetilde{\mathrm{MS}}_k + g_k(\mathrm{P} - \mathrm{I}_L) \quad \text{with} \quad I_L = \sum_{k=1}^{m_\lambda} w_k \widetilde{\mathrm{MS}}_k \tag{1}$$

- Interpolate the MS Image
- Add details $P - I_L$ controlled by injection gains g_k

Strategies for choosing the weights w_k and gains g_k

- Intensity, Hue, Saturation (IHS), GHIS
- Principal component analysis
- High-pass filter
- Optimization of a global distorsion using genetic algorithms
- Optimization by linear regression

Component Substitution

Some references

- R. Haydn *et al.*, “Application of the IHS color transform to the processing of multisensor data and image enhancement,” in *Proc. Int. Symp. Remote Sens. Arid and Semi-Arid Lands*, Cairo, Egypt, pp. 599-616, 1982.
- P. S. Chavez *et al.*, “Extracting spectral contrast in Landsat thematic mapper image data using selective principal component analysis,” *Photogramm. Eng. Remote Sens.*, vol. 55, no. 3, pp. 339-348, March 1989.
- P. S. Chavez *et al.*, “Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic,” *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295-303, March 1991.
- A. Garzelli *et al.*, “Fusion of panchromatic and multispectral images by genetic algorithms,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Denver, Colorado, pp. 3810-3813, 2006.
- B. Aiazzi *et al.*, “Improving component substitution pansharpening through multivariate regression of MS+PAN data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230-3239, Oct. 2007.

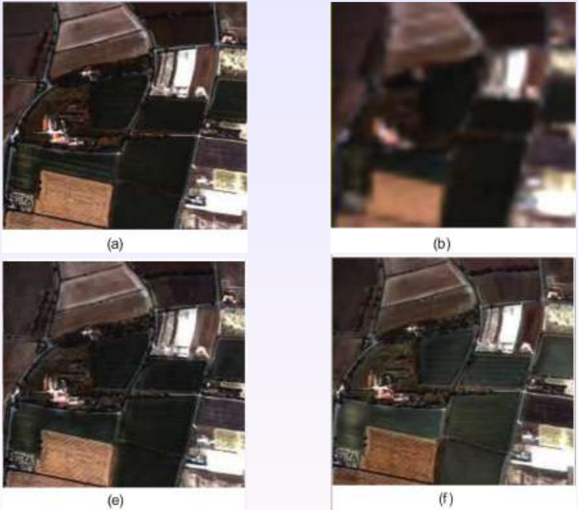
Image Characteristics

TABLE 2. CHARACTERISTIC OF THE THREE DATASETS.

DATASET	DIMENSIONS	SPATIAL RES	N	INSTRUMENT
Moffett	PAN 185×395 HS 37×79	20m 100m	224	AVIRIS
Camargue	PAN 500×500 HS 100×100	4m 20m	125	HyMap
Garons	PAN 400×400 HS 80×80	4m 20m	125	HyMap

Data Characteristics [Loncan2015]

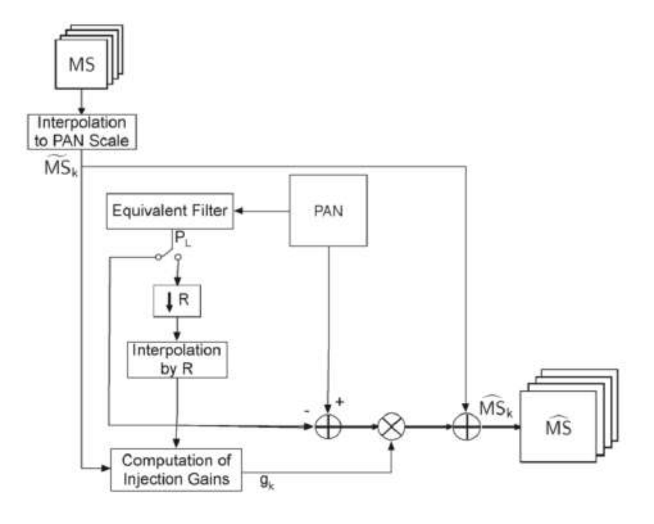
Component Substitution



Examples of fusion results using CS methods. (a) Reference, (b) Interpolated HS image, (e) Gram-Schmidt, (f) PCA.

Outline

MultiResolution Analysis (MRA)



Flowchart of MRA Methods [Vivone2015].

MultiResolution Analysis

Principle

$$\widehat{\mathrm{MS}}_k = \widetilde{\mathrm{MS}}_k + g_k(\mathrm{P} - \mathrm{P}_L), \quad k = 1, \ldots, m_\lambda \tag{2}$$

where P_L is a low-pass version of P.

Strategies for constructing P_L and choosing the gains g_k

- Smoothing filter-based intensity modulation (SFIM)

$$\mathrm{P}_L = \mathrm{P} * h_{\mathrm{LP}}$$

  where h_{LP} can be a box, Gaussian or Laplacian filter.
- Pyramidal decompositions: low-pass filter, wavelets, ...
- High-pass modulation paradigm for the gains

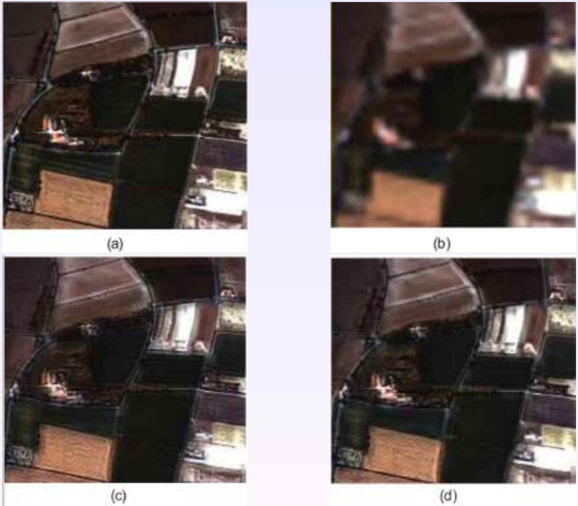
$$g_k = \widetilde{\mathrm{MS}}_k \left(\frac{1}{m_\lambda} \sum_{k=1}^{m_\lambda} \widetilde{\mathrm{MS}}_k \right)^{-1}, \quad k = 1, \ldots, m_\lambda.$$

MultiResolution Analysis

Some references

- J. G. Liu, "Smoothing filter based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461-3472, Dec. 2000.
- B. Garguet-Duport, *et al.*, "The use of multiresolution analysis and wavelet transform for merging SPOT panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 62, no. 9, pp. 1057-1066, Sep. 1996.
- M. González-Audícana *et al.*, "Comparison between Mallat's and the á trous discrete wavelet transform-based algorithms for the fusion of multispectral and panchromatic images," *Int. J. Remote Sens.*, vol. 26, no. 3, pp. 595-614, Feb. 2005.
- G. Vivone *et als.*, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geosci. Remote Sens. Lett.*, Denver, Colorado, pp. 3810-3813, 2006.
- X. Otazu *et als.*, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376-2385, Oct. 2005.
- B. Aiazzi *et als.*, "MTF-tailored multiscale fusion of high-resolution MS and PAN imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591-596, May 2006.

Multiresolution Analysis



Examples of fusion results using MRA methods. (a) Reference, (b) Interpolated HS image, (c) SFIM, (d) Generalized Laplacian Pyramid.

Outline

Bayesian Methods

Observation models

- Vectorized hyperspectral (HS) image

$$\mathbf{y}_{\mathrm{H}} = \mathbf{W}\boldsymbol{x} + \mathbf{n}_{\mathrm{H}}$$

  where $\mathbf{x} \in \mathbb{R}^{m_\lambda n}$, $\mathbf{y}_{\mathrm{H}}, \mathbf{n}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda m}$ (with $m < n$), $\mathbf{W} \in \mathbb{R}^{m_\lambda m \times m_\lambda n}$ performs spatial averaging and downsampling and $\mathbf{n}_{\mathrm{H}} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{n}_{\mathrm{H}}})$.
- Vectorized Panchromatic (PAN) or Multispectral (MS) image: $(\mathbf{y}_{\mathrm{M}}, \mathbf{x})$ jointly Gaussian. Thus

$$\mathbf{x}|\mathbf{y}_{\mathrm{M}} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{M}}, \mathbf{C}_{\mathrm{M}})$$

MAP estimator

- Direct form

$$\widetilde{\mathbf{x}} = \left[\mathbf{W}^T \mathbf{C}_{\mathbf{n}_{\mathrm{H}}}^{-1} \mathbf{W} + \mathbf{C}_{\mathrm{M}}^{-1}\right]^{-1} \left[\mathbf{W}^T \mathbf{C}_{\mathbf{n}_{\mathrm{H}}}^{-1} \mathbf{y}_{\mathrm{H}} + \mathbf{C}_{\mathrm{M}}^{-1} \boldsymbol{\mu}_{\mathrm{M}}\right]$$

- After matrix inversion lemma

$$\widetilde{\mathbf{x}} = \boldsymbol{\mu}_{\mathrm{M}} + \mathbf{C}_{\mathrm{M}} \mathbf{W}^T \left[\mathbf{W} \mathbf{C}_{\mathrm{M}} \mathbf{W}^T + \mathbf{C}_{\mathbf{n}_{\mathrm{H}}}\right]^{-1} (\mathbf{y}_{\mathrm{H}} - \mathbf{W}\boldsymbol{\mu}_{\mathrm{M}})$$

  Inversion of an $m_\lambda m \times m_\lambda m$ (instead of an $m_\lambda n \times m_\lambda n$) matrix.

Determination of μ_M and $\mathbf{C}_\mathrm{M}$

Using the previous observation models (Hardie, 2004)

- A priori mean of the target image estimated using a spectral interpolation of the PAN image
- Conditional independence $\Rightarrow$ block diagonal matrix $\mathbf{C}_\mathrm{M}$. Estimation of the diagonal matrices assuming adjacent pixels share the same covariance matrix. Adjacency determined by using clustering.

Using the stochastic mixing model (Eismann, 2005)

$$\mathbf{y}_{i,\mathrm{H}} = \sum_{k=1}^{K} a_{i,k}\mathbf{m}_k,$$

with $a_{i,k} > 0$, $\sum_{k=1}^{K} a_{i,k} = 1$ and $\mathbf{m}_k$ is a Gaussian vector.

Using a wavelet decomposition of the HS image (Zhang, 2009)

- Reduces spatial correlation
- Allows a separate estimation of the covariance matrices at different resolution levels

Bayesian Methods

References

- R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174-1184, Sep. 2004.
- M. T. Eismann and R. C. Hardie "Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 455-465, March 2005.
- Y. Zhang, S. De Backer and P. Scheunders, "Noise-Resistant Wavelet-Based Bayesian Fusion of Multispectral and Hyperspectral Images," *IEEE Trans. Geosci. and Remote Sens.*, vol. 47, no. 11, pp. 3834-3843, Nov. 2009.

Outline

Matrix Factorization

Observation models

- Hyperspectral (HS) image

$$\mathbf{Y}_{\mathrm{H}} = \mathbf{X}\mathbf{W} + \mathbf{N}_{\mathrm{H}}$$

where $\mathbf{X} \in \mathbb{R}^{m_\lambda \times n}$, $\mathbf{Y}_{\mathrm{H}}, \mathbf{N}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ (with $m < n$), $\mathbf{W} \in \mathbb{R}^{n \times m}$ performs spatial averaging and downsampling. Note that the distribution of $\mathbf{N}_{\mathrm{H}}$ does not need to be specified.

- Panchromatic (PAN) or Multispectral (MS) image

$$\mathbf{Y}_{\mathrm{M}} = \mathbf{R}\mathbf{X} + \mathbf{N}_{\mathrm{M}}$$

where $\mathbf{Y} \in \mathbb{R}^{n_\lambda \times n}$, $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ is the spectral response of the MS sensor and $\mathbf{N}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times m}$

Linear mixing model

- Target image

$$\mathbf{X} = \mathbf{M}\mathbf{A} + \mathbf{N}$$

where $\mathbf{M} \in \mathbb{R}^{m_\lambda \times p}$ and $\mathbf{A} \in \mathbb{R}^{p \times n}$ are the endmember and abundance matrices and p is the number of spectral signatures.

Matrix Factorization

New observation models

$$\mathbf{Y}_{\mathrm{H}} = \mathbf{M}\mathbf{W}_{\mathbf{A}} + \text{noise}, \quad \mathbf{Y}_{\mathrm{M}} = \mathbf{M}_{\mathbf{R}}\mathbf{A} + \text{noise}$$

where $\mathbf{W}_{\mathbf{A}} = \mathbf{A}\mathbf{W}$ is the spatially degraded abundance matrix and $\mathbf{M}_{\mathbf{R}} = \mathbf{R}\mathbf{M}$ is the spectrally degraded endmember matrix.

Strategy

Unmix the HS and MS images alternatively using nonnegative matrix factorizations (NMF)

- Step 1: estimate the endmember matrix $\mathbf{M}$ and $\mathbf{W}_{\mathbf{A}}$ by applying NMF to the HS image (initialized by the vertex component analysis (VCA))
- Step 2: estimate the abundance matrix $\mathbf{A}$ and $\mathbf{M}_{\mathbf{R}}$ by applying NMF to the MS image
- Step 2: fusion

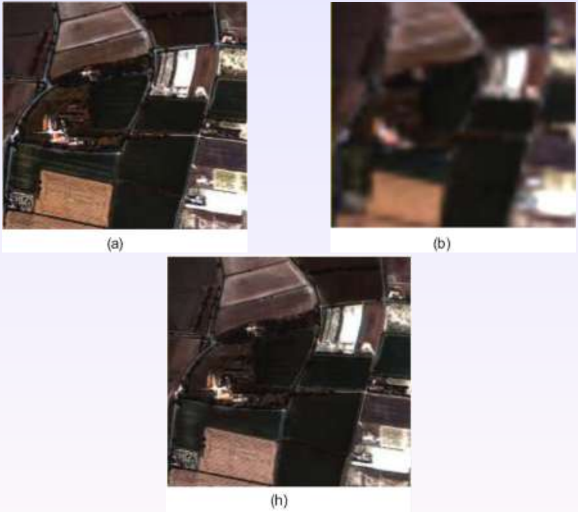
$$\widehat{X} = \widehat{\mathbf{M}}\widehat{\mathbf{A}}$$

Matrix Factorization

References

- O. Berné, A. Tielens, P. Pilleri, and C. Joblin, "Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy," in *Proc. IEEE Workshop on Hyperspectral Image Signal Process.*, Reykjavik, Iceland, pp. 1-4, June 2010.
- N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528-537, Feb. 2012.
- B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu, "Spatial and spectral image fusion using sparse matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1693-1704, 2014.

Matrix Factorization

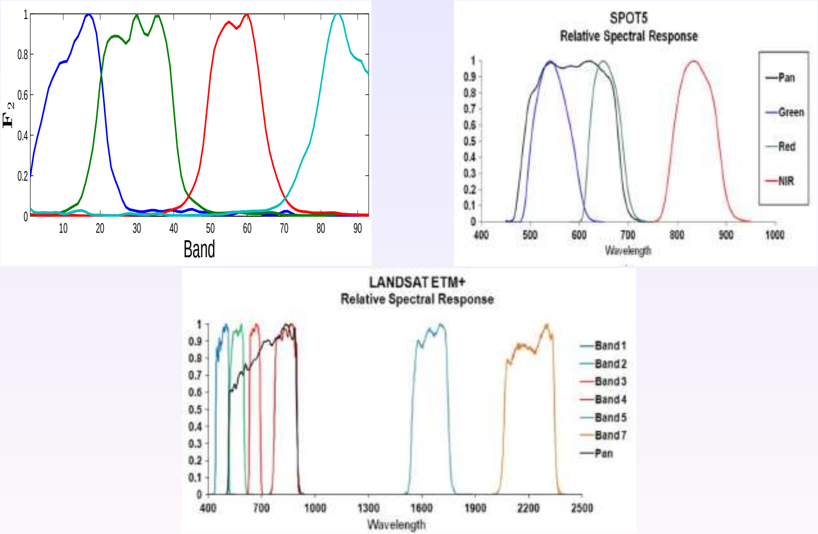


Examples of fusion results using matrix factorization (a) Reference, (b) Interpolated HS image, (h) Coupled NMF.

Summary

Outline

Motivations: Use the Spectral Response of the PAN or MS Sensor

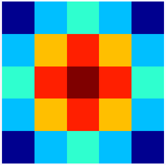


Spectral Responses. (a) IKONOS, (b) SPOT5, (c) LANDSAT.

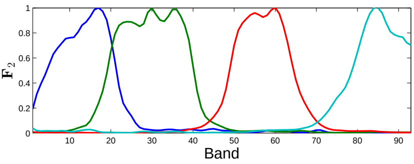
Forward model

$$\mathbf{Y}_{\mathrm{H}} = \mathbf{XBS} + \mathbf{N}_{\mathrm{H}}, \quad \mathbf{Y}_{\mathrm{M}} = \mathbf{RX} + \mathbf{N}_{\mathrm{M}}$$

- **$\mathbf{X} \in \mathbb{R}^{m_\lambda \times n}$: full resolution unknown image**
- $\mathbf{Y}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ and $\mathbf{Y}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$: observed HS and MS images
- $\mathbf{B} \in \mathbb{R}^{n \times n}$: cyclic convolution operator acting on the bands
- $\mathbf{S} \in \mathbb{R}^{n \times m}$: downsampling matrix
- $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$: spectral response of the MS sensor
- $\mathbf{N}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ and $\mathbf{N}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$: HS and MS noises, assumed to be a band-dependent Gaussian sequences



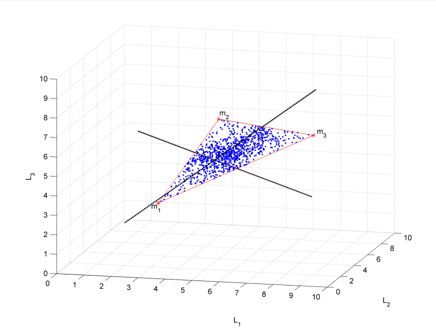
(d) Kernel of **B** (Spatial blurring)



(e) **R** (Spectral blurring)

Reparameterization

Dimensionality reduction



Projection of the data **X** in a lower-dimensional subspace ($\mathbb{R}^{\widetilde{m}_\lambda}$): $\mathbf{X} = \mathbf{HU}$, where **H** is an $\widetilde{m}_\lambda \times m_\lambda$ projection matrix[2].

[2] J. M. Bioucas-Dias *et al.*, "Hyperspectral subspace identification," *IEEE Trans. Geosci. and Remote Sens.*, vol. 46, no. 8, pp. 2435-2445, 2008.

Likelihoods

- Likelihood of the observations[3]

$$\mathbf{Y}_{\mathrm{H}}|\mathbf{U}, s^2_{\mathrm{H}} \sim \mathcal{MN}_{m_\lambda,m}(\mathbf{HUBS}, \mathrm{diag}\left(s^2_{\mathrm{H}}\right), \mathbf{I}_m)$$
$$\mathbf{Y}_{\mathrm{M}}|\mathbf{U}, s^2_{\mathrm{M}} \sim \mathcal{MN}_{n_\lambda,n}(\mathbf{RHU}, \mathrm{diag}\left(s^2_{\mathrm{M}}\right), \mathbf{I}_n)$$

where $s^2_{\mathrm{H}} = \left[s^2_{\mathrm{H},1}, \ldots, s^2_{\mathrm{H},m_\lambda}\right]^T$ and $s^2_{\mathrm{M}} = \left[s^2_{\mathrm{M},1}, \ldots, s^2_{\mathrm{M},n_\lambda}\right]^T$.

- Joint likelihood

$$f\left(\mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}|\mathbf{U}, \boldsymbol{s}^2\right) = f\left(\mathbf{Y}_{\mathrm{H}}|\mathbf{U}, s^2_{\mathrm{H}}\right) f\left(\mathbf{Y}_{\mathrm{M}}|\mathbf{U}, s^2_{\mathrm{M}}\right)$$

with $\boldsymbol{s}^2 = \left\{s^2_{\mathrm{H}}, s^2_{\mathrm{M}}\right\}$

[3] The probability density function of a matrix normal distribution is defined by

$$p(\mathbf{X}|\mathbf{M}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c) = \frac{\exp\left(-\frac{1}{2}\,\mathrm{tr}\left[\boldsymbol{\Sigma}_c^{-1}(\mathbf{X}-\mathbf{M})^T\boldsymbol{\Sigma}_r^{-1}(\mathbf{X}-\mathbf{M})\right]\right)}{(2\pi)^{np/2}|\boldsymbol{\Sigma}_c|^{n/2}|\boldsymbol{\Sigma}_r|^{p/2}}$$

Outline

Parameter Priors

- Pixel vectors in the lower dimensional subspace: independent conjugate Gaussian priors

$$\mathbf{U}|\bar{\mathbf{U}}, \boldsymbol{\Sigma} \sim \mathcal{MN}\left(\bar{\mathbf{U}}, \boldsymbol{\Sigma}, \mathbf{I}_n\right)$$

- Noise variances: independent conjugate inverse-gamma priors

$$s^2_{\mathrm{H},\ell} \;\&\; s^2_{\mathrm{M},\ell}|\nu, \gamma \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right)$$

Flexible distribution whose shape can be adjusted from (ν, γ)

Assumptions

- $\bar{\mathbf{U}}$: fixed using an interpolated hyperspectral image (obtained using splines) projected onto the subspace
- ν: fixed (disappears later)

Hyperparameter Prior

Hyperparameter vector: $\boldsymbol{\Phi} = \{\boldsymbol{\Sigma}, \gamma\}$

- Hyperparameter $\boldsymbol{\Sigma}$: Inverse-Wishart ($\mathcal{IW}$) distribution

$$\boldsymbol{\Sigma} \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}, \eta)$$

  where $\boldsymbol{\Psi}$ and η are fixed to provide a non-informative prior

- Hyperparameter γ: Jeffreys' non-informative prior

$$f(\gamma) \propto \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma)$$

Joint Posterior

Using Bayes theorem, the joint posterior distribution is

$$f\left(\boldsymbol{\theta}, \boldsymbol{\Phi} | \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}\right) \propto f\left(\mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}} | \boldsymbol{\theta}\right) f\left(\boldsymbol{\theta} | \boldsymbol{\Phi}\right) f\left(\boldsymbol{\Phi}\right)$$

where

- unknown parameters: $\boldsymbol{\theta} = \left\{\mathbf{U}, s_{\mathrm{H}}^2, s_{\mathrm{M}}^2\right\}$
- unknown hyperparameters: $\boldsymbol{\Phi} = \{\boldsymbol{\Sigma}, \gamma\}$

How can we estimate $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$?

- Marginalize the hyperparameter γ
- Sample according to the joint posterior $f\left(\mathbf{U}, \boldsymbol{s}^2, \Sigma | \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}\right)$ by using a block Gibbs sampler, which can be easily implemented since all the conditional distributions associated with $f\left(\mathbf{U}, \boldsymbol{s}^2, \Sigma | \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}\right)$ are simple.

Outline

Block Gibbs sampler[4]

for $t = 1$ to N_{MC} **do**

% Sampling the image covariance matrix

Sample $\Sigma^{(t)}$ from $f(\Sigma|\mathbf{U}^{(t-1)}, \boldsymbol{s}^{2^{(t-1)}}, \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}})$

% Sampling the multispectral noise variances

for $\ell = 1$ to n_λ **do**

Sample $s_{\mathrm{M},\ell}^{2(t)}$ from $f(s_{\mathrm{M},\ell}^{2}|\mathbf{U}, \mathbf{Y}_{\mathrm{M}})$,

end for

% Sampling the hyperspectral noise variances

for $\ell = 1$ to m_λ **do**

Sample $s_{\mathrm{H},\ell}^{2(t)}$ from $f(s_{\mathrm{H},\ell}^{2}|\mathbf{U}, \mathbf{Y}_{\mathrm{H}})$,

end for

% Sampling the high-resolved image

Sample $\mathbf{U}^{(t)}$ using a Hamiltonian Monte Carlo algorithm

end for

[4] Q. Wei *et al.*, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117-1127, Sept. 2015.

Conditional Distributions

- Covariance matrix of the image $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma}|\mathbf{u}, \boldsymbol{s}^2, \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}} \sim$$
$$\mathcal{W}^{-1}\left(\boldsymbol{\Psi} + \sum_{i=1}^{m_{\mathrm{x}} m_{\mathrm{y}}} (\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}}^{(i)})^T (\boldsymbol{u}_i - \boldsymbol{\mu}_{\boldsymbol{u}}^{(i)}), n + \eta\right)$$

- Noise variance vector $\boldsymbol{s}^2$

$$s_{\mathrm{H},\ell}^2|\mathbf{U}, \mathbf{Y}_{\mathrm{H}} \sim \mathcal{IG}\left(\frac{m}{2}, \frac{\left[\|\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS}\|_F^2\right]_\ell}{2}\right)$$
$$s_{\mathrm{M},\ell}^2|\mathbf{U}, \mathbf{Y}_{\mathrm{H}} \sim \mathcal{IG}\left(\frac{n}{2}, \frac{\left[\|\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU}\|_F^2\right]_\ell}{2}\right)$$

Conditional Distributions (Cont.)

- Highly-resolved image **U**

$$-\log f(\mathbf{U}|\Sigma, \boldsymbol{s}^2, \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}) = \frac{1}{2}\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS}\right)\|_F^2 + \frac{1}{2}\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU}\right)\|_F^2 + \frac{1}{2}\|\Sigma^{-\frac{1}{2}}\left(\mathbf{U} - \mu_{\mathbf{U}}\right)\|_F^2 + \mathrm{C}$$

  - Not a matrix normal distribution but a normal distribution in vector form: huge covariance matrix
  - Very difficult to draw samples directly from the conditional distribution w.r.t. **U**
  - A Hamiltonian Monte Carlo method[5] is used to sample this high dimensional Gaussian distribution.
  - Note: Other techniques based on perturbation-optimization strategies[6] might also be used.

[5] R. N. Neal, "MCMC using Hamiltonian dynamics," *Handbook of Markov Chain Monte Carlo*, S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng (editors), Chapman & Hall - CRC Press, pp. 113-162, 2010.

[6] F. Orieux *et al.*. "Sampling high-dimensional Gaussian distributions for general linear inverse problems," *IEEE Signal Process. Lett.*, vol. 19, no. 5, May 2012.

Hamiltonian Monte Carlo Methods

Classical Metropolis-Hastings moves

- Classical proposal: random walk
- Accept/reject procedure

Can be inefficient for sampling large vectors (low acceptance rate and mixing properties)

Deterministic gradient based methods

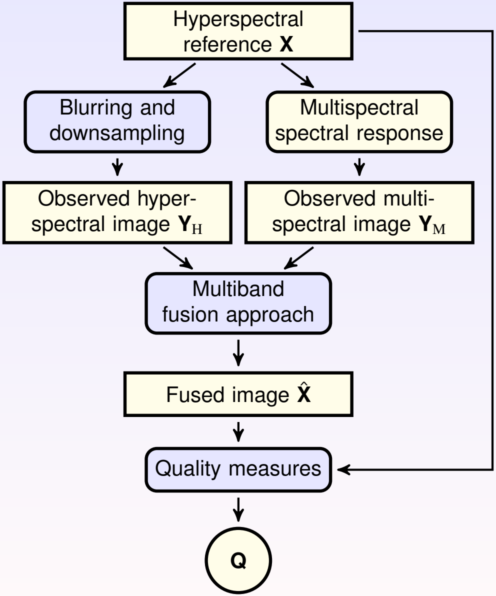
- Well adapted to update vector/matrix elements simultaneously
- Local behavior of a cost function

Hamiltonian Monte Carlo methods

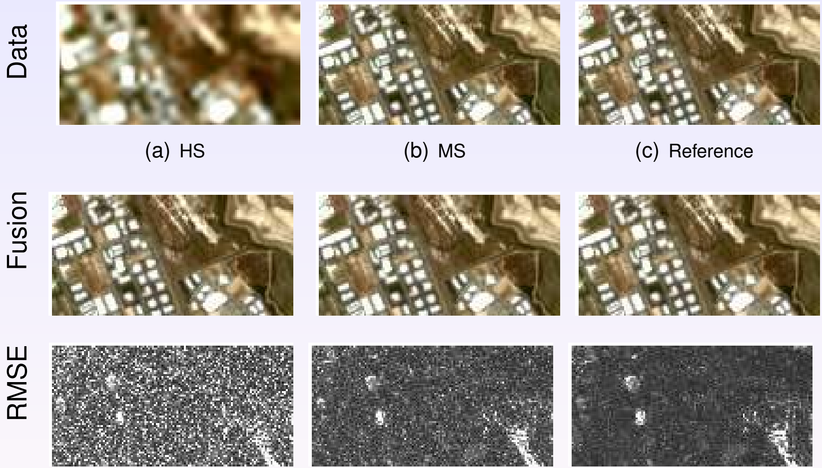
- Consideration of the local curvature of the target density to build an accurate proposal distribution for sampling vector/matrix elements simultaneously

Outline

Wald's Protocol



Qualitative Results (AVIRIS dataset)



(a) HS (b) MS (c) Reference

(d) MAP[Hardie2004] (e) Wavelet MAP[Zhang2012] (f) MCMC

Quantitative Performance Measures

- RMSE/RSNR (Root Mean Square Error): a similarity measure between the target image $\mathbf{X}$ and the fused image $\hat{\mathbf{X}}$

$$\mathrm{RMSE}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{nm_\lambda}\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2$$
$$\mathrm{RSNR}(\mathbf{X}, \hat{\mathbf{X}}) = \log \frac{1}{nm_\lambda}\frac{\|\mathbf{X}\|_F^2}{\mathrm{RMSE}}$$

  The smaller RMSE/larger RSNR, the better the fusion quality.

- SAM (Spectral Angle Mapper): spectral distortion between the actual and estimated images

$$\mathrm{SAM}(\boldsymbol{x}_n, \hat{\boldsymbol{x}}_n) = \arccos\left(\frac{\langle \boldsymbol{x}_n, \hat{\boldsymbol{x}}_n\rangle}{\|\boldsymbol{x}_n\|_2\|\hat{\boldsymbol{x}}_n\|_2}\right)$$

  The overall SAM is obtained by averaging the SAMs computed from all image pixels. The smaller the absolute value of SAM, the less important the spectral distortion.

Quantitative Performance Measures

- UIQI (Universal Image Quality Index): related to the correlation, luminance distortion and contrast distortion of the estimated image w.r.t. the reference image. The UIQI between two images **a** and **â** is

$$\mathrm{UIQI}(\mathbf{a}, \hat{\mathbf{a}}) = \frac{4\sigma^2_{a\hat{a}}\mu_a\mu_{\hat{a}}}{(\sigma^2_a + \sigma^2_{\hat{a}})(\mu^2_a + \mu^2_{\hat{a}})}$$

  where $(\mu_a, \mu_{\hat{a}}, \sigma^2_a, \sigma^2_{\hat{a}})$ are the sample means and variances of a and $\hat{a}$, and $\sigma^2_{a\hat{a}}$ is the sample covariance of $(a, \hat{a})$. The range of UIQI is $[-1, 1]$. The larger UIQI, the better.

- DD (degree of distortion): DD between two images **X** and $\hat{\mathbf{X}}$ is defined as

$$\mathrm{DD}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{nm_\lambda}\|\mathrm{vec}(\mathbf{X}) - \mathrm{vec}(\hat{\mathbf{X}})\|_1.$$

  The smaller DD, the better.

Quantitative Performance Measures

- ERGAS The relative dimensionless global error in synthesis (ERGAS) calculates the amount of spectral distortion in the image. This measure of fusion quality is defined as

$$\mathrm{ERGAS} = 100 \times \frac{1}{d^2} \sqrt{\frac{1}{m_\lambda} \sum_{i=1}^{m_\lambda} \left(\frac{\mathrm{RMSE}(i)}{\mu_i} \right)}$$

where $1/d^2$ is the ratio between the pixel sizes of the MS and HS images, μ_i is the mean of the ith band of the HS image, and m_λ is the number of HS bands. The smaller ERGAS, the smaller the spectral distortion.

Quantitative Results (AVIRIS dataset)

Table: Performance of HS+MS fusion methods in terms of: RSNR (db), UIQI, SAM (deg), ERGAS and DD($\times 10^{-2}$) (AVIRIS dataset).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time(s)
MAP[7]	23.33	0.9913	5.05	4.21	4.87	**1.6**
Wavelet[8]	25.53	0.9956	3.98	3.95	3.89	31
MCMC	**26.74**	**0.9966**	**3.40**	**3.77**	**3.33**	530

Advantages

- Samples generated by the proposed method can be used to compute uncertainties about the estimates (confidence intervals)
- Generalization to more complex problems (non-Gaussianities, nonlinearity, etc)
- Noise variance estimation

[7] Hardie *et al.*, "Application of the Stochastic Mixing Model to Hyperspectral Resolution Enhancement," *IEEE Trans. Image Process.*, vol. 13, no. 9, Sept. 2004.

[8] Zhang *et al.*, "Noise-Resistant Wavelet-Based Bayesian Fusion of Multispectral and Hyperspectral Images," *IEEE Trans. Geosci. and Remote Sens.*, vol. 47, no. 11, Nov. 2009.

Noise Variance Estimation

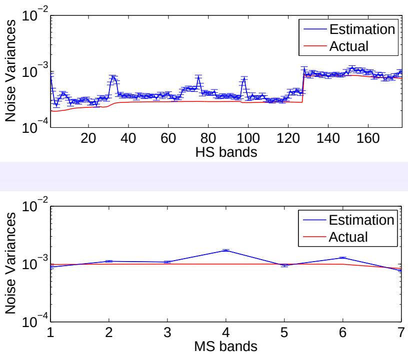


Figure: Noise variances and their MMSE estimates. (Top) HS image. (Bottom) MS image.

- Good estimation performance
- Track noise variance variations with good performance

Outline

Accelerating With Optimization

The negative logarithm of the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\boldsymbol{\mathcal{Y}})$ is given as

$$
\begin{aligned}
&L(\mathbf{U}, \boldsymbol{s}^2, \boldsymbol{\Sigma}) \\
&= -\log p(\boldsymbol{\theta}, \boldsymbol{\Sigma}|\boldsymbol{\mathcal{Y}}) \\
&= -\log p(\mathbf{Y}_{\mathrm{H}}|\boldsymbol{\theta}) - \log p(\mathbf{Y}_{\mathrm{M}}|\boldsymbol{\theta}) - \sum_{l=1}^{n} \log p(\boldsymbol{u}_l|\boldsymbol{\Sigma}) \\
&- \sum_{i=1}^{m_\lambda} \log p\left(s_{\mathrm{H},i}^2\right) - \sum_{j=1}^{n_\lambda} \log p\left(s_{\mathrm{M},j}^2\right) - \log p(\boldsymbol{\Sigma}) - C
\end{aligned}
$$

- MAP estimator: minimizing the function $L(\mathbf{U}, \boldsymbol{s}^2, \boldsymbol{\Sigma})$ with respect to $\mathbf{U}$, $\boldsymbol{s}^2$ and $\boldsymbol{\Sigma}$ iteratively
- use a Block coordinated descent (BCD) algorithm [9]

[9] Q. Wei *et al.*, "Bayesian fusion of multispectral and hyperspectral images using a block coordinate descent Method", in *IEEE GRSS Workshop on Hyperspectral Image and SIgnal Processing: Evolution in Remote Sensing (WHISPERS)*, Tokyo, Japan, Jun. 2015.

Block Coordinated Descent for HS and MS Image Fusion

Input: $\mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}, \widetilde{m}_{\lambda}, \mathbf{B}, \mathbf{S}, \mathbf{R}, \boldsymbol{s}_0^2, \Sigma_0$

- $\hat{\mathbf{H}} \leftarrow \mathrm{PCA}(\mathbf{Y}_{\mathrm{H}}, \widetilde{m}_{\lambda})$; `/* Subspace transform matrix */`

for $t = 1, 2, \ldots$ **to** *stopping rule* **do**

$\mathbf{U}_t = \arg\min_{\mathbf{U}} L(\mathbf{U}, \boldsymbol{s}_{t-1}^2, \Sigma_{t-1})$; `/* Optimize w.r.t.` $\mathbf{U}$ `*/`

$\boldsymbol{s}_t^2 = \arg\min_{\boldsymbol{s}^2} L(\mathbf{U}_t, \boldsymbol{s}^2, \Sigma_{t-1})$; `/* Optimize w.r.t.` $\boldsymbol{s}^2$ `*/`

$\Sigma_t = \arg\min_{\boldsymbol{\Sigma}} L(\mathbf{U}_t, \boldsymbol{s}_t^2, \Sigma)$; `/* Optimize w.r.t.` $\boldsymbol{\Sigma}$ `*/`

end

Output: $\hat{\mathbf{X}} = \hat{\mathbf{H}}\hat{\mathbf{U}}$ (High resolution HS image)

Remarks

The convergence is guaranteed[10].

[10] D. P. Bertsekas. Nonlinear Programming. Athena Scientific Belmont, 1999.

Minimization w.r.t. U

Using the linear model, dimensionality reduction, fusing the HS and MS images can be formulated as finding **U** minimizing the cost function

$$\begin{aligned} L_{\mathbf{U}}(\mathbf{U}) = \quad & \frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS}\right)\|_F^2 + \frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU}\right)\|_F^2 \\ & + \frac{1}{2}\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\mathbf{U} - \boldsymbol{\mu}_{\mathbf{U}}\right)\|_F^2. \end{aligned}$$

- First two terms: data fidelity terms for the HS+MS images (likelihoods)
- Last term: penalty ensuring appropriate regularization (prior)

Difficulties

- Large dimensionality of **U**
- Diagonalization of the linear operators $\mathbf{H}(\cdot)\mathbf{BS}$ not possible

Alternating Direction Method of Multipliers (ADMM)

Idea: transform the unconstrained optimization with respect to **U** into a constrained one via a variable splitting "trick", and then attack this constrained problem using an augmented Lagrangian (AL) method[11]

- Splittings: $\mathbf{H}_1 = \mathbf{UB}$, $\mathbf{H}_2 = \mathbf{U}$ and $\mathbf{H}_3 = \mathbf{U}$
- Respective scaled dual variable: $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$

$$
\begin{aligned}
& L(\mathbf{U}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) \qquad \boxed{\text{Deconvolution}} \\
= \quad & \frac{1}{2}\left\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{H}_1\mathbf{S})\right\|_F^2 + \boxed{\frac{\mu}{2}\left\|\mathbf{UB} - \mathbf{H}_1 - \mathbf{G}_1\right\|_F^2} + \\
& \frac{1}{2}\left\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{R}\mathbf{H}\mathbf{H}_2)\right\|_F^2 + \boxed{\frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_2 - \mathbf{G}_2\right\|_F^2} + \\
& \frac{1}{2}\left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\mu}_{\mathbf{U}} - \mathbf{H}_3\right)\right\|_F^2 + \boxed{\frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_3 - \mathbf{G}_3\right\|_F^2}
\end{aligned}
$$

[11] M. Afonso *et al.*, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3,pp. 681-695, 2011.

Alternating Direction Method of Multipliers (ADMM)

Idea: transform the unconstrained optimization with respect to **U** into a constrained one via a variable splitting "trick", and then attack this constrained problem using an augmented Lagrangian (AL) method

- Splittings: $\mathbf{H}_1 = \mathbf{UB}$, $\mathbf{H}_2 = \mathbf{U}$ and $\mathbf{H}_3 = \mathbf{U}$
- Respective scaled dual variable: $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$

$$L(\mathbf{U}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) \qquad \boxed{\textit{Upsampling}}$$

$$= \boxed{\frac{1}{2}\left\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{H}_1\mathbf{S})\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{UB} - \mathbf{H}_1 - \mathbf{G}_1\right\|_F^2} +$$
$$\frac{1}{2}\left\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{RH}\mathbf{H}_2)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_2 - \mathbf{G}_2\right\|_F^2 +$$
$$\frac{1}{2}\left\|\Sigma^{-\frac{1}{2}}(\mu_{\mathbf{U}} - \mathbf{H}_3)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_3 - \mathbf{G}_3\right\|_F^2$$

Alternating Direction Method of Multipliers (ADMM)
Idea: transform the unconstrained optimization with respect to **U** into a constrained one via a variable splitting “trick”, and then attack this constrained problem using an augmented Lagrangian (AL) method

- Splittings: $\mathbf{H}_1 = \mathbf{UB}$, $\mathbf{H}_2 = \mathbf{U}$ and $\mathbf{H}_3 = \mathbf{U}$
- Respective scaled dual variable: $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$

Spectral Unmixing

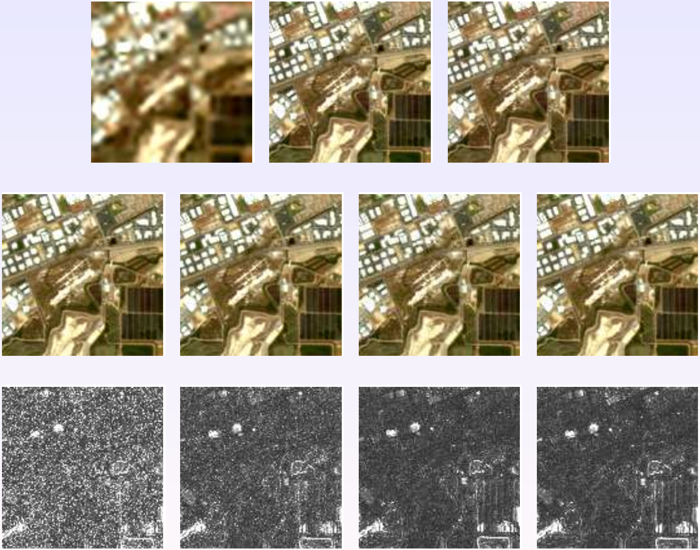
$$
\begin{aligned}
& L(\mathbf{U}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) \\
= \quad & \frac{1}{2}\left\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{H}_1\mathbf{S})\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{UB} - \mathbf{H}_1 - \mathbf{G}_1\right\|_F^2 + \\
& \boxed{\frac{1}{2}\left\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{R}\mathbf{H}\mathbf{H}_2)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_2 - \mathbf{G}_2\right\|_F^2} + \\
& \frac{1}{2}\left\|\Sigma^{-\frac{1}{2}}\left(\boldsymbol{\mu}_{\mathbf{U}} - \mathbf{H}_3\right)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_3 - \mathbf{G}_3\right\|_F^2
\end{aligned}
$$

Alternating Direction Method of Multipliers (ADMM)
Idea: transform the unconstrained optimization with respect to **U** into a constrained one via a variable splitting "trick", and then attack this constrained problem using an augmented Lagrangian (AL) method

- Splittings: $\mathbf{H}_1 = \mathbf{UB}$, $\mathbf{H}_2 = \mathbf{U}$ and $\mathbf{H}_3 = \mathbf{U}$
- Respective scaled dual variable: $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$

$$L(\mathbf{U}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3, \mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3) \qquad \boxed{\textit{Denoising}}$$

$$\begin{aligned} = \quad & \frac{1}{2}\left\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{H}\mathbf{H}_1\mathbf{S})\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{UB} - \mathbf{H}_1 - \mathbf{G}_1\right\|_F^2 + \\ & \frac{1}{2}\left\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}} - \mathbf{R}\mathbf{H}\mathbf{H}_2)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_2 - \mathbf{G}_2\right\|_F^2 + \\ & \boxed{\frac{1}{2}\left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\left(\boldsymbol{\mu}_{\mathbf{U}} - \mathbf{H}_3\right)\right\|_F^2 + \frac{\mu}{2}\left\|\mathbf{U} - \mathbf{H}_3 - \mathbf{G}_3\right\|_F^2} \end{aligned}$$



(a) MAP (b) Wavelet MAP (c) MCMC (d) BCD

Table: Performance of the compared fusion methods: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-2}) and time (in second) (AVIRIS dataset).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
MAP	23.14	0.9932	5.147	3.524	4.958	3
Wavelet MAP	24.91	0.9956	4.225	3.282	4.120	72
MCMC	25.92	0.9971	3.733	2.926	3.596	6228
BCD	25.85	0.9970	3.738	2.946	3.600	96

- Promising results for the considered quality measures
- Significant reduction in computation time: Save a lot of time!

Summary

State-of-the-Art

Revisited Bayesian Fusion

Sparse Prior Based on Dictionary Learning
- Sparse Regularization
- Alternate Optimization Scheme

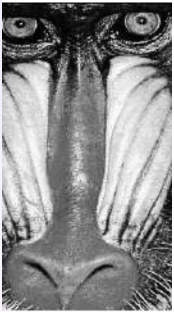
Fast fUsion Based on a Sylvester Equation (FUSE)

Conclusions

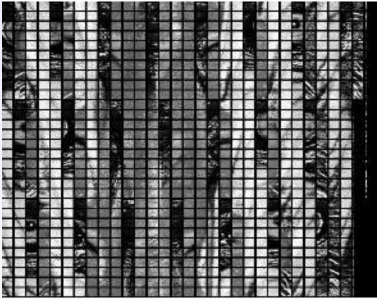
Outline

Sparse Regularization

Motivation
Self-similarity property of natural image patches



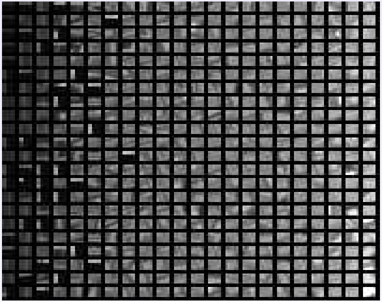
image



patches

Remote Sensing Images

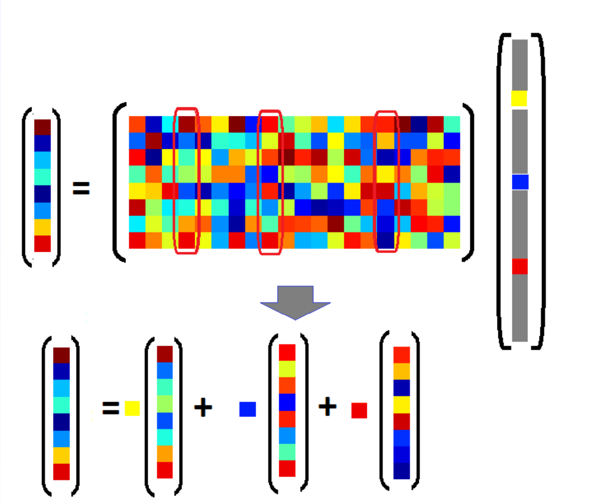
image



patches

Sparse Regularization

The patches of the target image **U** can be sparsely approximated on an over-complete dictionary (with columns referred to as atoms).



Penalized Inverse Problem

Based on the previous observation models and dimensionality reduction, fusing the HS and MS images can be formulated as the following inverse problem

$$\min_{\mathbf{U}} \underbrace{\frac{1}{2}\left\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}}-\mathbf{HUBS})\right\|_F^2}_{\substack{\text{HS data term} \\ \propto \ln p(\mathbf{Y}_{\mathrm{H}}|\mathbf{U})}} + \underbrace{\frac{1}{2}\left\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}}-\mathbf{RHU})\right\|_F^2}_{\substack{\text{MS data term} \\ \propto \ln p(\mathbf{Y}_{\mathrm{M}}|\mathbf{U})}} + \underbrace{\lambda\phi(\mathbf{U})}_{\substack{\text{regularizer} \\ \propto \ln p(\mathbf{U})}},$$

Sparse Regularization

Regularizer

$$\phi(\mathbf{U}) = \frac{1}{2}\left\|\mathbf{U} - \bar{\mathbf{U}}\left(\mathbf{D}, \mathbf{A}\right)\right\|_F^2$$

Separating each band of the target image leads to

$$\phi(\mathbf{U}) = \frac{1}{2}\sum_{i=1}^{\widetilde{m}_\lambda}\left\|\mathbf{U}_i - \mathcal{P}\left(\mathbf{D}_i\mathbf{A}_i\right)\right\|_F^2$$

- $\mathbf{U}_i \in \mathbb{R}^n$ is the ith band (or row) of $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$
- $\mathbf{D}_i \in \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{at}}$ is the dictionary dedicated to the ith band of $\mathbf{U}$ (n_p is the patch size and n_at is the number of atoms) and $\mathbf{D} = \left[\mathbf{D}_1, \cdots, \mathbf{D}_{\widetilde{m}_\lambda}\right]$
- $\mathbf{A}_i \in \mathbb{R}^{n_\mathrm{at} \times n_\mathrm{pat}}$ is the ith band's code (n_pat is the number of patches associated with the ith band) and $\mathbf{A} = \left[\mathbf{A}_1, \cdots, \mathbf{A}_{\widetilde{m}_\lambda}\right]$
- $\mathcal{P}(\cdot)$ is a linear operator that averages the overlapping patches of each band to restore the target image

How can we obtain the dictionary **D** and the code **A**?

Dictionary Learning and Sparse Coding

Dictionary Learning

Learn[12] the set of over-complete dictionaries $\mathbf{D} = \left[\mathbf{D}_1, \cdots, \mathbf{D}_{\widetilde{m}_\lambda}\right]$: applying a DL algorithm on the rough estimation of $\mathbf{U}$ (constructed from the MS and HS images)

- K-SVD method
- Online Dictionary Learning (ODL) method

Sparse Coding

- Orthogonal Matching Pursuit (OMP): to estimate the sparse code $\mathbf{A}_i$ (with $n_{\max}$ coefficients) for each band $\mathbf{U}_i$
- Support $(\boldsymbol{\Omega}_i \subset \mathbb{N}^2, i = 1, \cdots, \widetilde{m}_\lambda)$: The positions of the non-zero elements of the code $\mathbf{A}_i$ are also identified

[12] M. Elad *et al.*, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

Re-estimation of the Sparse Code

Inspired by hierarchical models frequently encountered in Bayesian inference, we propose to include the code $\mathbf{A}$ within the estimation process.

$$\phi(\mathbf{U}, \mathbf{A}) = \frac{1}{2} \sum_{i=1}^{\widetilde{m}_\lambda} \left\| \mathbf{U}_i - \mathcal{P}\left(\mathbf{D}_i \mathbf{A}_i\right) \right\|_F^2 + \mu_a \left\| \mathbf{A}_i \right\|_0 \quad \text{NP hard!}$$

where $\|.\|_0$ is the ℓ_0 counting function (or ℓ_0 norm) and μ_a is a regularization parameter.

By fixing the supports $\boldsymbol{\Omega}_i$, the ℓ_0 norm reduces to a constant. Hence,

$$\phi(\mathbf{U}, \mathbf{A}) = \frac{1}{2} \sum_{i=1}^{\widetilde{m}_\lambda} \left\| \mathbf{U}_i - \mathcal{P}\left(\mathbf{D}_i \mathbf{A}_i\right) \right\|_F^2 \text{ s.t. } \mathbf{A}_{i,\setminus \boldsymbol{\Omega}_i} = 0$$

where $\mathbf{A}_{i,\setminus \boldsymbol{\Omega}_i} = \{\mathbf{A}_i(l,k) \mid (l,k) \notin \boldsymbol{\Omega}_i\}$.

Final Optimization Problem

Joint optimization with respect to **U** and **A**

$$\begin{aligned}\min_{\mathbf{U},\mathbf{A}} L(\mathbf{U},\mathbf{A}) \quad &= \frac{1}{2}\left\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS})\right\|_F^2 + \frac{1}{2}\left\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU}\right\|_F^2 + \\ &\frac{\lambda}{2}\sum_{i=1}^{\widetilde{m}_\lambda}\left(\left\|\mathbf{U}_i - \mathcal{P}\left(\mathbf{D}_i\mathbf{A}_i\right)\right\|_F^2\right), \text{ s.t. } \mathbf{A}_{i,\setminus\boldsymbol{\Omega}_i} = 0\end{aligned}$$

Solution

- Solved by minimizing w.r.t. **U** and **A** alternatively
- Each sub-problem is strictly convex

Outline

Optimization with Respect to U

$$\min_{\mathbf{U}} L(\mathbf{U}) = \frac{1}{2}\left\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS}\right)\right\|_F^2 + \frac{1}{2}\left\|\boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU}\right)\right\|_F^2 + \frac{\lambda}{2}\sum_{i=1}^{\widetilde{m}_\lambda}\left(\left\|\mathbf{U}_i - \mathcal{P}\left(\mathbf{D}_i\mathbf{A}_i\right)\right\|_F^2\right),$$

Difficulties

- Large dimensionality of $\mathbf{U}$
- Diagonalization of the linear operators $\mathbf{H}(\cdot)\mathbf{BS}$ and $\mathcal{P}(\cdot)$ not possible

Solution

Alternating Direction Method of Multipliers (ADMM)

Optimization with Respect to A

Optimization with respect to $\mathbf{A}_i$ ($i = 1, \cdots, \widetilde{m}_\lambda$) conditional on $\mathbf{U}_i$

$$\min_{\mathbf{A}_i} \left\| \mathbf{U}_i - \mathcal{P}(\mathbf{D}_i \mathbf{A}_i) \right\|_F^2 \text{ s.t. } \mathbf{A}_{i,\setminus \boldsymbol{\Omega}_i} = 0$$

Remarks

- The optimization with respect to $\mathbf{A}_i$ considers only the non-zero elements of $\mathbf{A}_i$, denoted as $\mathbf{A}_{i,\boldsymbol{\Omega}_i} = \{\mathbf{A}_i(l,k) \mid (l,k) \in \boldsymbol{\Omega}_i\}$
- Standard least square (LS) problem which can be solved analytically

Alternate Optimization Scheme[13]

Input: $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{Y}_{\mathrm{M}}$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{R}$, $\mathrm{SNR}_{\mathrm{H}}$, $\mathrm{SNR}_{\mathrm{M}}$, $\widetilde{m}_{\lambda}$, $n_{\max}$

- Approximate $\bar{\mathbf{U}}$ using $\mathbf{Y}_{\mathrm{M}}$ and $\mathbf{Y}_{\mathrm{H}}$ /* *Rough estimation of* $\mathbf{U}$ */
- $\hat{\mathbf{D}} \leftarrow \mathrm{ODL}(\bar{\mathbf{U}})$ /* *Online dictionary learning* */
- $\hat{\mathbf{A}} \leftarrow \mathrm{OMP}(\hat{\mathbf{D}}, \bar{\mathbf{U}}, n_{\max})$ /* *Sparse coding* */
- $\hat{\Omega} \leftarrow \hat{\mathbf{A}} \neq 0$ /* *Computing support* */
- $\hat{\mathbf{H}} \leftarrow \mathrm{PCA}(\mathbf{Y}_{\mathrm{H}}, \widetilde{m}_{\lambda})$ /* *Computing subspace transform matrix* */

/* *Start alternate optimization* */
for $t = 1, 2, \ldots$ **to** *stopping rule* **do**
  $\hat{\mathbf{U}}_t \in \{\mathbf{U} : L(\mathbf{U}, \hat{\mathbf{A}}_{t-1}) \leq L(\hat{\mathbf{U}}_{t-1}, \hat{\mathbf{A}}_{t-1})\}$ /* *solved with ADMM* */
  $\hat{\mathbf{A}}_t \in \{\mathbf{A} : L(\hat{\mathbf{U}}_t, \mathbf{A}) \leq L(\hat{\mathbf{U}}_t, \hat{\mathbf{A}}_{t-1})\}$ /* *solved with LS* */
end
$\hat{\mathbf{X}} = \hat{\mathbf{H}}\hat{\mathbf{U}}$
Output: $\hat{\mathbf{X}}$ (high resolution HS image)

[13] Q. Wei *et al.*, "Hyperspectral and multispectral image fusion based on a sparse representation", *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 7, pp. 3658-3668, July 2015.

Image Characteristics

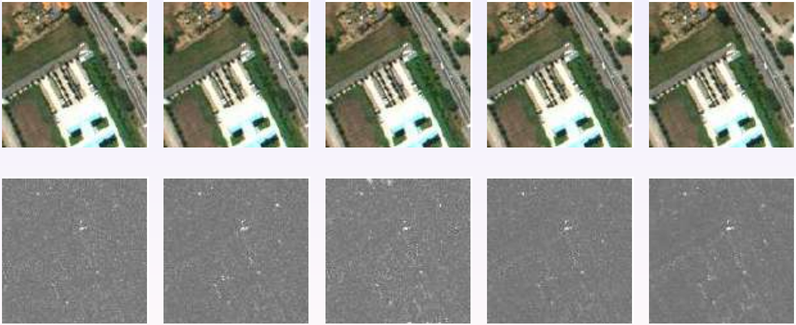
TABLE 2. CHARACTERISTIC OF THE THREE DATASETS.

DATASET	DIMENSIONS	SPATIAL RES	N	INSTRUMENT
Moffett	PAN 185 × 395 HS 37 × 79	20m 100m	224	AVIRIS
Camargue	PAN 500 × 500 HS 100 × 100	4m 20m	125	HyMap
Garons	PAN 400 × 400 HS 80 × 80	4m 20m	125	HyMap

Data Characteristics [Loncan2015]

Qualitative Results (Pavia Dataset)

(a) Ref (b) HS (c) MS



(d) MAP (e) Wavelet (f) CNMF (g) Gaussian (h) Sparse

Quantitative Results (Pavia Dataset)

Table: Performance of different MS + HS fusion methods (Pavia dataset): RMSE (in 10^{-2}), UIQI, SAM (in degree), ERGAS, DD (in 10^{-3}) and Time (in second).

Methods	RMSE	UIQI	SAM	ERGAS	DD	Time
MAP	1.148	0.9875	1.962	1.029	8.666	3
Wavelet MAP	1.099	0.9885	1.849	0.994	8.349	75
CNMF	1.119	0.9857	2.039	1.089	9.007	14
Gaussian	1.011	0.9903	1.653	0.911	7.598	6003
Sparse	0.947	0.9913	1.492	0.850	7.010	282

The proposed method provides promising results for the considered quality measures.

Summary

State-of-the-Art

Revisited Bayesian Fusion

Sparse Prior Based on Dictionary Learning

Fast fUsion Based on a Sylvester Equation (FUSE)
- From Maximum Likelihood...
- ... to a Maximum a Posteriori Estimator
- Comparison with State-of-The-Art Fusion Methods (HS Pansharpening)

Conclusions

Outline

State-of-the-Art

Revisited Bayesian Fusion

Sparse Prior Based on Dictionary Learning

Fast fUsion Based on a Sylvester Equation (FUSE)
- From Maximum Likelihood...
- ... to a Maximum a Posteriori Estimator
- Comparison with State-of-The-Art Fusion Methods (HS Pansharpening)

Conclusions

Transforming Optimization to Solving a Sylvester Equation

Forward model

$$\mathbf{Y}_{\mathrm{H}} = \mathbf{XBS} + \mathbf{N}_{\mathrm{H}}, \quad \mathbf{Y}_{\mathrm{M}} = \mathbf{RX} + \mathbf{N}_{\mathrm{M}}$$

$$\text{s.t. } \mathbf{X} = \mathbf{HU}$$

Negative log-likelihood (in subspace)

$$\begin{aligned} -\log p\left(\mathcal{Y}|\mathbf{U}\right) &= -\log p\left(\mathbf{Y}_{\mathrm{H}}|\mathbf{U}\right) - \log p\left(\mathbf{Y}_{\mathrm{M}}|\mathbf{U}\right) + C \\ &= \tfrac{1}{2}\|\Lambda_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS}\right)\|^2 + \tfrac{1}{2}\|\Lambda_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU}\right)\|^2 + C \end{aligned}$$

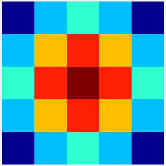
Minimizing the likelihood w.r.t. $\mathbf{U}$ $\Leftrightarrow$ solve a generalized Sylvester matrix equation

$$\boxed{\left[\mathbf{H}^H \Lambda_{\mathrm{H}}^{-1}\mathbf{H}\right]\mathbf{U}\left[\mathbf{BS}\left(\mathbf{BS}\right)^H\right] + \left[\left(\mathbf{RH}\right)^H \Lambda_{\mathrm{M}}^{-1}\left(\mathbf{RH}\right)\right]\mathbf{U} = \text{term ind. on } \mathbf{U}}$$

- $\mathbf{BS}\left(\mathbf{BS}\right)^H$ is not diagonalizable!

Assumption 1

The blurring matrix $\mathbf{B}$ *is a block circulant matrix with circulant blocks (BCCB).*



Assumption 2

The decimation matrix $\mathbf{S}$ *corresponds to downsampling the original signal and its conjugate transpose* $\mathbf{S}^H$ *interpolates the decimated signal with zeros, e.g.,*

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

These two assumptions are used to compute an explicit solution of the Sylvester equation.

$$\left[\mathbf{H}^H \Lambda_{\mathrm{H}}^{-1} \mathbf{H}\right] \mathbf{U} \left[\mathbf{BS}\left(\mathbf{BS}\right)^H\right] + \left[\left(\mathbf{RH}\right)^H \Lambda_{\mathrm{M}}^{-1} \left(\mathbf{RH}\right)\right] \mathbf{U} = \text{term ind. on } \mathbf{U}$$

3 *Main Steps*

- Left multiply by $\left(\mathbf{H}^H \Lambda_{\mathrm{H}}^{-1} \mathbf{H}\right)^{-1}$: $\mathbf{UC_2} + \mathbf{C_1 U} = \mathbf{C_3}$, where $\mathbf{C_2} = \mathbf{BS}\left(\mathbf{BS}\right)^{\mathbf{H}}$.

> ***Lemma 1***
>
> *The equality $\mathbf{F}^H \underline{\mathbf{S}}\mathbf{F} = \frac{1}{d}\mathbf{J}_d \otimes \boldsymbol{I}_m$ holds, where $\mathbf{F}$ is the DFT matrix, $\underline{\mathbf{S}} = \mathbf{SS}^H$, $\mathbf{J}_d$ is the $d \times d$ matrix of ones and $\boldsymbol{I}_m$ is the $m \times m$ identity matrix.*

- Diagonalize $\mathbf{C_1}$ and use Lemma 1 to simplify $\mathbf{C}_2$:

$$\bar{\mathbf{U}}\mathbf{M} + \Lambda_C \bar{\mathbf{U}} = \bar{\mathbf{C}}_3$$

with a diagonal matrix Λ_C and $\mathbf{M} = \frac{1}{d}\begin{bmatrix} \sum_{i=1}^{d} \underline{\mathbf{D}}_i & \underline{\mathbf{D}}_2 & \cdots & \underline{\mathbf{D}}_d \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}$

$\underline{\mathbf{D}}_i$: $m \times m$ diagonal matrix, d: downsampling ratio, m: number of image pixels

Theorem 2

[a] *Let* $(\bar{\mathbf{C}}_3)_{l,j}$ *denotes the* j*th* *block* *of the* l*th* *band* *of* $\bar{\mathbf{C}}_3$ *for any* $l = 1, \cdots, \widetilde{m}_\lambda$. *Then, the solution* $\bar{\mathbf{U}}$ *of the SE can be decomposed as*

$$\bar{\mathbf{U}} = \begin{bmatrix} \bar{\mathbf{u}}_{1,1} & \bar{\mathbf{u}}_{1,2} & \cdots & \bar{\mathbf{u}}_{1,d} \\ \bar{\mathbf{u}}_{2,1} & \bar{\mathbf{u}}_{2,2} & \cdots & \bar{\mathbf{u}}_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathbf{u}}_{\widetilde{m}_\lambda,1} & \bar{\mathbf{u}}_{\widetilde{m}_\lambda,2} & \cdots & \bar{\mathbf{u}}_{\widetilde{m}_\lambda,d} \end{bmatrix}$$

with

$$\bar{\mathbf{u}}_{l,j} = \begin{cases} (\bar{\mathbf{C}}_3)_{l,j} \left(\frac{1}{d} \sum_{i=1}^{d} \underline{\mathbf{D}}_i + \lambda_C^l \boldsymbol{I}_m \right)^{-1}, & j = 1, \\ \frac{1}{\lambda_C^l} \left[(\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d} \bar{\mathbf{u}}_{l,1} \underline{\mathbf{D}}_j \right], & j = 2, \cdots, d. \end{cases}$$

[a] Q. Wei *et al.*, “Fast multi-band image fusion based on solving a Sylvester equation”, *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109-4121, Nov. 2015.

Fast FUsion Based on a Sylvester Equation (FUSE)

Input: $\mathbf{Y}_\mathrm{M}, \mathbf{Y}_\mathrm{H}, \boldsymbol{\Lambda}_\mathrm{M}, \boldsymbol{\Lambda}_\mathrm{H}, \mathbf{R}, \mathbf{B}, \mathbf{S}, \mathbf{H}$

- $\mathbf{D} \leftarrow \mathrm{Dec}\left(\mathbf{B}\right)$ and $\underline{\mathbf{D}} = \mathbf{D}^*\mathbf{D}$ /*Circulant matrix: $\mathbf{B} = \mathbf{F}\mathbf{D}\mathbf{F}^H$ */
- $\mathbf{C}_1 \leftarrow \left(\mathbf{H}^H \boldsymbol{\Lambda}_\mathrm{H}^{-1}\mathbf{H}\right)^{-1}\left((\mathbf{R}\mathbf{H})^H \boldsymbol{\Lambda}_\mathrm{L}^{-1}\mathbf{R}\mathbf{H}\right)$;
- $(\mathbf{Q}, \boldsymbol{\Lambda}_C) \leftarrow \mathrm{EigDec}\left(\mathbf{C}_1\right)$ /* Eigen-dec of $\mathbf{C}_1$: $\mathbf{C}_1 = \mathbf{Q}\boldsymbol{\Lambda}_C\mathbf{Q}^{-1}$ */
- $\bar{\mathbf{C}}_3 \leftarrow$
  $\mathbf{Q}^{-1}\left(\mathbf{H}^H \boldsymbol{\Lambda}_\mathrm{H}^{-1}\mathbf{H}\right)^{-1}\left(\mathbf{H}^H \boldsymbol{\Lambda}_\mathrm{H}^{-1}\mathbf{Y}_\mathrm{H}(\mathbf{B}\mathbf{S})^H + (\mathbf{R}\mathbf{H})^H \boldsymbol{\Lambda}_\mathrm{L}^{-1}\mathbf{Y}_\mathrm{M}\right)\mathbf{B}\mathbf{F}\mathbf{P}^{-1}$;

for $l = 1$ **to** $\widetilde{m}_\lambda$ **do**

$$\bar{\mathbf{u}}_{l,1} = (\bar{\mathbf{C}}_3)_{l,1}\left(\frac{1}{d}\sum_{i=1}^{d}\underline{\mathbf{D}}_i + \lambda_C^l \mathbf{I}_n\right)^{-1};$$

  for $j = 2$ **to** d **do**

$$\bar{\mathbf{u}}_{l,j} = \frac{1}{\lambda_C^l}\left((\bar{\mathbf{C}}_3)_{l,j} - \frac{1}{d}\bar{\mathbf{u}}_{l,1}\underline{\mathbf{D}}_j\right);$$

  end

end

Output: $\mathbf{X} = \mathbf{H}\mathbf{Q}\bar{\mathbf{U}}\mathbf{P}\mathbf{D}^{-1}\mathbf{F}^H$

Outline

From ML to MAP Estimators

Generalization to Bayesian estimators[14]

- $\phi(\mathbf{X})$: Gaussian prior based on interpolation[15]
- $\phi(\mathbf{X})$: Sparse representation based on dictionary learning[16]
- $\phi(\mathbf{X})$: Total variation (TV)[17]

---

[14] Q. Wei *et al.*, "Fast multi-band image fusion based on solving a Sylvester equation", *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109-4121, Nov. 2015.

[15] Q. Wei *et al.*, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117-1127, Sept. 2015.

[16] Q. Wei *et al.*, "Hyperspectral and multispectral image fusion based on a sparse representation", *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 7, pp. 3658-3668, July 2015.

[17] M. Simões *et al.*, "A convex formulation for hyperspectral image superresolution via subspace-based regularization", *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 6, pp. 3373-3388, June 2015.

Gaussian Prior

Gaussian prior: Sylvester equation embedded in BCD (FUSE-BCD)

Input: $\mathbf{Y}_{\mathrm{H}}$, $\mathbf{Y}_{\mathrm{M}}$, $\widetilde{m}_{\lambda}$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{R}$, s_0^2, Σ_0

- $\hat{\mathbf{H}} \leftarrow \mathrm{PCA}(\mathbf{Y}_{\mathrm{H}}, \widetilde{m}_{\lambda})$; `/* Subspace transform matrix */`

for $t = 1, 2, \ldots$ **to** *stopping rule* **do**

$\mathbf{U}_t = \arg\min_{\mathbf{U}} L(\mathbf{U}, s_{t-1}^2, \Sigma_{t-1})$; `/* Sylvester equation */`

$s_t^2 = \arg\min_{s^2} L(\mathbf{U}_t, s^2, \Sigma_{t-1})$; `/* Optimize w.r.t.` s^2 `*/`

$\Sigma_t = \arg\min_{\Sigma} L(\mathbf{U}_t, s_t^2, \Sigma)$; `/* Optimize w.r.t.` Σ `*/`

end

Output: $\hat{\mathbf{X}} = \hat{\mathbf{H}}\hat{\mathbf{U}}$ (High resolution HS image)

Sparse Representation

Sparse prior: Sylvester equation embedded in BCD (FUSE-BCD)

Input: $\mathbf{Y}_\mathrm{H}$, $\mathbf{Y}_\mathrm{M}$, $\mathbf{B}$, $\mathbf{S}$, $\mathbf{R}$, SNR_H, SNR_M, $\widetilde{m}_\lambda$, n_max
Output: $\hat{\mathbf{X}}$ (high resolution HS image)

- Approximate $\bar{\mathbf{U}}$ using $\mathbf{Y}_\mathrm{M}$ and $\mathbf{Y}_\mathrm{H}$ /* *Rough estimation of* $\mathbf{U}$ */
- $\hat{\mathbf{D}} \leftarrow \mathrm{ODL}(\bar{\mathbf{U}})$ /* *Online dictionary learning* */
- $\hat{\mathbf{A}} \leftarrow \mathrm{OMP}(\hat{\mathbf{D}}, \bar{\mathbf{U}}, n_\mathrm{max})$ /* *Sparse coding* */
- $\hat{\Omega} \leftarrow \hat{\mathbf{A}} \neq 0$ /* *Computing support* */
- $\hat{\mathbf{H}} \leftarrow \mathrm{PCA}(\mathbf{Y}_\mathrm{H}, \widetilde{m}_\lambda)$ /* *Computing subspace transform matrix* */

/* *Start alternate optimization* */
for $t = 1, 2, \ldots$ **to** *stopping rule* **do**

$\hat{\mathbf{U}}_t \in \{\mathbf{U} : L(\mathbf{U}, \hat{\mathbf{A}}_{t-1}) \leq L(\hat{\mathbf{U}}_{t-1}, \hat{\mathbf{A}}_{t-1})\}$ /* *solved with SE* */

$\hat{\mathbf{A}}_t \in \{\mathbf{A} : L(\hat{\mathbf{U}}_t, \mathbf{A}) \leq L(\hat{\mathbf{U}}_t, \hat{\mathbf{A}}_{t-1})\}$ /* *solved with LS* */

end

$\hat{\mathbf{X}} = \hat{\mathbf{H}}\hat{\mathbf{U}}$

Non-Gaussian Prior

Non-Gaussian prior, such as (TV)[18]

$$\arg\min_{\mathbf{U}} \underbrace{\frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}}-\mathbf{HUBS}\right)\|_F^2}_{\text{HS data term}} + \underbrace{\frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}}-\mathbf{RHU}\right)\|_F^2}_{\text{MS data term}} + \underbrace{\lambda \mathrm{TV}\left(\mathbf{U}\right)}_{\text{regularizer}}.$$

can be equivalently solved as

$$\arg\min_{\mathbf{U},\mathbf{V}} \frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{H}}-\mathbf{HUBS}\right)\|_F^2 + \frac{1}{2}\|\boldsymbol{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{M}}-\mathbf{RHU}\right)\|_F^2 + \lambda \mathrm{TV}\left(\mathbf{V}\right)$$

$$\text{s.t. } \mathbf{U}=\mathbf{V}$$

- ADMM algorithm: Sylvester equation + proximity operator
- Sylvester equation embedded in ADMM (FUSE-ADMM)

[18] M. Simões *et al.*, "A convex formulation for hyperspectral image superresolution via subspace-based regularization", *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 6, pp. 3373-3388, June 2015.

Performance and Computational Times

Table: Performance of HS+MS fusion methods: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-3}) and time (in second).

Regularization	Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
supervised naive Gaussian	ADMM	29.321	0.9906	1.555	0.888	7.115	126.83
	FUSE	29.372	0.9908	1.551	0.879	7.092	0.38
unsupervised naive Gaussian	ADMM-BCD	29.084	0.9902	1.615	0.913	7.341	99.55
	FUSE-BCD	29.077	0.9902	1.623	0.913	7.368	1.09
sparse representation	ADMM-BCD	29.582	0.9911	1.423	0.872	6.678	162.88
	FUSE-BCD	29.688	0.9913	1.431	0.856	6.672	73.66
TV	ADMM	29.473	0.9912	1.503	0.861	6.922	134.21
	FUSE-ADMM	29.631	0.9915	1.477	0.845	6.788	90.99

- The computational time is decreased significantly!

Outline

State-of-the-Art

Revisited Bayesian Fusion

Sparse Prior Based on Dictionary Learning

Fast fUsion Based on a Sylvester Equation (FUSE)
- From Maximum Likelihood...
- ... to a Maximum a Posteriori Estimator
- Comparison with State-of-The-Art Fusion Methods (HS Pansharpening)

Conclusions

Simulation Scenarios

Table: Characteristics of the three datasets[19]

dataset	dimensions	spatial res	N	instrument
Moffett	PAN 185×395 HS 37×79	20m 100m	224	AVIRIS
Camargue	PAN 500×500 HS 100×100	4m 20m	125	HyMap
Garons	PAN 400×400 HS 80×80	4m 20m	125	HyMap

[19] L. Loncan, L. B. Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. Licciardi, M. Simoes, J-Y. Tourneret, M. Veganzones, G. Vivone, Q. Wei and N. Yokoya, "Hyperspectral pansharpening: a review", *IEEE Geosci. and Remote Sens. Mag.*, vol. 3, no. 3, pp. 27-46, Sept. 2015.

Visual Results

Figure: Camargue. (a) Ref, (b) interpolation, (c) SFIM, (d) MTF GLP HPM, (e) GSA, (f) PCA, (g) GFPCA, (h) CNMF, (i) Bayesian Sparse, (j) HySure.

Camargue

Table: Quality measures for the Camargue dataset[20]

method	CC	SAM	RMSE	ERGAS	Time(sec)
SFIM	0.91886	4.2895	637.1451	3.4159	3.47
MTF-GLP	0.92397	4.3378	622.4711	3.2666	4.26
MTF-GLP-HPM	0.92599	4.2821	611.9161	3.2497	4.25
GS	0.91262	4.4982	665.0173	3.5490	8.29
GSA	0.92826	4.1950	587.1322	3.1940	8.73
PCA	0.90350	5.1637	710.3275	3.8680	8.92
GFPCA	0.89042	4.8472	745.6006	4.0001	8.51
CNMF	0.93000	4.4187	591.3190	3.1762	47.54
Supervised Gaussian	0.95195	3.6428	489.5634	2.6286	7.35
Sparse represent.	0.95882	3.3345	448.1721	2.4712	485.13
HySure	0.94650	3.8767	511.8525	2.8181	296.27

[20] red: best green: second best blue: third best

Moffet Field

Table: Quality measures for the Moffett field dataset[21]

method	**CC**	**SAM**	**RMSE**	**ERGAS**	**Time**(sec)
SFIM	0.92955	9.5271	365.2577	6.5429	1.26
MTF-GLP	0.93919	9.4599	352.1290	6.0491	1.86
MTF-GLP-HPM	0.93817	9.3567	354.8167	6.1992	1.71
GS	0.90521	14.1636	443.4351	7.5952	4.77
GSA	0.93857	11.2758	363.7090	6.2359	5.52
PCA	0.89580	14.6132	463.2204	7.9283	3.46
GFPCA	0.91614	11.3363	404.2979	7.0619	2.58
CNMF	0.95496	9.4177	314.4632	5.4200	10.98
Supervised Gaussian	0.97785	7.1308	220.0310	3.7807	1.31
Sparse represent.	0.98168	6.6392	200.3365	3.4281	133.61
HySure	0.97059	7.6351	254.2005	4.3582	140.05

[21] red: best green: second best blue: third best

Garons

Table: Quality measures for the Garons dataset[22]

method	CC	SAM	RMSE	ERGAS	Time(sec)
SFIM	0.77052	6.7356	1036.4695	5.1702	2.74
MTF-GLP	0.80124	6.6155	956.3047	4.8245	4.00
MTF-GLP-HPM	0.79989	6.6905	962.1076	4.8280	2.98
GS	0.80347	6.6627	1037.6446	5.1373	5.56
GSA	0.80717	6.7719	928.6229	4.7076	5.99
PCA	0.81452	6.6343	1021.8547	5.0166	6.09
GFPCA	0.63390	7.4415	1312.0373	6.3416	4.36
CNMF	0.82993	6.9522	893.9194	4.4927	23.98
Supervised Gaussian	0.86857	5.8749	784.1298	3.9147	3.07
Sparse represent.	0.87834	5.6377	750.3510	3.7629	259.44
HySure	0.86080	6.0224	778.1051	4.0454	177.60

[22] red: best green: second best blue: third best

Summary

State-of-the-Art

Revisited Bayesian Fusion

Sparse Prior Based on Dictionary Learning

Fast fUsion Based on a Sylvester Equation (FUSE)

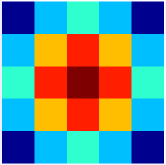
Conclusions

Summary

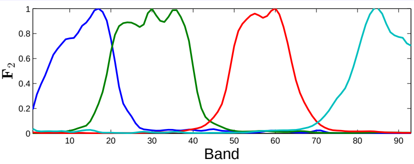
- Fusion of multi band images formulated as a linear inverse problem, which exploits explicitly the forward model
- Constrain the estimated image in a lower-dimensional space
- Definition of multiple priors within a (hierarchical) Bayesian framework
  - Gaussian prior
  - Sparse prior from dictionary learning
- Estimation of noise variances is possible with the proposed algorithm
- The spectral response of the MS image can be included in the estimation at the price of a higher computational complexity
- Toward fast fusion by solving a Sylvester equation: can be generalized to various priors

Future Work

- Blind Hyperspectral Unmixing
  - joint estimation of the HS and MS degradation operators: **B** and **R**



(a) Kernel of **B** (Spatial blurring)



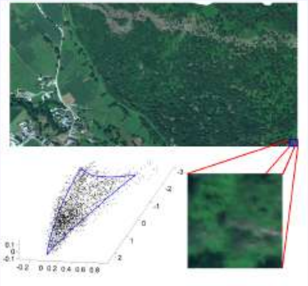
(b) **R** (Spectral blurring)

  - incorporating other physical models: unmixing, MRF, etc.
- Tensor Analysis?
  C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos and W.-K. Ma, "Hyperspectral Super-Resolution: A Coupled Tensor Factorization Approach," https://arxiv.org/pdf/1804.05307.pdf, April 2018.

Future Work

- Real data
  - Misregistration: different sensors, platforms
  - Nonlinear degradations: translation, rotation, stretching

(c) Nonlinear Unmixing



(d) Myth or Reality?

N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin and A. O. Hero, "Nonlinear unmixing of hyperspectral images: models and algorithms," *IEEE Signal Processing Magazine*, Jan. 2014.

  - Regularization parameters: included within the estimation scheme

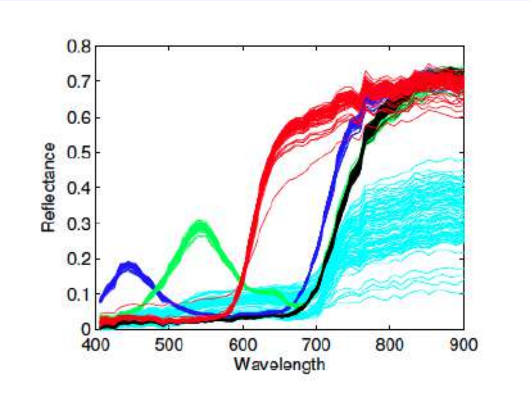
Challenge 1: Multi-temporal Images

Fusion of Snoopy and Nishino Japanese Islands - Pleiades Images

Thanks to the CNES of Toulouse for providing these Pleiades images

Challenge 2: Endmember variability

How to account for endmember variability in the fusion of hyperspectral and PAN/Multispectral Images?



Thanks to Alina Zare from the university of Missouri-Columbia for providing this image

Thanks to my student QI WEI!!

Qi WEI

Qi Wei (S'13) was born in Shanxi, China, in 1989. He received the B.Sc. degree in electrical engineering from Beihang University (BUAA), Beijing, China, in 2010. He is currently working toward the Ph.D. degree with the National Polytechnic Institute of Toulouse (INP-ENSEEIHT, University of Toulouse), Toulouse, France.

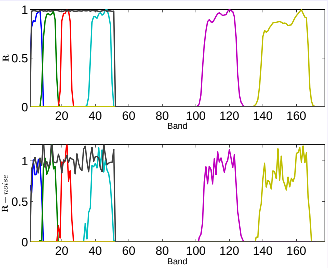
He is also currently with the Signal and Communications Group, IRIT Laboratory, University of Toulouse. From February to August 2012, he was an Exchange Master Student with the Signal Processing and Communications Group, Department of Signal Theory and Communications (TSC), Universitat Politècnica de Catalunya, Barcelona, Spain. His research has been focused on statistical signal processing, particularly on inverse problems in image processing.

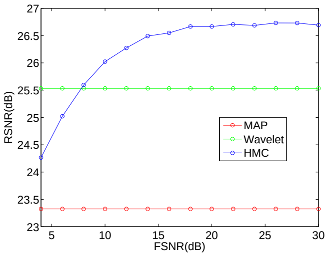
Thanks for your attention! Questions?

Robustness with respect to R

FSNR: defined to adjust the knowledge of **R**

$$\mathrm{FSNR} = 10 \log_{10}\left(\frac{\|\mathbf{R}\|_F^2}{m_\lambda n_{\lambda,2} s_2^2}\right)$$



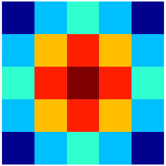


When FSNR is above 8dB, the proposed method outperforms the MAP and wavelet-based MAP methods.

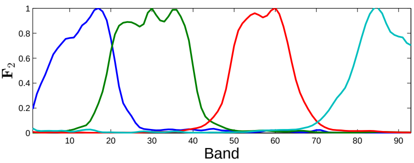
How to proceed when R is unknown?

$$\mathbf{Y}_{\mathrm{H}} = \mathbf{XBS} + \mathbf{N}_{\mathrm{H}}, \quad \mathbf{Y}_{\mathrm{M}} = \mathbf{RX} + \mathbf{N}_{\mathrm{M}}$$

- **$\mathbf{X} \in \mathbb{R}^{m_\lambda \times n}$: full resolution unknown image**
- $\mathbf{Y}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ and $\mathbf{Y}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$: observed HS and MS images
- $\mathbf{B} \in \mathbb{R}^{n \times n}$: cyclic convolution operator acting on the bands
- $\mathbf{S} \in \mathbb{R}^{n \times m}$: downsampling matrix
- $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$: spectral response of the MS sensor
- $\mathbf{N}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ and $\mathbf{N}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times n}$: HS and MS noises, assumed to be a band-dependent Gaussian sequence



(g) Kernel of **B** (Hyperspectral)



(h) **R** (Multispectral)

Block Gibbs sampler with unknown R

for $t = 1$ to N_{MC} **do**

% Sampling the image covariance matrix

Sample $\boldsymbol{\Sigma}_{\mathbf{u}}^{(t)}$ from $f(\boldsymbol{\Sigma}|\mathbf{U}^{(t-1)}, \boldsymbol{s}^{2^{(t-1)}}, \mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}})$

% Sampling the multispectral noise variances

for $\ell = 1$ to n_λ **do**

Sample $s_{\mathrm{M},\ell}^{2(t)}$ from $f(s_{\mathrm{M},\ell}^{2}|\mathbf{U}^{(t-1)}, \mathbf{Y}_{\mathrm{M}})$,

end for

% Sampling the hyperspectral noise variances

for $\ell = 1$ to m_λ **do**

Sample $s_{\mathrm{H},\ell}^{2(t)}$ from $f(s_{\mathrm{H},\ell}^{2}|\mathbf{U}^{(t-1)}, \mathbf{Y}_{\mathrm{H}})$,

end for

% Sampling the pseudo spectral response

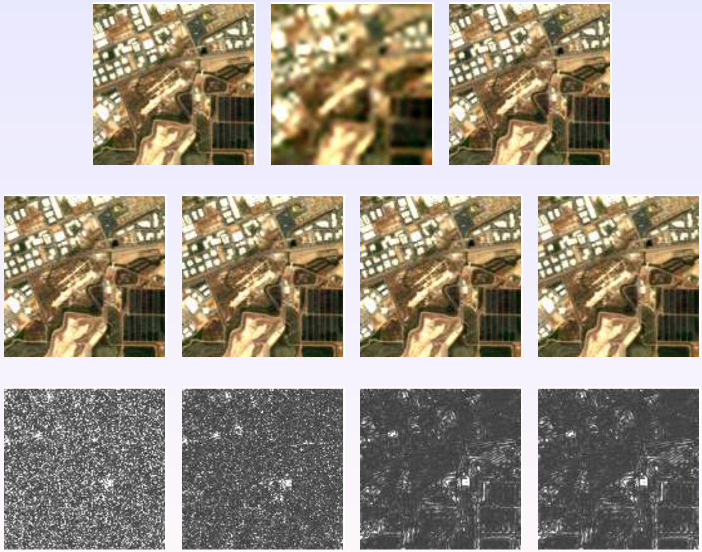
Sample $\underline{\mathbf{R}}$ from $f(\underline{\mathbf{R}}|\mathbf{U}^{(t-1)}, {s_{\mathrm{M}}^{2}}^{(t)}, \mathbf{Y}_{\mathrm{M}})$[23]

% Sampling the high-resolved image

Sample $\mathbf{U}^{(t)}$ using a Hamiltonian Monte Carlo algorithm

end for

[23] Q. Wei *et al.*, "Bayesian fusion of multispectral and hyperspectral images with unknown sensor spectral response", in *Proc. ICIP*, Paris, France, Oct. 2014.



(a) MAP (b) Wavelet MAP (c) HMC (known $\mathbf{R}$) (d) HMC ($\widehat{\mathbf{R}}$)

Quantitative fusion results

Table: Performance of the compared fusion methods: RSNR (in dB), UIQI, SAM (in degree), ERGAS, DD (in 10^{-2}) and Time (in second)(AVIRIS dataset).

Methods	RSNR	UIQI	SAM	ERGAS	DD	Time
MAP	16.655	0.9336	5.739	3.930	2.354	3
Wavelet MAP	19.501	0.9626	4.186	2.897	1.698	73
MCMC with known $\mathbf{R}$	21.913	0.9771	3.094	2.231	1.238	8811
MCMC with mismatched $\mathbf{R}$[24]	21.804	0.9764	3.130	2.260	1.257	8388
MCMC with $\mathbf{R}$ estimated	21.897	0.9769	3.101	2.234	1.244	10471

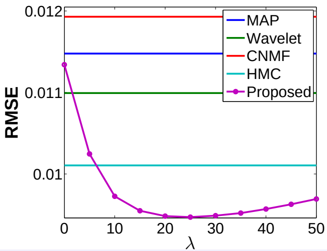
[24] FSNR= 10dB.

Estimation of the Spectral Response R

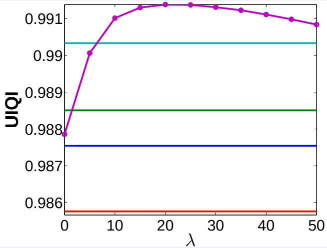
References

- M. Simões, J. Bioucas-Dias, L. B. Almeida and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization,"*IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373-3388, June 2015.
- N. Yokoya, N. Mayumi, and A. Iwasaki "Cross-calibration for data fusion of EO-1/Hyperion and Terra/ASTER," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sensing*, vol. 6, pp. 419-426, Apr. 2013.
- X. Otazu, M. González-Audícana, O. Fors and J. Núñez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods ," *IEEE Trans. Geosci. and Remote Sens.*, vol. 47, no. 11, pp. 3834-3843, Nov. 2009.

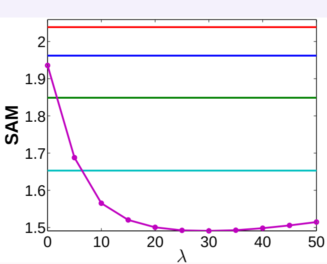
Performance versus λ



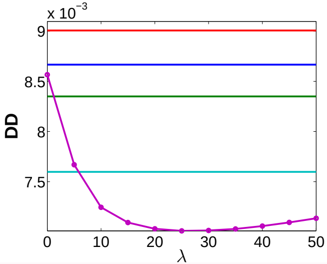
(e) RMSE



(f) UIQI



(g) SAM



(h) DD

Variational approaches

Observation models

- Hyperspectral (HS) image

$$\mathbf{Y}_{\mathrm{H}} = \mathbf{X}\mathbf{W} + \mathbf{N}_{\mathrm{H}}$$

where $\mathbf{X} \in \mathbb{R}^{m_\lambda \times n}$, $\mathbf{Y}_{\mathrm{H}}, \mathbf{N}_{\mathrm{H}} \in \mathbb{R}^{m_\lambda \times m}$ (with $m < n$), $\mathbf{W} \in \mathbb{R}^{n \times m}$ performs spatial averaging and downsampling. Note that the distribution of $\mathbf{N}_{\mathrm{H}}$ does not need to be specified.

- Panchromatic (PAN) or Multispectral (MS) image

$$\mathbf{Y}_{\mathrm{M}} = \mathbf{R}\mathbf{X} + \mathbf{N}_{\mathrm{M}}$$

where $\mathbf{Y} \in \mathbb{R}^{n_\lambda \times n}$, $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ is the spectral response of the MS sensor and $\mathbf{N}_{\mathrm{M}} \in \mathbb{R}^{n_\lambda \times m}$

- Dimensionality reduction

$$\mathbf{X} = \mathbf{H}\mathbf{U}$$

Variational approaches

HySure Algorithm

$$\min_{\mathbf{U}} L(\mathbf{U}) = \frac{1}{2}\|\mathbf{Y}_{\mathrm{H}} - \mathbf{HUBS}\|_F^2 + \frac{\lambda_M}{2}\|\mathbf{Y}_{\mathrm{M}} - \mathbf{RHU}\|_F^2 + \frac{\lambda}{2}\phi(\mathbf{UD}_h, \mathbf{UD}_v)$$

where $\mathbf{D}_h$ and $\mathbf{D}_v$ compute the horizontal and vertical differences and ϕ is a vector total variation ensuring smooth spatial variations.

References

- M. Simões, J. Bioucas-Dias, L. B. Almeida and J. Chanussot, “A Convex Formulation for Hyperspectral Image Superresolution (HySure) via Subspace-based Regularization,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373-3388, June 2015.
- X. He, L. Condat, J. Bioucas-Dias and J. Chanussot,“A New Pansharpening Method Based on Spatial and Spectral Sparsity Priors,” *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4160-4171, Sep. 2014.