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

Jean-Yves Tourneret\textsuperscript{(1)}

\textsuperscript{(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
Bayesian Fusion of Multi-band Images

Fusion of Multiband Images

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

<table>
<thead>
<tr>
<th>Name</th>
<th>AVIRIS (HS)</th>
<th>SPOT-5 (MS)</th>
<th>Pleiades (MS)</th>
<th>WorldView-3 (MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Res. (m)</td>
<td>20</td>
<td>10</td>
<td>2</td>
<td>1.24</td>
</tr>
<tr>
<td># bands</td>
<td>224</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

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
Review papers


Classes of Existing Methods

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

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
Component Substitution

Flowchart of CS Methods [Vivone2015]
Component Substitution

**Principle**

\[
\tilde{MS}_k = \tilde{MS}_k + g_k(P - I_L) \quad \text{with} \quad I_L = \sum_{k=1}^{m_\lambda} w_k \tilde{MS}_k
\] (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 distortion using [genetic algorithms](#)
- Optimization by linear regression
Component Substitution

Some references


## Image Characteristics

<table>
<thead>
<tr>
<th>DATASET</th>
<th>DIMENSIONS</th>
<th>SPATIAL RES</th>
<th>N</th>
<th>INSTRUMENT</th>
</tr>
</thead>
</table>
| 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

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
MultiResolution Analysis (MRA)

Flowchart of MRA Methods [Vivone2015].
MultiResolution Analysis

\textit{Principle}

\[ \tilde{M}_k = \tilde{M}_k + g_k (P - M_L), \quad k = 1, \ldots, m \lambda \]

where \( P_L \) is a low-pass version of \( P \).

\textit{Strategies for constructing} \( P_L \) \textit{and choosing the gains} \( g_k \)

\begin{itemize}
  \item Smoothing filter-based intensity modulation (SFIM)
    \[ P_L = P \ast h_{LP} \]
    where \( h_{LP} \) can be a box, Gaussian or Laplacian filter.
  \item Pyramidal decompositions: low-pass filter, wavelets, ...
  \item High-pass modulation paradigm for the gains
    \[ g_k = \tilde{M}_k \left( \frac{1}{m\lambda} \sum_{k=1}^{m\lambda} \tilde{M}_k \right)^{-1}, \quad k = 1, \ldots, m\lambda. \]
\end{itemize}
MultiResolution Analysis

Some references


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

State-of-the-Art

Component Substitution
MultiResolution Analysis
Bayesian Methods
Matrix Factorization

Revisited Bayesian Fusion

Sparse Prior Based on Dictionary Learning

Fast fUision Based on a Sylvester Equation (FUSE)

Conclusions
Bayesian Methods

Observation models

- Vectorized hyperspectral (HS) image

\[ y_H = Wx + n_H \]

where \( x \in \mathbb{R}^{m \times n}, y_H, n_H \in \mathbb{R}^{m \times m} \) (with \( m < n \)), \( W \in \mathbb{R}^{m \times m} \) performs spatial averaging and downsampling and \( n_H \sim \mathcal{N}(0, C_{n_H}) \).

- Vectorized Panchromatic (PAN) or Multispectral (MS) image: \((y_M, x)\) jointly Gaussian. Thus

\[ x | y_M \sim \mathcal{N}(\mu_M, C_M) \]

MAP estimator

- Direct form

\[ \tilde{x} = \left( W^T C_{n_H}^{-1} W + C_{M}^{-1} \right)^{-1} \left( W^T C_{n_H}^{-1} y_H + C_{M}^{-1} \mu_M \right) \]

- After matrix inversion lemma

\[ \tilde{x} = \mu_M + C_M W^T \left( W C_M W^T + C_{n_H} \right)^{-1} (y_H - W \mu_M) \]

Inversion of an \( m \times m \) matrix (instead of an \( m \times n \) matrix).
Determination of $\mu_M$ and $C_M$

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 $C_M$. 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)

$$y_{i,H} = \sum_{k=1}^{K} a_{i,k} m_k,$$

with $a_{i,k} > 0$, $\sum_{k=1}^{K} a_{i,k} = 1$ and $m_k$ 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


Bayesian Fusion of Multi-band Images

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
Matrix Factorization

**Observation models**

- **Hyperspectral (HS) image**
  \[ Y_H = XW + N_H \]
  where \( X \in \mathbb{R}^{m \times n} \), \( Y_H, N_H \in \mathbb{R}^{m \times m} \) (with \( m < n \)), \( W \in \mathbb{R}^{n \times m} \) performs spatial averaging and downsampling. Note that the distribution of \( N_H \) does not need to be specified.

- **Panchromatic (PAN) or Multispectral (MS) image**
  \[ Y_M = RX + N_M \]
  where \( Y \in \mathbb{R}^{n \times n} \), \( R \in \mathbb{R}^{n \times m} \) is the spectral response of the MS sensor and \( N_M \in \mathbb{R}^{n \times m} \).

**Linear mixing model**

- **Target image**
  \[ X = MA + N \]
  where \( M \in \mathbb{R}^{m \times p} \) and \( A \in \mathbb{R}^{p \times n} \) are the endmember and abundance matrices and \( p \) is the number of spectral signatures.
Matrix Factorization

New observation models

\[ Y_H = MW_A + \text{noise}, \quad Y_M = MR_A + \text{noise} \]

where \( W_A = AW \) is the spatially degraded abundance matrix and \( MR = RM \) 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 \( M \) and \( W_A \) by applying NMF to the HS image (initialized by the vertex component analysis (VCA))
- Step 2: estimate the abundance matrix \( A \) and \( MR \) by applying NMF to the MS image
- Step 2: fusion

\[ \hat{X} = \hat{M}A \]
Matrix Factorization

References


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

State-of-the-Art

Revisited Bayesian Fusion
  Problem Statement
  Hierarchical Bayesian model
  Block Gibbs sampler
  Simulation Results
  Accelerating with optimization

Sparse Prior Based on Dictionary Learning

Fast fUsion Based on a Sylvester Equation (FUSE)

Conclusions
Outline

State-of-the-Art

Revisited Bayesian Fusion
  Problem Statement
    Hierarchical Bayesian model
    Block Gibbs sampler
  Simulation Results
  Accelerating with optimization

Sparse Prior Based on Dictionary Learning

Fast fUsion Based on a Sylvester Equation (FUSE)

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

Spectral Responses. (a) IKONOS, (b) SPOT5, (c) LANDSAT.
Bayesian Fusion of Multi-band Images

Revisited Bayesian Fusion

Forward model

\[ Y_H = XBS + N_H, \quad Y_M = RX + N_M \]

- \( X \in \mathbb{R}^{m \times n} \): full resolution unknown image
- \( Y_H \in \mathbb{R}^{m \times m} \) and \( Y_M \in \mathbb{R}^{n \times n} \): observed HS and MS images
- \( B \in \mathbb{R}^{n \times n} \): cyclic convolution operator acting on the bands
- \( S \in \mathbb{R}^{n \times m} \): downsampling matrix
- \( R \in \mathbb{R}^{n \times m} \): spectral response of the MS sensor
- \( N_H \in \mathbb{R}^{m \times m} \) and \( N_M \in \mathbb{R}^{n \times n} \): 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 $\mathbf{X}$ in a lower-dimensional subspace ($\mathbb{R}^{\tilde{m}_\lambda}$): $\mathbf{X} = \mathbf{H}\mathbf{U}$, where $\mathbf{H}$ is an $\tilde{m}_\lambda \times m_\lambda$ projection matrix$^2$.

---

Likelihoods

- **Likelihood of the observations**

\[
\begin{align*}
Y_H | U, s^2_H & \sim \mathcal{MN}_{m, \lambda, m}(\text{HUBS}, \text{diag } (s^2_H), I_m) \\
Y_M | U, s^2_M & \sim \mathcal{MN}_{n, \lambda, n}(\text{RHU}, \text{diag } (s^2_M), I_n)
\end{align*}
\]

where 
\[
s^2_H = \left[ s^2_{H,1}, \ldots, s^2_{H,m} \right]^T \quad \text{and} \quad s^2_M = \left[ s^2_{M,1}, \ldots, s^2_{M,n} \right]^T.
\]

- **Joint likelihood**

\[
f (Y_H, Y_M | U, s^2) = f (Y_H | U, s^2_H) f (Y_M | U, s^2_M)
\]

with 
\[
s^2 = \{ s^2_H, s^2_M \}
\]

---

The probability density function of a **matrix normal** distribution is defined by

\[
p(\mathbf{X} | \mathbf{M}, \Sigma_r, \Sigma_c) = \frac{\exp \left( -\frac{1}{2} \text{tr} \left[ \Sigma_c^{-1} \left( \mathbf{X} - \mathbf{M} \right)^T \Sigma_r^{-1} \left( \mathbf{X} - \mathbf{M} \right) \right] \right)}{(2\pi)^{np/2} |\Sigma_c|^{n/2} |\Sigma_r|^{p/2}}
\]
Outline

State-of-the-Art

Revisited Bayesian Fusion
  Problem Statement
  Hierarchical Bayesian model
  Block Gibbs sampler
  Simulation Results
  Accelerating with optimization

Sparse Prior Based on Dictionary Learning

Fast fUsion Based on a Sylvester Equation (FUSE)

Conclusions
Parameter Priors

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

\[
\mathbf{U} | \mathbf{\bar{U}}, \Sigma \sim \mathcal{MN} (\mathbf{\bar{U}}, \Sigma, \mathbf{I}_n)
\]

- Noise variances: independent conjugate inverse-gamma priors

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

Flexible distribution whose shape can be adjusted from \((\nu, \gamma)\)

Assumptions

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

- \(\nu\): fixed (disappears later)
Hyperparameter Prior

Hyperparameter vector: \( \Phi = \{\Sigma, \gamma\} \)

- Hyperparameter \( \Sigma \): Inverse-Wishart (IW) distribution
  \[ \Sigma \sim \mathcal{W}^{-1}(\Psi, \eta) \]
  where \( \Psi \) and \( \eta \) are fixed to provide a non-informative prior

- Hyperparameter \( \gamma \): Jeffreys’ non-informative prior
  \[ f(\gamma) \propto \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma) \]
Using Bayes theorem, the joint posterior distribution is

\[ f(\theta, \Phi | Y_H, Y_M) \propto f(Y_H, Y_M | \theta) f(\theta | \Phi) f(\Phi) \]

where

- unknown parameters: \( \theta = \{U, s^2_H, s^2_M\} \)
- unknown hyperparameters: \( \Phi = \{\Sigma, \gamma\} \)

How can we estimate \( \theta \) and \( \Phi \)?

- Marginalize the hyperparameter \( \gamma \)
- Sample according to the joint posterior \( f(U, s^2, \Sigma | Y_H, Y_M) \) by using a block Gibbs sampler, which can be easily implemented since all the conditional distributions associated with \( f(U, s^2, \Sigma | Y_H, Y_M) \) are simple.
Outline

State-of-the-Art

Revisited Bayesian Fusion
  Problem Statement
  Hierarchical Bayesian model
  Block Gibbs sampler
  Simulation Results
  Accelerating with optimization

Sparse Prior Based on Dictionary Learning

Fast fUsion Based on a Sylvester Equation (FUSE)

Conclusions
Bayesian Fusion of Multi-band Images

Revisited Bayesian Fusion

**Block Gibbs sampler\(^4\)**

\[
\text{for } t = 1 \text{ to } N_{\text{MC}} \text{ do}
\]
\[
\text{\% Sampling the image covariance matrix}
\]
\[
\text{Sample } \Sigma^{(t)} \text{ from } f(\Sigma|U^{(t-1)}, s^{2(t-1)}, Y_H, Y_M)
\]
\[
\text{\% Sampling the multispectral noise variances}
\]
\[
\text{for } \ell = 1 \text{ to } n \text{ do}
\]
\[
\text{Sample } s^{2(t)}_{M,\ell} \text{ from } f(s^{2}_{M,\ell}|U, Y_M),
\]
\[
\text{end for}
\]
\[
\text{\% Sampling the hyperspectral noise variances}
\]
\[
\text{for } \ell = 1 \text{ to } m \text{ do}
\]
\[
\text{Sample } s^{2(t)}_{H,\ell} \text{ from } f(s^{2}_{H,\ell}|U, Y_H),
\]
\[
\text{end for}
\]
\[
\text{\% Sampling the high-resolved image}
\]
\[
\text{Sample } U^{(t)} \text{ using a Hamiltonian Monte Carlo algorithm}
\]
\[
\text{end for}
\]

Conditional Distributions

- **Covariance matrix of the image** $\Sigma$

$$\Sigma | u, s^2, Y_H, Y_M \sim \mathcal{W}^{-1} \left( \Psi + \sum_{i=1}^{m_x m_y} (u_i - \mu_u^{(i)})^T (u_i - \mu_u^{(i)}), n + \eta \right)$$

- **Noise variance vector** $s^2$

$$s_{H,\ell}^2 | U, Y_H \sim \mathcal{IG} \left( \frac{m}{2}, \frac{\| Y_H - \text{HUBS} \|_F^2}{2} \right)$$

$$s_{M,\ell}^2 | U, Y_H \sim \mathcal{IG} \left( \frac{n}{2}, \frac{\| Y_M - \text{RHU} \|_F^2}{2} \right)$$
Conditional Distributions (Cont.)

- Highly-resolved image $\mathbf{U}$

\[
- \log f(\mathbf{U}|\Sigma, s^2, \mathbf{Y}_H, \mathbf{Y}_M) = \frac{1}{2} \| \Lambda_H^{-\frac{1}{2}} (\mathbf{Y}_H - \text{HUBS}) \|_F^2 + \\
\frac{1}{2} \| \Lambda_M^{-\frac{1}{2}} (\mathbf{Y}_M - \text{RHU}) \|_F^2 + \frac{1}{2} \| \Sigma^{-\frac{1}{2}} (\mathbf{U} - \mu_U) \|_F^2 + C
\]

- 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. $\mathbf{U}$
- 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.

---


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

State-of-the-Art

Revisited Bayesian Fusion

Problem Statement
Hierarchical Bayesian model
Block Gibbs sampler

Simulation Results
Accelerating with optimization

Sparse Prior Based on Dictionary Learning

Fast fUsion Based on a Sylvester Equation (FUSE)

Conclusions
Bayesian Fusion of Multi-band Images

Revisited Bayesian Fusion

Wald’s Protocol

Hyperspectral reference $X$

Blurring and downsampling

Observed hyperspectral image $Y_H$

Multispectral spectral response

Observed multispectral image $Y_M$

Multiband fusion approach

Fused image $\hat{X}$

Quality measures

$Q$
Qualitative Results (AVIRIS dataset)

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

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

RMSE
Quantitative Performance Measures

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

$$\text{RMSE}(X, \hat{X}) = \frac{1}{nm} \| X - \hat{X} \|_F^2$$

$$\text{RSNR}(X, \hat{X}) = \log \frac{1}{nm} \frac{\| X \|_F^2}{\text{RMSE}}$$

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

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

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

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 $\hat{a}$ is

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

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 $\hat{X}$ is defined as

$$
DD(X, \hat{X}) = \frac{1}{nm} \| \text{vec}(X) - \text{vec}(\hat{X}) \|_1.
$$

The smaller DD, the better.
**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

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

where \(1/d^2\) is the ratio between the pixel sizes of the MS and HS images, \(\mu_i\) is the mean of the \(i\)th band of the HS image, and \(m_\lambda\) is the number of HS bands. The smaller ERGAS, the smaller the spectral distortion.
Quantitative Results (AVIRIS dataset)

<table>
<thead>
<tr>
<th>Methods</th>
<th>RSNR</th>
<th>UIQI</th>
<th>SAM</th>
<th>ERGAS</th>
<th>DD</th>
<th>Time(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP(^7)</td>
<td>23.33</td>
<td>0.9913</td>
<td>5.05</td>
<td>4.21</td>
<td>4.87</td>
<td>1.6</td>
</tr>
<tr>
<td>Wavelet(^8)</td>
<td>25.53</td>
<td>0.9956</td>
<td>3.98</td>
<td>3.95</td>
<td>3.89</td>
<td>31</td>
</tr>
<tr>
<td>MCMC</td>
<td>26.74</td>
<td>0.9966</td>
<td>3.40</td>
<td>3.77</td>
<td>3.33</td>
<td>530</td>
</tr>
</tbody>
</table>

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


Noise Variance Estimation

![Graphs showing noise variances and their MMSE estimates.](image)

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

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

State-of-the-Art

Revisited Bayesian Fusion
  Problem Statement
  Hierarchical Bayesian model
  Block Gibbs sampler
  Simulation Results
  Accelerating with optimization

Sparse Prior Based on Dictionary Learning

Fast fUsion Based on a Sylvester Equation (FUSE)

Conclusions
Accelerating With Optimization

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

$$L(\mathbf{U}, \mathbf{s}^2, \Sigma) = -\log p(\theta, \Sigma|\mathbf{Y})$$

$$= -\log p(\mathbf{Y}_H|\theta) - \log p(\mathbf{Y}_M|\theta) - \sum_{i=1}^{n} \log p(\mathbf{U}_i|\Sigma)$$

$$- \sum_{i=1}^{m_{\lambda}} \log p\left(s_{H,i}^2\right) - \sum_{j=1}^{n_{\lambda}} \log p\left(s_{M,j}^2\right) - \log p(\Sigma) - C$$

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

---

Block Coordinated Descent for HS and MS Image Fusion

Input: \( Y_H, Y_M, \tilde{m}_\lambda, B, S, R, s_0^2, \Sigma_0 \)

1. \( \hat{H} \leftarrow \text{PCA}(Y_H, \tilde{m}_\lambda) \); /* Subspace transform matrix */

2. \textbf{for} \( t = 1, 2, \ldots \) to stopping rule \textbf{do}

3. \quad \( U_t = \arg \min_U L(U, s_{t-1}^2, \Sigma_{t-1}) ; \) /* Optimize w.r.t. \( U \) */

4. \quad \( s_t^2 = \arg \min_{s^2} L(U_t, s^2, \Sigma_{t-1}) ; \) /* Optimize w.r.t. \( s^2 \) */

5. \quad \( \Sigma_t = \arg \min_\Sigma L(U_t, s_t^2, \Sigma) ; \) /* Optimize w.r.t. \( \Sigma \) */

\textbf{end}

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

Remarks
The convergence is guaranteed\(^{10}\).

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

$$L_U(U) = \frac{1}{2} \| \Lambda_{H}^{-\frac{1}{2}} (Y_H - HUBS) \|_F^2 + \frac{1}{2} \| \Lambda_{M}^{-\frac{1}{2}} (Y_M - RHU) \|_F^2 + \frac{1}{2} \| \Sigma^{-\frac{1}{2}} (U - \mu_U) \|_F^2.$$ 

- 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 $H(\cdot)BS$ not possible
Bayesian Fusion of Multi-band Images

Revisited Bayesian Fusion

**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:** $H_1 = UB$, $H_2 = U$ and $H_3 = U$
- **Respective scaled dual variable:** $G_1$, $G_2$, $G_3$

\[
L(U, H_1, H_2, H_3, G_1, G_2, G_3) = \frac{1}{2} \left\| \Lambda_H^{-\frac{1}{2}} (Y_H - HH_1 S) \right\|^2_F + \frac{\mu}{2} \left\| UB - H_1 - G_1 \right\|^2_F + \frac{1}{2} \left\| \Lambda_M^{-\frac{1}{2}} (Y_M - RHH_2) \right\|^2_F + \frac{\mu}{2} \left\| U - H_2 - G_2 \right\|^2_F + \frac{1}{2} \left\| \Sigma^{-\frac{1}{2}} (\mu U - H_3) \right\|^2_F + \frac{\mu}{2} \left\| U - H_3 - G_3 \right\|^2_F
\]

**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:** $H_1 = UB$, $H_2 = U$ and $H_3 = U$
- **Respective scaled dual variable:** $G_1$, $G_2$, $G_3$

$$L(U, H_1, H_2, H_3, G_1, G_2, G_3) = \begin{align*}
\frac{1}{2} & \| \Lambda_H^{-\frac{1}{2}} (Y_H - HH_1 S) \|_F^2 + \frac{\mu}{2} \| UB - H_1 - G_1 \|_F^2 + \\
\frac{1}{2} & \| \Lambda_M^{-\frac{1}{2}} (Y_M - RHH_2) \|_F^2 + \frac{\mu}{2} \| U - H_2 - G_2 \|_F^2 + \\
\frac{1}{2} & \| \Sigma^{-\frac{1}{2}} (\mu U - H_3) \|_F^2 + \frac{\mu}{2} \| U - H_3 - G_3 \|_F^2
\end{align*}$$
**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**: \( H_1 = UB, \ H_2 = U \) and \( H_3 = U \)
- **Respective scaled dual variable**: \( G_1, G_2, G_3 \)

\[
L(U, H_1, H_2, H_3, G_1, G_2, G_3) = \frac{1}{2} \left\| \Lambda_H^{-\frac{1}{2}} (Y_H - HH_1 S) \right\|_F^2 + \frac{\mu}{2} \left\| UB - H_1 - G_1 \right\|_F^2 + \\
\frac{1}{2} \left\| \Lambda_M^{-\frac{1}{2}} (Y_M - RHH_2) \right\|_F^2 + \frac{\mu}{2} \left\| U - H_2 - G_2 \right\|_F^2 + \\
\frac{1}{2} \left\| \Sigma^{-\frac{1}{2}} (\mu U - H_3) \right\|_F^2 + \frac{\mu}{2} \left\| U - H_3 - G_3 \right\|_F^2
\]
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: $H_1 = UB$, $H_2 = U$ and $H_3 = U$
- Respective scaled dual variable: $G_1, G_2, G_3$

\[
\begin{align*}
L(U, H_1, H_2, H_3, G_1, G_2, G_3) &\text{Denoising} \\
= &\frac{1}{2} \left\| \Lambda_H^{-\frac{1}{2}} (Y_H - HH_S) \right\|_F^2 + \frac{\mu}{2} \left\| UB - H_1 - G_1 \right\|_F^2 + \\
&\frac{1}{2} \left\| \Lambda_M^{-\frac{1}{2}} (Y_M - RHH_2) \right\|_F^2 + \frac{\mu}{2} \left\| U - H_2 - G_2 \right\|_F^2 + \\
&\frac{1}{2} \left\| \Sigma^{-\frac{1}{2}} (\mu U - H_3) \right\|_F^2 + \frac{\mu}{2} \left\| U - H_3 - G_3 \right\|_F^2
\end{align*}
\]
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).

<table>
<thead>
<tr>
<th>Methods</th>
<th>RSNR</th>
<th>UIQI</th>
<th>SAM</th>
<th>ERGAS</th>
<th>DD</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP</td>
<td>23.14</td>
<td>0.9932</td>
<td>5.147</td>
<td>3.524</td>
<td>4.958</td>
<td>3</td>
</tr>
<tr>
<td>Wavelet MAP</td>
<td>24.91</td>
<td>0.9956</td>
<td>4.225</td>
<td>3.282</td>
<td>4.120</td>
<td>72</td>
</tr>
<tr>
<td>MCMC</td>
<td>25.92</td>
<td>0.9971</td>
<td>3.733</td>
<td>2.926</td>
<td>3.596</td>
<td>6228</td>
</tr>
<tr>
<td>BCD</td>
<td>25.85</td>
<td>0.9970</td>
<td>3.738</td>
<td>2.946</td>
<td>3.600</td>
<td>96</td>
</tr>
</tbody>
</table>

- 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

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
**Motivation**

Self-similarity property of natural image patches
Remote Sensing Images

image

patches
Sparse Regularization

The patches of the target image $\mathbf{U}$ can be \textit{sparsely} approximated on an \textit{over-complete} dictionary (with columns referred to as \textit{atoms}).
Based on the previous observation models and dimensionality reduction, fusing the HS and MS images can be formulated as the following inverse problem

\[
\min_U \frac{1}{2} \left\| \Lambda_H^{-\frac{1}{2}} (Y_H - \text{HUBS}) \right\|_F^2 + \frac{1}{2} \left\| \Lambda_H^{-\frac{1}{2}} (Y_M - \text{RHU}) \right\|_F^2 + \lambda \phi(U),
\]

- **HS data term** $\propto \ln p(Y_H | U)$
- **MS data term** $\propto \ln p(Y_M | U)$
- **regularizer** $\propto \ln p(U)$
Sparse Regularization

Regularizer

\[ \phi(U) = \frac{1}{2} \| U - \bar{U}(D, A) \|_F^2 \]

Separating each band of the target image leads to

\[ \phi(U) = \frac{1}{2} \sum_{i=1}^{\tilde{m}_\lambda} \| U_i - \mathcal{P}(D_iA_i) \|_F^2 \]

- **\( U_i \in \mathbb{R}^n \) is the \( i \)th band (or row) of \( U \in \mathbb{R}^{\tilde{m}_\lambda \times n} \)**
- **\( D_i \in \mathbb{R}^{n_p \times n_{at}} \) is the dictionary dedicated to the \( i \)th band of \( U \) (\( n_p \) is the patch size and \( n_{at} \) is the number of atoms) and \( D = [D_1, \ldots, D_{\tilde{m}_\lambda}] \)**
- **\( A_i \in \mathbb{R}^{n_{at} \times n_{pat}} \) is the \( i \)th band’s code (\( n_{pat} \) is the number of patches associated with the \( i \)th band) and \( A = [A_1, \ldots, A_{\tilde{m}_\lambda}] \)**
- **\( \mathcal{P}(\cdot) \) 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\textsuperscript{12} the set of over-complete dictionaries \( \mathbf{D} = [\mathbf{D}_1, \cdots, \mathbf{D}_{\tilde{m}_\lambda}] \): 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 \( n_{\text{max}} \) coefficients) for each band \( \mathbf{U}_i \)
- Support (\( \Omega_i \subset \mathbb{N}^2, i = 1, \cdots, \tilde{m}_\lambda \)): The positions of the non-zero elements of the code \( \mathbf{A}_i \) are also identified

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.

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

where $\| . \|_0$ is the $\ell_0$ counting function (or $\ell_0$ norm) and $\mu_a$ is a regularization parameter.

By fixing the supports $\Omega_i$, the $\ell_0$ norm reduces to a constant. Hence,

\[
\phi(\mathbf{U}, \mathbf{A}) = \frac{1}{2} \sum_{i=1}^{\tilde{m}_\lambda} \| \mathbf{U}_i - \mathcal{P}(\mathbf{D}_i \mathbf{A}_i) \|_F^2 \quad \text{s.t. } \mathbf{A}_i, \mathcal{O}_i = 0
\]

where $\mathbf{A}_i, \mathcal{O}_i = \{ \mathbf{A}_i(l, k) \mid (l, k) \not\in \Omega_i \}$.
Final Optimization Problem

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

$$
\min_{U,A} L(U, A) = \frac{1}{2} \| \Lambda_{H}^{-\frac{1}{2}} (Y_{H} - \text{HUBS}) \|_{F}^{2} + \frac{1}{2} \| \Lambda_{M}^{-\frac{1}{2}} Y_{M} - \text{RHU} \|_{F}^{2} + \\
\frac{\lambda}{2} \sum_{i=1}^{\tilde{m}} \left( \| U_{i} - \mathcal{P}(D_{i}A_{i}) \|_{F}^{2} \right), \text{ s.t. } A_{i,\backslash\Omega_{i}} = 0
$$

Solution

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

State-of-the-Art

Revisited Bayesian Fusion

Sparse Prior Based on Dictionary Learning
  Sparse Regularization
  Alternate Optimization Scheme

Fast fUision Based on a Sylvester Equation (FUSE)

Conclusions
Optimization with Respect to $U$

$$\min_U L(U) = \frac{1}{2} \left\| \Lambda_H^{-\frac{1}{2}} (Y_H - HUBS) \right\|_F^2 + \frac{1}{2} \left\| \Lambda_M^{-\frac{1}{2}} (Y_M - RHU) \right\|_F^2 +$$

$$\frac{\lambda}{2} \sum_{i=1}^{\tilde{m}} \left( \left\| U_i - P(D_i A_i) \right\|_F^2 \right),$$

**Difficulties**

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

**Solution**

Alternating Direction Method of Multipliers (ADMM)
Optimization with Respect to $A$

Optimization with respect to $A_i$ ($i = 1, \cdots, \tilde{m}_\lambda$) conditional on $U_i$

$$\min_{A_i} \|U_i - \mathcal{P}(D_i A_i)\|_F^2 \text{ s.t. } A_i, \Omega_i = 0$$

Remarks

- The optimization with respect to $A_i$ considers only the non-zero elements of $A_i$, denoted as $A_{i,\Omega_i} = \{A_i(l, k) \mid (l, k) \in \Omega_i\}$
- Standard least square (LS) problem which can be solved analytically
Alternate Optimization Scheme$^{13}$

**Input:** $Y_H, Y_M, B, S, R, \text{SNR}_H, \text{SNR}_M, \tilde{m}_\lambda, n_{\text{max}}$

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

/* Start alternate optimization */

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

\[
\hat{U}_t \in \{U : L(U, \hat{A}_{t-1}) \leq L(\hat{U}_{t-1}, \hat{A}_{t-1})\} \quad \text{/* solved with ADMM */}
\hat{A}_t \in \{A : L(\hat{U}_t, A) \leq L(\hat{U}_t, \hat{A}_{t-1})\} \quad \text{/* solved with LS */}
```

end

**Output:** $\hat{X}$ (high resolution HS image)

---

### Image Characteristics

#### Table 2. Characteristic of the Three Datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Dimensions</th>
<th>Spatial Res</th>
<th>N</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moffett</td>
<td>PAN 185 × 395</td>
<td>20m</td>
<td>224</td>
<td>AVIRIS</td>
</tr>
<tr>
<td></td>
<td>HS 37 × 79</td>
<td>100m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Camargue</td>
<td>PAN 500 × 500</td>
<td>4m</td>
<td>125</td>
<td>HyMap</td>
</tr>
<tr>
<td></td>
<td>HS 100 × 100</td>
<td>20m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garons</td>
<td>PAN 400 × 400</td>
<td>4m</td>
<td>125</td>
<td>HyMap</td>
</tr>
<tr>
<td></td>
<td>HS 80 × 80</td>
<td>20m</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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)

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

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).

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

\[ Y_H = XBS + N_H, \quad Y_M = RX + N_M \]

s.t. \( X = HU \)

Negative log-likelihood (in subspace)

\[
- \log p(Y|U) = - \log p(Y_H|U) - \log p(Y_M|U) + C \\
= \frac{1}{2} \| \Lambda_H^{-\frac{1}{2}} (Y_H - HUBS) \|^2 + \frac{1}{2} \| \Lambda_M^{-\frac{1}{2}} (Y_M - RHU) \|^2 + C
\]

Minimizing the likelihood w.r.t. \( U \) ⇔ solve a generalized Sylvester matrix equation

\[
\begin{bmatrix}
H^H \Lambda_H^{-1} H \\
BS (BS)^H
\end{bmatrix}
U
\begin{bmatrix}
BS (BS)^H \\
(RH)^H \Lambda_M^{-1} (RH)
\end{bmatrix}
U = \text{term ind. on } U
\]

▸ BS (BS)^H is not diagonalizable!
Assumption 1

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

Assumption 2

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

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

These two assumptions are used to compute an **explicit solution** of the Sylvester equation.
\[
\begin{bmatrix}
H^H \Lambda_H^{-1} H
\end{bmatrix}
U
\begin{bmatrix}
BS (BS)^H
\end{bmatrix}
+ \begin{bmatrix}
(RH)^H \Lambda_M^{-1} (RH)
\end{bmatrix}
U = \text{term ind. on } U
\]

3 Main Steps

- Left multiply by \((H^H \Lambda_H^{-1} H)^{-1}\): \(UC_2 + C_1 U = C_3\), where \(C_2 = BS (BS)^H\).

**Lemma 1**

The equality \(F^H S F = \frac{1}{d} J_d \otimes I_m\) holds, where \(F\) is the DFT matrix, \(S = SS^H\), \(J_d\) is the \(d \times d\) matrix of ones and \(I_m\) is the \(m \times m\) identity matrix.

- Diagonalize \(C_1\) and use Lemma 1 to simplify \(C_2\):

\[
\tilde{U}M + \Lambda_C \tilde{U} = \tilde{C}_3
\]

with a diagonal matrix \(\Lambda_C\) and \(M = \frac{1}{d} \begin{bmatrix}
\sum_{i=1}^{d} D_i & D_2 & \cdots & D_d
\end{bmatrix}
\]

\(D_i; m \times m\) diagonal matrix, \(d\): downsampling ratio, \(m\): number of image pixels
\textbf{Theorem 2}

\textit{Let} $(\bar{C}_3)_{l,j}$ \textit{denotes the} $j$\textit{th block of the} $l$\textit{th band of} $\bar{C}_3$ \textit{for any} $l = 1, \ldots, \tilde{m}_\lambda$. \textit{Then, the solution} $\bar{U}$ \textit{of the SE can be decomposed as}

\[
\bar{U} = \begin{bmatrix}
\bar{u}_{1,1} & \bar{u}_{1,2} & \cdots & \bar{u}_{1,d} \\
\bar{u}_{2,1} & \bar{u}_{2,2} & \cdots & \bar{u}_{2,d} \\
\vdots & \vdots & \ddots & \vdots \\
\bar{u}_{\tilde{m}_\lambda,1} & \bar{u}_{\tilde{m}_\lambda,2} & \cdots & \bar{u}_{\tilde{m}_\lambda,d}
\end{bmatrix}
\]

\textit{with}

\[
\bar{u}_{l,j} = \begin{cases}
(\bar{C}_3)_{l,j} \left( \frac{1}{d} \sum_{i=1}^{d} D_i + \lambda_C l_m \right)^{-1}, & j = 1, \\
\frac{1}{\lambda_C} \left[ (\bar{C}_3)_{l,j} - \frac{1}{d} \bar{u}_{l,1} D_j \right], & j = 2, \ldots, d.
\end{cases}
\]

Fast Fusion Based on a Sylvester Equation (FUSE)

Input: $Y_M, Y_H, \Lambda_M, \Lambda_H, R, B, S, H$

- $D \leftarrow \text{Dec} (B)$ and $D = D^* D$ /* Circulant matrix: $B = FDF^H$ */

- $C_1 \leftarrow \left( H^H \Lambda_H^{-1} H \right)^{-1} \left( (RH)^H \Lambda_L^{-1} RH \right)$;

- $(Q, \Lambda_C) \leftarrow \text{EigDec} (C_1)$ /* Eigen-dec of $C_1$: $C_1 = Q \Lambda_C Q^{-1}$ */

- $\tilde{C}_3 \leftarrow$

  $$Q^{-1} \left( H^H \Lambda_H^{-1} H \right)^{-1} \left( H^H \Lambda_H^{-1} Y_H (BS)^H + (RH)^H \Lambda_L^{-1} Y_M \right) \text{BFP}^{-1};$$

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

  $$\tilde{u}_{l,1} = (\tilde{C}_3)_{l,1} \left( \frac{1}{d} \sum_{i=1}^d D_i + \lambda_C^l I_n \right)^{-1};$$

  for $j = 2$ to $d$ do

    $$\tilde{u}_{l,j} = \frac{1}{\lambda_C^l} \left( (\tilde{C}_3)_{l,j} - \frac{1}{d} \tilde{u}_{l,1} D_j \right);$$

  end

end

Output: $X = HQ\tilde{U}\tilde{P}D^{-1}F^H$
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
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)\(^{17}\)

---


Gaussian Prior

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

Input: $Y_H, Y_M, \tilde{m}_\lambda, B, S, R, s_0^2, \Sigma_0$

- $H \leftarrow \text{PCA}(Y_H, \tilde{m}_\lambda)$; /* Subspace transform matrix */

for $t = 1, 2, \ldots$ to stopping rule do
  $U_t = \arg\min_U L(U, s_{t-1}^2, \Sigma_{t-1})$; /* Sylvester equation */
  $s_t^2 = \arg\min_{s^2} L(U_t, s^2, \Sigma_{t-1})$; /* Optimize w.r.t. $s^2$ */
  $\Sigma_t = \arg\min_\Sigma L(U_t, s_t^2, \Sigma)$; /* Optimize w.r.t. $\Sigma$ */
end

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

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

**Input:** \( Y_H, Y_M, B, S, R, \text{SNR}_H, \text{SNR}_M, \tilde{m}_\lambda, n_{\text{max}} \)

**Output:** \( \hat{X} \) (high resolution HS image)

- Approximate \( \tilde{U} \) using \( Y_M \) and \( Y_H \) /* Rough estimation of \( U \) */
- \( \hat{D} \leftarrow \text{ODL}(\tilde{U}) \) /* Online dictionary learning */
- \( \hat{A} \leftarrow \text{OMP}(\hat{D}, \tilde{U}, n_{\text{max}}) \) /* Sparse coding */
- \( \hat{\Omega} \leftarrow \hat{A} \neq 0 \) /* Computing support */
- \( \hat{H} \leftarrow \text{PCA}(Y_H, \tilde{m}_\lambda) \) /* Computing subspace transform matrix */

/* Start alternate optimization */

\[
\begin{align*}
\text{for } t = 1, 2, \ldots \text{ to stopping rule do} \quad & \\
\quad \hat{U}_t \in \{U : L(U, \hat{A}_{t-1}) \leq L(\hat{U}_{t-1}, \hat{A}_{t-1})\} \quad & /* \text{solved with SE} */ \\
\quad \hat{A}_t \in \{A : L(\hat{U}_t, A) \leq L(\hat{U}_t, \hat{A}_{t-1})\} \quad & /* \text{solved with LS} */ \\
\text{end} \\
\hat{X} = \hat{H}\hat{U}
\end{align*}
\]
Non-Gaussian Prior

Non-Gaussian prior, such as \((TV)^{18}\)

\[
\arg\min_u \frac{1}{2} \| \Lambda_H^{-\frac{1}{2}} (Y_H - HUBS) \|_F^2 + \frac{1}{2} \| \Lambda_M^{-\frac{1}{2}} (Y_M - RHU) \|_F^2 + \lambda TV (U).
\]

\text{HS data term} \quad \text{MS data term}

This can be equivalently solved as

\[
\arg\min_{u,v} \frac{1}{2} \| \Lambda_H^{-\frac{1}{2}} (Y_H - HUBS) \|_F^2 + \frac{1}{2} \| \Lambda_M^{-\frac{1}{2}} (Y_M - RHU) \|_F^2 + \lambda TV (V)
\]

s.t. \(U = V\)

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

---

### 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).

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

- 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\(^{19}\)

<table>
<thead>
<tr>
<th>dataset</th>
<th>dimensions</th>
<th>spatial res</th>
<th>N</th>
<th>instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moffett</td>
<td>PAN 185 × 395</td>
<td>20m</td>
<td>224</td>
<td>AVIRIS</td>
</tr>
<tr>
<td></td>
<td>HS 37 × 79</td>
<td>100m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Camargue</td>
<td>PAN 500 × 500</td>
<td>4m</td>
<td>125</td>
<td>HyMap</td>
</tr>
<tr>
<td></td>
<td>HS 100 × 100</td>
<td>20m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garons</td>
<td>PAN 400 × 400</td>
<td>4m</td>
<td>125</td>
<td>HyMap</td>
</tr>
<tr>
<td></td>
<td>HS 80 × 80</td>
<td>20m</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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

20 red: best  green: second best  blue: third best
## Moffet Field

**Table:** Quality measures for the Moffett field dataset\(^{21}\)

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

\(^{21}\) *red:* best \hspace{1em} *green:* second best \hspace{1em} *blue:* third best
## Garons

**Table:** Quality measures for the Garons dataset[^22]

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

[^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 \( R \)

![Kernel of \( B \)](image1)

- incorporating **other physical models**: unmixing, MRF, etc.

- **Tensor Analysis?**
  C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos and W.-K. Ma,
Future Work

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

(c) Nonlinear Unmixing  
(d) Myth or Reality?


- **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 \( R \)

\[
\text{FSNR} = 10 \log_{10} \left( \frac{\| R \|_F^2}{m_\lambda n_\lambda s_2^2} \right)
\]

When FSNR is above 8dB, the proposed method outperforms the MAP and wavelet-based MAP methods.
How to proceed when $R$ is unknown?

$$Y_H = XBS + N_H, \quad Y_M = RX + N_M$$

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

(g) Kernel of $B$ (Hyperspectral)

(h) $R$ (Multispectral)
Bayesian Fusion of Multi-band Images

Block Gibbs sampler with unknown $R$

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

% Sampling the image covariance matrix
Sample $\Sigma_u^{(t)}$ from $f(\Sigma|U^{(t-1)}, s^{2(t-1)}, Y_H, Y_M)$

% Sampling the multispectral noise variances
for $\ell = 1$ to $n_\lambda$ do
  Sample $s_{M,\ell}^{2(t)}$ from $f(s_{M,\ell}^{2(t)}|U^{(t-1)}, Y_M)$,
end for

% Sampling the hyperspectral noise variances
for $\ell = 1$ to $m_\lambda$ do
  Sample $s_{H,\ell}^{2(t)}$ from $f(s_{H,\ell}^{2(t)}|U^{(t-1)}, Y_H)$,
end for

% Sampling the pseudo spectral response
Sample $R$ from $f(R|U^{(t-1)}, s_{M}^{2(t)}, Y_M)^{23}$

% Sampling the high-resolved image
Sample $U^{(t)}$ using a Hamiltonian Monte Carlo algorithm
end for

---

Bayesian Fusion of Multi-band Images

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

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

$^{24}$FSNR = 10dB.
Estimation of the Spectral Response R

References


Performance versus $\lambda$

(e) RMSE

(f) UIQI

(g) SAM

(h) DD

Figure: Performance of the proposed fusion algorithm versus $\lambda$. 
Variational approaches

Observation models

- **Hyperspectral (HS) image**
  \[ Y_H = XW + N_H \]
  where \( X \in \mathbb{R}^{m \times n} \), \( Y_H \), \( N_H \in \mathbb{R}^{m \times m} \) (with \( m < n \)), \( W \in \mathbb{R}^{n \times m} \) performs spatial averaging and downsampling. Note that the distribution of \( N_H \) does not need to be specified.

- **Panchromatic (PAN) or Multispectral (MS) image**
  \[ Y_M = RX + N_M \]
  where \( Y \in \mathbb{R}^{n \times n} \), \( R \in \mathbb{R}^{n \times m} \) is the spectral response of the MS sensor and \( N_M \in \mathbb{R}^{n \times m} \)

- **Dimensionality reduction**
  \[ X =HU \]
Variational approaches

HySure Algorithm

$$\min_{\mathbf{U}} L(\mathbf{U}) = \frac{1}{2} \| \mathbf{Y}_H - \text{HUBS} \|_F^2 + \frac{\lambda}{2} \| \mathbf{Y}_M - \text{RHU} \|_F^2 + \frac{\lambda}{2} \phi(\mathbf{UD}_h, \mathbf{UD}_v)$$

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

References
