Hyperspectral Super-Resolution: A Tensor-Based Approach

Xiao Fu

School of Electrical Engineering and Computer Science
Oregon State University

Ack: Charilaos Kanatsoulis, Nikos Sidiropoulos, Wing-Kin Ma

June 13, 2018

Oregon State University

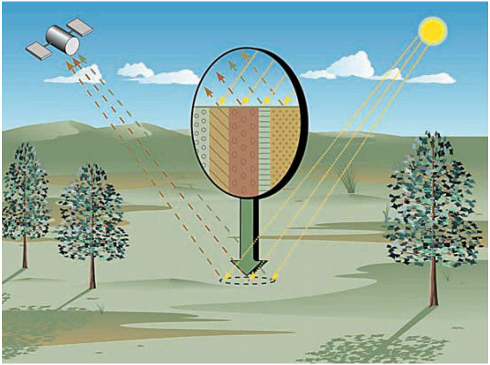
main campus, Corvallis, Oregon

OSU EECS

1. Oregon State University is the flagship and the largest public university in Oregon.
2. OSU is located in the west coast (one hour's drive to the gorgeous Oregon Coast and Portland).
3. Corvallis ranks the top 5 safest and nicest college town in the U.S.
4. OSU is celebrating its 150th anniversary; OSU is known for its strong engineering programs.
5. The school of EECS is the home of more than 60 faculty members, covering most areas in EE and CS.
6. OSU's CS program ranks 37 of the United States according to CSrankings.
7. The Artificial Intelligence (AI) program ranks 21 all over the country.

Hyperspectral Imaging

1. Hyperspectral sensor records EM scattering patterns of distinct materials over hundreds of spectral bands (from visible to near-infrared wavelength) [Keshava *et al.* '02].



Hyperspectral Pixels

1. Every pixel of a hyperspectral image is a high-dimensional vector:

$$\boldsymbol{y}_\ell \in \mathbb{R}^{K_H},\ \ell = 1, \ldots, L_H,$$

where

- ℓ is the pixel index
- K_H is the number of frequency bands
- $y_{m,\ell}$ is the spectral intensity of pixel ℓ at frequency m

Hyperspectral Cube

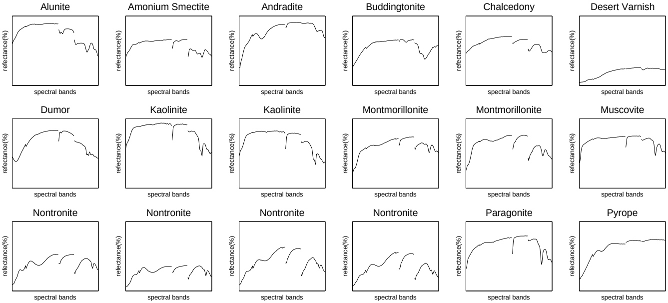
1. Hyperspectral images come as cubes:

Hyperspectral Cube

1. Hyperspectral images come as cubes:

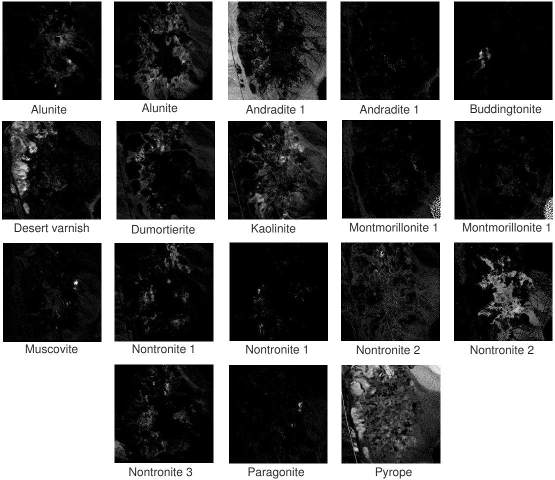
2. Why is this important?

1. Spectral information is rich.



† the spectra are from the U.S.G.S. library.

1. greatly helps material identification on the ground.



† figure from [Chan *et al.* '11].

Who're interested in HSI?

NASA SPINOFF

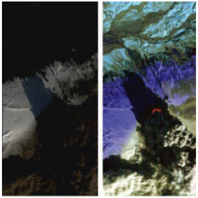
NASA TECHNOLOGY TRANSFER PROGRAM

Hyperspectral Systems Increase Imaging Capabilities

Space Technology Hall of Fame icon
Health and Medicine

Originating Technology/NASA Contribution

While the human eye can see a range of phenomena in the world, there is a larger range that it cannot see. Without the aid of technology, people are limited to seeing wavelengths of visible light, a tiny range within the electromagnetic spectrum. Hyperspectral imaging, however, allows people to get a glimpse at how objects look in the ultraviolet (UV) and infrared wavelengths —the ranges on either side of visible light on the spectrum.



The Hyperion instrument onboard the Earth Observing-1 spacecraft obtained these images of Iceland's Eyjafjallajökull volcano. The left-hand image was created with visible wavelengths; the right-hand picture is an infrared image.

Hyperspectral imaging is the process of scanning and displaying an image within a section of the electromagnetic spectrum. To

Who're interested in HSI?

BATTLEFIELD INTELLIGENCE

Hyperspectral sensor lets drones see through camouflage, spot explosives

BY JOEY CHENG 2014

The Air Force is planning to test a high-powered spectral sensor for unmanned aerial vehicles capable of spotting such things on the ground as improvised explosives or camouflaged targets by identifying what those objects are made of.

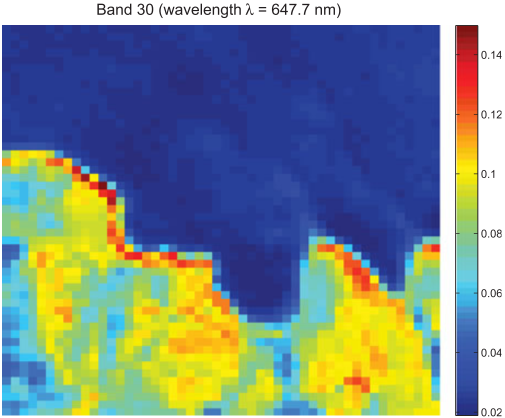
The Air Force Life Cycle Management Center has **announced plans** to negotiate a contract with Raytheon Co. to test a podded version of the Airborne Cueing and Exploitation System-Hyperspectral (ACES-HY) on the Predator UAV.

Other Application Domains

1. many applications in
   - Geology
   - outer space exploration
   - agriculture/forest inspection
   - mine detection
   - food/medicine security
   - ...

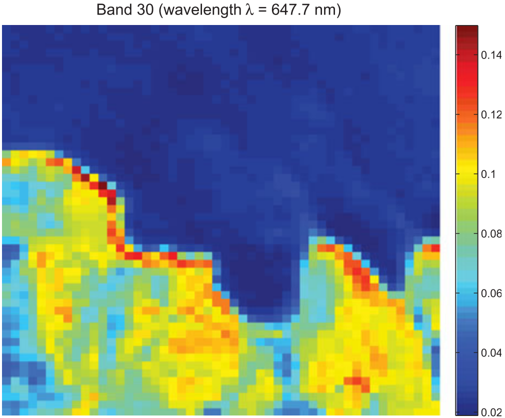
What We Do Not Like ...

1. What we do not like about hyperspectral images is ...



What We Do Not Like ...

1. What we do not like about hyperspectral images is ...



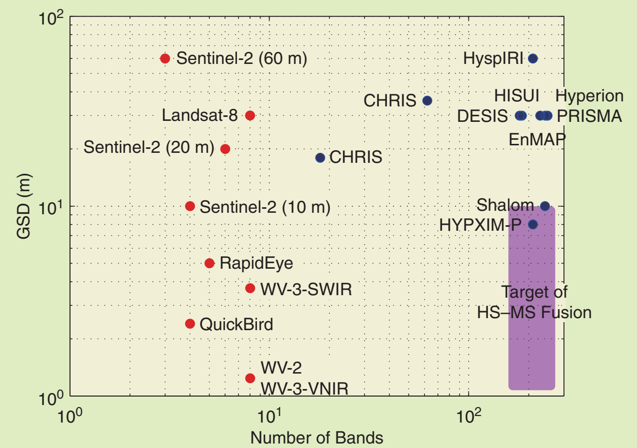
2. The spatial resolution is really NOT eye-pleasing.

Spatial-Spectral Resolution Trade-Off

1. Measuring hyperspectral pixels is a very complicated process and many factors play role [Akgun *et al.* '05]:
    - optics
    - EM reflection mechanism
    - hardware limitations, e.g., sampling strategy and sensor dynamic range.
    - ...
2. directly improving the sensors could be very costly.
3. the spectral and spatial resolutions pose an (inevitable) trade-off in sensor manufacturing [Yokoya *et al.*, '17].

1. The so-called **multispectral images** (MSIs) have very good spatial resolution, but every pixel is only measured at several bands (single digits).

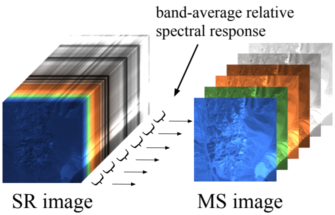
Typical Hyperspectral/Multispectral Sensors

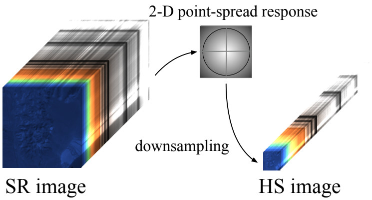


† [Yokoya *et al.*, '17]; GSD: ground sampling distance

The Fusion Problem

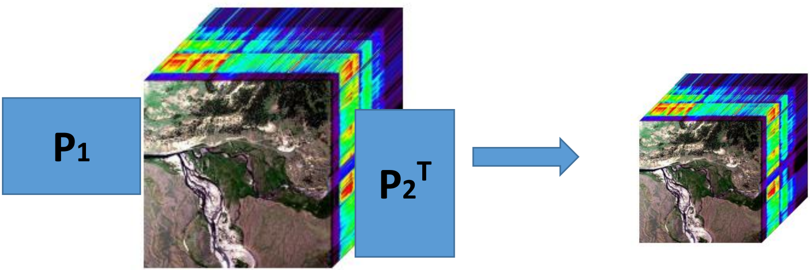
1. Natural idea: how about fusing HSI and MSI?
2. Consider *co-registered* HSI $\underline{\boldsymbol{Y}}_H \in \mathbb{R}^{I_H \times J_H \times K_H}$ and MSI $\underline{\boldsymbol{Y}}_M \in \mathbb{R}^{I_M \times J_M \times K_M}$.
3. Note that we have $I_H J_H \ll I_M J_M$ and $K_H \gg K_M$.
4. Assume there is a super-resolution image (SRI) $\underline{\boldsymbol{Y}}_S \in \mathbb{R}^{I_M \times J_M \times K_H}$, and the MSI and HSI are degraded from the SRI.





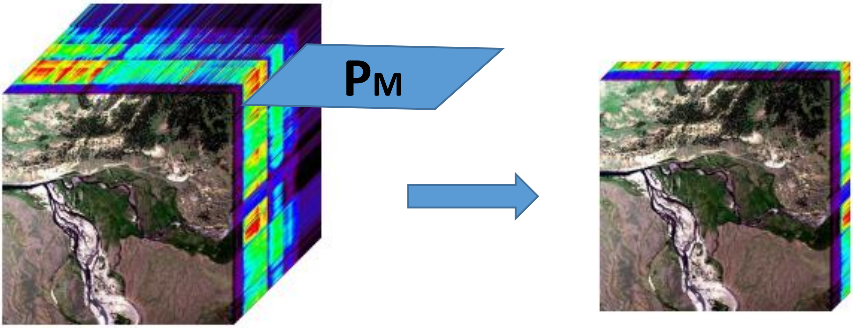
Spatial Degradation

1. Illustration of spatial degradation (SRI$\Longrightarrow$HSI): 2-D blurring and downsampling.

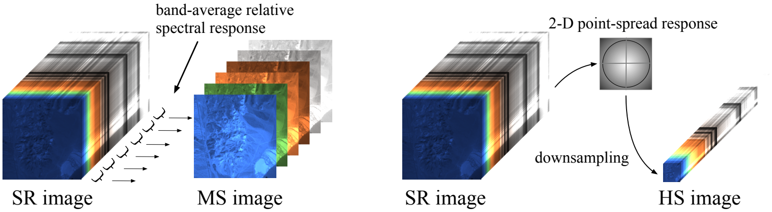


Spectral Degradation

1. Illustration of spectral degradation (SRI$\Longrightarrow$MSI): spectral aggregation.



Problem Statement



1. **The hyperspectral Super-Resolution (HSR) problem**: given the HSI and MSI, recover the SRI that has the spatial resolution of the MSI and the spectral resolution of the HSI.

Prior Art

1. Much effort has been invested to this research area:
   - early methods based on **component substitution** [Carper *et al.* '90]; cumbersome and unreliable.
   - **multi-resolution analysis** [Vivone *et al.* '14]; advanced version of component substitution.
   - state-of-art: **coupled matrix factorization** [Wei *et al.* '17, Simoes *et al.* '15, Yokoya *at al* '12, Wei *et al.* '15]

The LMM

1. The linear mixture model (LMM) of hperspectral pixels:

$$\boldsymbol{y}_\ell \approx \sum_{r=1}^{R} \boldsymbol{a}_r s_{r,\ell} = \boldsymbol{A}_H \boldsymbol{s}_\ell,\ \ \ell = 1, \ldots, L_H (= I_H J_H)$$

where $\boldsymbol{A}_H = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R] \in \mathbb{R}^{K_H \times R}$ is the **spectral signature matrix** and $\boldsymbol{s}_\ell = [s_{1,\ell}, \ldots, s_{R,\ell}]^\top$ is the **abundance vector**.

The LMM

1. The linear mixture model (LMM) of hperspectral pixels:

$$\boldsymbol{y}_\ell \approx \sum_{r=1}^{R} \boldsymbol{a}_r s_{r,\ell} = \boldsymbol{A}_H \boldsymbol{s}_\ell,\ \ \ell = 1, \ldots, L_H (= I_H J_H)$$

   where $\boldsymbol{A}_H = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R] \in \mathbb{R}^{K_H \times R}$ is the **spectral signature matrix** and $\boldsymbol{s}_\ell = [s_{1,\ell}, \ldots, s_{R,\ell}]^\top$ is the **abundance vector**.

2. Putting all pixels into a matrix $\boldsymbol{Y}_H = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{L_H}] \in \mathbb{R}^{K_H \times L_H}$:

$$\boldsymbol{Y}_H \approx \boldsymbol{A}_H \boldsymbol{S}_H,$$

   where $\boldsymbol{S}_H = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{L_H}] \in \mathbb{R}^{R \times L_H}$.

The LMM

1. The linear mixture model (LMM) of hperspectral pixels:

$$\boldsymbol{y}_\ell \approx \sum_{r=1}^{R} \boldsymbol{a}_r s_{r,\ell} = \boldsymbol{A}_H \boldsymbol{s}_\ell,\ \ \ell = 1, \ldots, L_H (= I_H J_H)$$

   where $\boldsymbol{A}_H = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R] \in \mathbb{R}^{K_H \times R}$ is the **spectral signature matrix** and $\boldsymbol{s}_\ell = [s_{1,\ell}, \ldots, s_{R,\ell}]^\top$ is the **abundance vector**.

2. Putting all pixels into a matrix $\boldsymbol{Y}_H = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{L_H}] \in \mathbb{R}^{K_H \times L_H}$:

$$\boldsymbol{Y}_H \approx \boldsymbol{A}_H \boldsymbol{S}_H,$$

   where $\boldsymbol{S}_H = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{L_H}] \in \mathbb{R}^{R \times L_H}$.

3. Note that $R \ll \min\{K_H, L_H\}$ and thus the matrix has low rank.

The LMM

1. The linear mixture model (LMM) of hperspectral pixels:

$$\boldsymbol{y}_\ell \approx \sum_{r=1}^{R} \boldsymbol{a}_r s_{r,\ell} = \boldsymbol{A}_H \boldsymbol{s}_\ell,\ \ell = 1, \ldots, L_H (= I_H J_H)$$

   where $\boldsymbol{A}_H = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R] \in \mathbb{R}^{K_H \times R}$ is the **spectral signature matrix** and $\boldsymbol{s}_\ell = [s_{1,\ell}, \ldots, s_{R,\ell}]^\top$ is the **abundance vector**.

2. Putting all pixels into a matrix $\boldsymbol{Y}_H = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{L_H}] \in \mathbb{R}^{K_H \times L_H}$:

$$\boldsymbol{Y}_H \approx \boldsymbol{A}_H \boldsymbol{S}_H,$$

   where $\boldsymbol{S}_H = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{L_H}] \in \mathbb{R}^{R \times L_H}$.

3. Note that $R \ll \min\{K_H, L_H\}$ and thus the matrix has low rank.
4. K_H: number of bands of HSI; $L_H = I_H J_H$: number of pixels of HSI.

1. The MSI admits a similar model

$$\boldsymbol{Y}_M \approx \boldsymbol{A}_M \boldsymbol{S}_M, \ \in \mathbb{R}^{K_M \times L_M}$$

2. This time R can be larger than $\min\{K_M, L_M\}$.
3. K_M: number of bands of MSI; $L_M = I_M J_M$: number of pixels of MSI.
4. We also have a virtual matricized version of the SRI:

$$\boldsymbol{Y}_S = \boldsymbol{A}_H \boldsymbol{S}_M \in \mathbb{R}^{K_H \times I_M J_M}$$

Matrix Factorization-Based HSR: Main Idea

1. In the matrix form, the degradation model can be written as

$$\boldsymbol{Y}_H = \boldsymbol{Y}_S \boldsymbol{P}_H, \quad \boldsymbol{Y}_M = \boldsymbol{P}_M \boldsymbol{Y}_S,$$

where $\boldsymbol{P}_H = \boldsymbol{P}_1 \otimes \boldsymbol{P}_2$, and $\otimes$ denotes the Kronecker product.

Matrix Factorization-Based HSR: Main Idea

1. In the matrix form, the degradation model can be written as

$$\boldsymbol{Y}_H = \boldsymbol{Y}_S \boldsymbol{P}_H, \quad \boldsymbol{Y}_M = \boldsymbol{P}_M \boldsymbol{Y}_S,$$

where $\boldsymbol{P}_H = \boldsymbol{P}_1 \otimes \boldsymbol{P}_2$, and $\otimes$ denotes the Kronecker product.

2. One possible formulation:

$$\underset{\boldsymbol{Y}_S}{\text{minimize}} \ \|\boldsymbol{Y}_H - \boldsymbol{Y}_S \boldsymbol{P}_H\|_F^2 + \|\boldsymbol{Y}_M - \boldsymbol{P}_M \boldsymbol{Y}_S\|_F^2$$

Matrix Factorization-Based HSR: Main Idea

1. In the matrix form, the degradation model can be written as

$$\boldsymbol{Y}_H = \boldsymbol{Y}_S \boldsymbol{P}_H, \quad \boldsymbol{Y}_M = \boldsymbol{P}_M \boldsymbol{Y}_S,$$

   where $\boldsymbol{P}_H = \boldsymbol{P}_1 \otimes \boldsymbol{P}_2$, and $\otimes$ denotes the Kronecker product.

2. One possible formulation:

$$\underset{\boldsymbol{Y}_S}{\text{minimize}} \ \ \|\boldsymbol{Y}_H - \boldsymbol{Y}_S \boldsymbol{P}_H\|_F^2 + \|\boldsymbol{Y}_M - \boldsymbol{P}_M \boldsymbol{Y}_S\|_F^2$$

3. What's the problem here?

Matrix Factorization-Based HSR: Main Idea

1. The coupled matrix factorization idea [Wei *et al.* '17, Simoes *et al.* '15, Yokoya *at al* '12, Wei *et al.* '15]:

$$\underset{\boldsymbol{A}_H,\boldsymbol{S}_M}{\text{minimize}} \quad \left\| \boldsymbol{Y}_H - \underbrace{\boldsymbol{A}_H \boldsymbol{S}_M}_{\boldsymbol{Y}_S} \boldsymbol{P}_H \right\|_F^2 + \left\| \boldsymbol{Y}_M - \boldsymbol{P}_M \underbrace{\boldsymbol{A}_H \boldsymbol{S}_M}_{\boldsymbol{Y}_S} \right\|_F^2$$

Matrix Factorization-Based HSR: Main Idea

1. The coupled matrix factorization idea [Wei *et al.* '17, Simoes *et al.* '15, Yokoya *at al* '12, Wei *et al.* '15]:

$$\underset{\boldsymbol{A}_H,\boldsymbol{S}_M}{\text{minimize}} \quad \Big\| \boldsymbol{Y}_H - \underbrace{\boldsymbol{A}_H \boldsymbol{S}_M}_{\boldsymbol{Y}_S} \boldsymbol{P}_H \Big\|_F^2 + \Big\| \boldsymbol{Y}_M - \boldsymbol{P}_M \underbrace{\boldsymbol{A}_H \boldsymbol{S}_M}_{\boldsymbol{Y}_S} \Big\|_F^2$$

2. was to recover $I_M J_M K_H$ unknowns from $I_H J_H K_H + I_M J_M K_M$ equations; becomes recovering $(I_M J_M + K_H)R$ unknowns after low-rank parametrization.

Matrix Factorization-Based HSR: Main Idea

1. The coupled matrix factorization idea [Wei *et al.* '17, Simoes *et al.* '15, Yokoya *at al* '12, Wei *et al.* '15]:

$$\underset{\boldsymbol{A}_H, \boldsymbol{S}_M}{\text{minimize}} \quad \left\| \boldsymbol{Y}_H - \underbrace{\boldsymbol{A}_H \boldsymbol{S}_M}_{\boldsymbol{Y}_S} \boldsymbol{P}_H \right\|_F^2 + \left\| \boldsymbol{Y}_M - \boldsymbol{P}_M \underbrace{\boldsymbol{A}_H \boldsymbol{S}_M}_{\boldsymbol{Y}_S} \right\|_F^2$$

2. was to recover $I_M J_M K_H$ unknowns from $I_H J_H K_H + I_M J_M K_M$ equations; becomes recovering $(I_M J_M + K_H)R$ unknowns after low-rank parametrization.
3. the workhorse in this area; a large amount of variants exist.

Matrix Factorization-Based HSR: Main Idea

1. The coupled matrix factorization idea [Wei *et al.* '17, Simoes *et al.* '15, Yokoya *at al* '12, Wei *et al.* '15]:

$$\underset{\boldsymbol{A}_H,\boldsymbol{S}_M}{\text{minimize}} \quad \left\| \boldsymbol{Y}_H - \underbrace{\boldsymbol{A}_H \boldsymbol{S}_M}_{\boldsymbol{Y}_S} \boldsymbol{P}_H \right\|_F^2 + \left\| \boldsymbol{Y}_M - \boldsymbol{P}_M \underbrace{\boldsymbol{A}_H \boldsymbol{S}_M}_{\boldsymbol{Y}_S} \right\|_F^2$$

2. was to recover $I_M J_M K_H$ unknowns from $I_H J_H K_H + I_M J_M K_M$ equations; becomes recovering $(I_M J_M + K_H)R$ unknowns after low-rank parametrization.
3. the workhorse in this area; a large amount of variants exist.
4. makes sense: a special case of **low-rank matrix sensing**.

Matrix Factorization-Based HSR: Main Idea

1. The coupled matrix factorization idea [Wei *et al.* '17, Simoes *et al.* '15, Yokoya *at al* '12, Wei *et al.* '15]:

$$\underset{\boldsymbol{A}_H,\boldsymbol{S}_M}{\text{minimize}} \quad \left\|\boldsymbol{Y}_H - \underbrace{\boldsymbol{A}_H\boldsymbol{S}_M}_{\boldsymbol{Y}_S}\boldsymbol{P}_H\right\|_F^2 + \left\|\boldsymbol{Y}_M - \boldsymbol{P}_M\underbrace{\boldsymbol{A}_H\boldsymbol{S}_M}_{\boldsymbol{Y}_S}\right\|_F^2$$

2. was to recover $I_M J_M K_H$ unknowns from $I_H J_H K_H + I_M J_M K_M$ equations; becomes recovering $(I_M J_M + K_H)R$ unknowns after low-rank parametrization.
3. the workhorse in this area; a large amount of variants exist.
4. makes sense: a special case of **low-rank matrix sensing**.
5. works to a certain extent, but many issues exist.

Issues

1. Fundamentally, the matrix factorization based super-resolution is an inverse problem:

$$\boldsymbol{Y}_H \approx (\boldsymbol{A}_H \boldsymbol{S}_M)\boldsymbol{P}_H, \;\; \boldsymbol{Y}_M \approx \boldsymbol{P}_M(\boldsymbol{A}_H \boldsymbol{S}_M),$$

   where $\boldsymbol{P}_H$ and $\boldsymbol{P}_M$ are **structured** compressing matrices.
2. There is no guarantee that $\boldsymbol{Y}_S$ can be found this way.
3. Performance heavily relies on initialization, prior information, and regularization, and it varies significantly from case to case.
4. has to assume knowledge of $\boldsymbol{P}_H$ and $\boldsymbol{P}_M$—hardly available in practice.

Proposed Approach

1. Our approach [Kanatsoulis, Fu, Sidiropoulos and Ma '18]:

Proposed Approach

1. Our approach [Kanatsoulis, Fu, Sidiropoulos and Ma '18]:
   - **tensor-based**: no matricization.

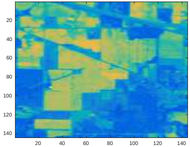
Proposed Approach

1. Our approach [Kanatsoulis, Fu, Sidiropoulos and Ma '18]:
   - **tensor-based**: no matricization.
   - **identifiability-guaranteed**: $\underline{\boldsymbol{Y}}_S$ is provably identifiable under certain conditions.

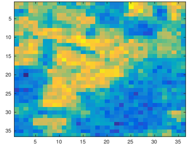
Proposed Approach

1. Our approach [Kanatsoulis, Fu, Sidiropoulos and Ma '18]:
   - **tensor-based**: no matricization.
   - **identifiability-guaranteed**: $\underline{\boldsymbol{Y}}_S$ is provably identifiable under certain conditions.
   - **(semi-)blind**: no need to know the spatial degradation operator.

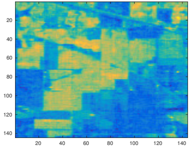
Sneak Peek



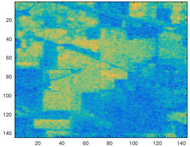
(a) SRI



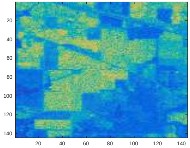
(b) HSI



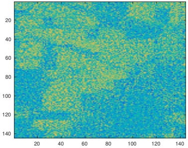
(c) STEREO



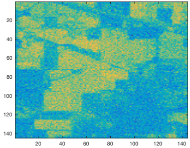
(d) HySure



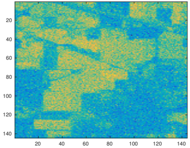
(e) CNMF



(f) FUMI



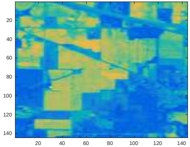
(g) FUSE



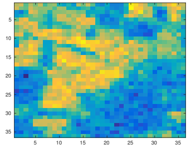
(h) FUSE-SPARSE

1. The proposed method (STEREO) looks very promising.

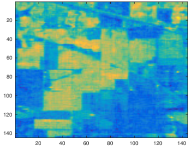
Sneak Peek



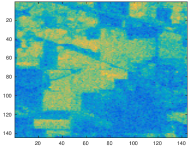
(a) SRI



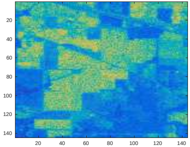
(b) HSI



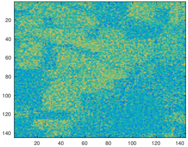
(c) STEREO



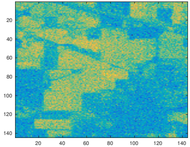
(d) HySure



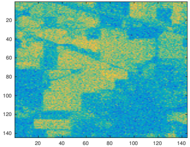
(e) CNMF



(f) FUMI



(g) FUSE



(h) FUSE-SPARSE

1. The proposed method (STEREO) looks very promising.
2. But what is a tensor?

Tensor Preliminaries

1. A tensor $\underline{\boldsymbol{X}}$ is a multi-way array whose elements are indexed by $(i, j, k, \ell, \ldots)$, i.e., more than two indices.
2. A third-order tensor $\underline{\boldsymbol{X}} \in \mathbb{R}^{I \times J \times K}$ is a "shoe box":

3. Tensors have many similarities of matrices but also **striking differences**.

Tensor Preliminaries

1. **matrix rank**: the minimum number R such that

$$\boldsymbol{X} = \sum_{r=1}^{R} \boldsymbol{A}(:,r) \circ \boldsymbol{B}(:,r) = \boldsymbol{A}\boldsymbol{B}^{\top},$$

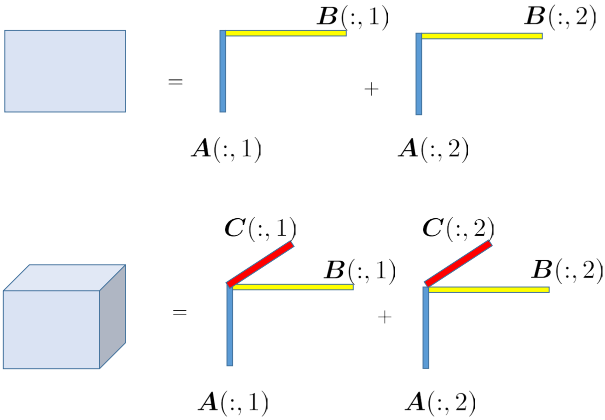
where $\circ$ denotes the outer product, $\boldsymbol{A} \in \mathbb{R}^{I \times R}$, $\boldsymbol{B} \in \mathbb{R}^{J \times R}$.

2. **tensor rank**: the minimum number R such that

$$\underline{\boldsymbol{X}} = \sum_{r=1}^{R} \boldsymbol{A}(:,r) \circ \boldsymbol{B}(:,r) \circ \boldsymbol{C}(:,r) = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!]$$

where $\boldsymbol{A} \in \mathbb{R}^{I \times R}$, $\boldsymbol{B} \in \mathbb{R}^{J \times R}$, and $\boldsymbol{C} \in \mathbb{R}^{K \times R}$.

Tensor Preliminaries: Rank Decomposition



Tensor Preliminaries

1. for matrices:
   - The rank decomposition is very nonunique: $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{B}^\top = \boldsymbol{A}\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{B}^\top$.
   - $R \leq \min\{I, J\}$: the rank has to be $\leq$ the outer dimensions.
2. for tensors:
   - The rank decomposition is **essentially unique** under mild conditions.
   - R can largely exceed I, J, K.

Tensor Preliminaries - Uniqueness

1. For example:

Theorem (Chiantini *et al.* 2012)

Let $\underline{\boldsymbol{X}} = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!]$ with $\boldsymbol{A} : I \times R$, $\boldsymbol{B} : J \times R$, and $\boldsymbol{C} : K \times R$. Assume that $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are drawn from from a jointly continuous distribution (over $\mathbb{R}^{(I+J+K)R}$). Also assume $I \geq J \geq K$ without loss of generality. If $R \leq \frac{1}{16}JK$, then the decomposition of $\underline{\boldsymbol{X}}$ in terms of $\boldsymbol{A}, \boldsymbol{B}$, and $\boldsymbol{C}$ is essentially unique, almost surely.

2. **essential uniqueness** means that if $\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{C}}$ also satisfy $\underline{\boldsymbol{X}} = [\![\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \tilde{\boldsymbol{C}}]\!]$, we can only have $\boldsymbol{A} = \tilde{\boldsymbol{A}}\boldsymbol{\Pi}\boldsymbol{\Lambda}_1$, $\boldsymbol{B} = \tilde{\boldsymbol{B}}\boldsymbol{\Pi}\boldsymbol{\Lambda}_2$, and $\boldsymbol{C} = \tilde{\boldsymbol{C}}\boldsymbol{\Pi}\boldsymbol{\Lambda}_3$, where $\boldsymbol{\Pi}$ is a permutation matrix and $\boldsymbol{\Lambda}_i$ is a full rank diagonal matrix such that $\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_3 = \boldsymbol{I}$.
3. **How mild the condition is?** $I = J = K = 100$ and $R \leq 625$.
4. bottom line: $R \leq \mathcal{O}(JK)$ [Sidiropoulos, De Lathauwer, Fu *et al.* '17].

1. Two basic operations:
   - **Matricization**:

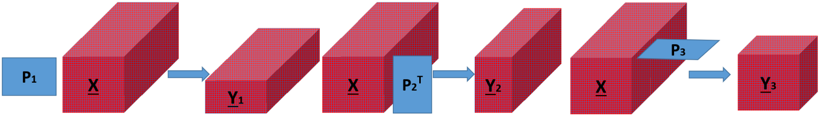
$$\mathbf{X}^{(1)} = (\mathbf{C} \odot \mathbf{B})\mathbf{A}^\top \in \mathbb{R}^{KJ \times I}$$

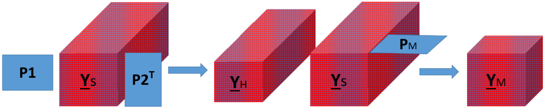
$$\mathbf{X}^{(2)} = (\mathbf{C} \odot \mathbf{A})\mathbf{B}^\top \in \mathbb{R}^{KI \times J}$$

$$\mathbf{X}^{(3)} = (\boldsymbol{B} \odot \boldsymbol{A})\boldsymbol{C}^\top \in \mathbb{R}^{IJ \times K}$$

     where $\odot$ denotes the Khatri-Rao product.
   - **Mode product**:

$$\underline{\tilde{\boldsymbol{X}}} = \underline{\boldsymbol{X}} \times_1 \boldsymbol{P}_1 \times_2 \boldsymbol{P}_2 \times_3 \boldsymbol{P}_3 = [\![\boldsymbol{P}_1\boldsymbol{A}, \boldsymbol{P}_2\boldsymbol{B}, \boldsymbol{P}_3\boldsymbol{C}]\!]$$





1. Let $\underline{\boldsymbol{Y}}_H \in \mathbb{R}^{I_H \times J_H \times K_H}$ denote the HSI cube and $\underline{\boldsymbol{Y}}_M \in \mathbb{R}^{I_M \times J_M \times K_M}$ the MSI cube. $(I_H J_H \ll I_M J_M, K_M \ll K_H)$
2. SRI⇒MSI: $\underline{\boldsymbol{Y}}_M = \underline{\boldsymbol{Y}}_S \times_3 \boldsymbol{P}_M$.
3. SRI⇒HSI: $\underline{\boldsymbol{Y}}_H = \underline{\boldsymbol{Y}}_S \times_1 \boldsymbol{P}_1 \times_2 \boldsymbol{P}_2$.
4. Let $\underline{\boldsymbol{Y}}_S = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!]$ for some unknown $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and rank R.

$$\underline{\boldsymbol{Y}}_M = \underline{\boldsymbol{Y}}_S \times_3 \boldsymbol{P}_M = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M \boldsymbol{C}]\!]$$
$$\underline{\boldsymbol{Y}}_H = \underline{\boldsymbol{Y}}_S \times_1 \boldsymbol{P}_1 \times_2 \boldsymbol{P}_2 = [\![\boldsymbol{P}_1 \boldsymbol{A}, \boldsymbol{P}_2 \boldsymbol{B}, \boldsymbol{C}]\!]$$

1. Major insight:

1. Major insight:
   1. If $\underline{\boldsymbol{Y}}_M = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M\boldsymbol{C}]\!]$ and $\underline{\boldsymbol{Y}}_H = [\![\boldsymbol{P}_1\boldsymbol{A}, \boldsymbol{P}_2\boldsymbol{B}, \boldsymbol{C}]\!]$ are unique decompositions, then $\boldsymbol{A}$ and $\boldsymbol{B}$ can be identified from $\underline{\boldsymbol{Y}}_M$ and $\boldsymbol{C}$ can be identified from $\underline{\boldsymbol{Y}}_H$, respectively.

1. Major insight:
   1. If $\underline{\boldsymbol{Y}}_M = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M\boldsymbol{C}]\!]$ and $\underline{\boldsymbol{Y}}_H = [\![\boldsymbol{P}_1\boldsymbol{A}, \boldsymbol{P}_2\boldsymbol{B}, \boldsymbol{C}]\!]$ are unique decompositions, then $\boldsymbol{A}$ and $\boldsymbol{B}$ can be identified from $\underline{\boldsymbol{Y}}_M$ and $\boldsymbol{C}$ can be identified from $\underline{\boldsymbol{Y}}_H$, respectively.
   2. The "compressed tensors" admit unique rank decomposition under mild conditions.

1. Major insight:
   1. If $\underline{\boldsymbol{Y}}_M = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M\boldsymbol{C}]\!]$ and $\underline{\boldsymbol{Y}}_H = [\![\boldsymbol{P}_1\boldsymbol{A}, \boldsymbol{P}_2\boldsymbol{B}, \boldsymbol{C}]\!]$ are unique decompositions, then $\boldsymbol{A}$ and $\boldsymbol{B}$ can be identified from $\underline{\boldsymbol{Y}}_M$ and $\boldsymbol{C}$ can be identified from $\underline{\boldsymbol{Y}}_H$, respectively.
   2. The "compressed tensors" admit unique rank decomposition under mild conditions.
   3. If $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are not known, it does not hurt the identifiability.

1. Major insight:
    1. If $\underline{\boldsymbol{Y}}_M = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M\boldsymbol{C}]\!]$ and $\underline{\boldsymbol{Y}}_H = [\![\boldsymbol{P}_1\boldsymbol{A}, \boldsymbol{P}_2\boldsymbol{B}, \boldsymbol{C}]\!]$ are unique decompositions, then $\boldsymbol{A}$ and $\boldsymbol{B}$ can be identified from $\underline{\boldsymbol{Y}}_M$ and $\boldsymbol{C}$ can be identified from $\underline{\boldsymbol{Y}}_H$, respectively.
    2. The "compressed tensors" admit unique rank decomposition under mild conditions.
    3. If $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are not known, it does not hurt the identifiability.
    4. more good news: if one tensor is identifiable, it is enough to identify $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ under mild conditions, if $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and $\boldsymbol{P}_M$ are known.

Identifiability: Formal Result

1. Consider an estimator of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$:

$$\underset{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}}{\text{minimize}} \ \left\| \underline{\boldsymbol{Y}}_H - [\![\boldsymbol{P}_1\boldsymbol{A}, \boldsymbol{P}_2\boldsymbol{B}, \boldsymbol{C}]\!] \right\|_F^2 + \lambda \left\| \underline{\boldsymbol{Y}}_M - [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M\boldsymbol{C}]\!] \right\|_F^2 ,$$

Theorem (Kanatsoulis, Fu, Sidiropoulos and Ma '18)

Let $\underline{\boldsymbol{Y}}_H = [\![\boldsymbol{P}_1\boldsymbol{A}, \boldsymbol{P}_2\boldsymbol{B}, \boldsymbol{C}]\!]$ and $\underline{\boldsymbol{Y}}_M = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M\boldsymbol{C}]\!]$. Assume without loss of generality that $I_M \geq J_M \geq K_M$. Also assume that $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are drawn from some continuous distribution, that $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and $\boldsymbol{P}_M$ have full rank, and that $(\boldsymbol{A}^\star, \boldsymbol{B}^\star, \boldsymbol{C}^\star)$ is an optimal solution to the above problem. Then, $\hat{\underline{\boldsymbol{Y}}}_S(i,j,k) = \sum_{f=1}^F \boldsymbol{A}^\star(i,f)\boldsymbol{B}^\star(j,f)\boldsymbol{C}^\star(k,f)$ recovers the ground-truth $\underline{\boldsymbol{Y}}_S$ almost surely if $R \leq \min(2^{\lfloor\gamma\rfloor-2}, I_H J_H)$, where $\gamma = \log_2(J_M K_M)$.

Identifiability: The Semi-Blind Case

1. Consider an estimator of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ when the spatial degradation operators are unknown:

$$\underset{\boldsymbol{A},\boldsymbol{B},\tilde{\boldsymbol{A}},\tilde{\boldsymbol{B}},\boldsymbol{C}}{\text{minimize}} \left\| \underline{\boldsymbol{Y}}_H - \left[\!\left[\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \boldsymbol{C} \right]\!\right] \right\|_F^2 + \lambda \left\| \underline{\boldsymbol{Y}}_M - [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M \boldsymbol{C}]\!] \right\|_F^2 .$$

Theorem (Kanatsoulis, Fu, Sidiropoulos and Ma '18)

Let $\underline{\boldsymbol{Y}}_H = [\![\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{B}}, \boldsymbol{C}]\!]$ and $\underline{\boldsymbol{Y}}_M = [\![\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}_M\boldsymbol{C}]\!]$. Assume without loss of generality that $I_H \geq J_H \geq K_H$ and $I_M \geq J_M \geq K_M$. Also assume that $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are drawn from some continuous distribution, that $\boldsymbol{P}_1$, $\boldsymbol{P}_2$ and $\boldsymbol{P}_M$ have full rank, and that $(\tilde{\boldsymbol{A}}^\star, \tilde{\boldsymbol{B}}^\star, \boldsymbol{A}^\star, \boldsymbol{B}^\star, \boldsymbol{C}^\star)$ is an optimal solution to the above. Then, $\hat{\underline{\boldsymbol{Y}}}_S = [\![\boldsymbol{A}^\star, \boldsymbol{B}^\star, \boldsymbol{C}^\star]\!]$ recovers the ground-truth $\underline{\boldsymbol{Y}}_S$ almost surely if $R \leq \min\{2^{\lfloor \gamma_1 \rfloor - 2}, 2^{\lfloor \gamma_2 \rfloor - 2}\}$, where $\gamma_1 = \log_2(J_M K_M)$ and $\gamma_2 = \log_2(J_H K_H)$.

Remarks

1. take-home: if the HSI and MSI are low-rank tensors, we have a decent chance for establishing identifiability of the SRI.

Remarks

1. take-home: if the HSI and MSI are low-rank tensors, we have a decent chance for establishing identifiability of the SRI.
2. and real HSIs and MSIs are indeed with low tensor rank.

Table: The NMSE of using a low-rank tensor model to approximate a subimage of the AVIRIS Cuprite data of size $512 \times 614 \times 187$ (identifiable when $R \leq 5,984$).

rank	300	400	500	600	700	800
NMSE	0.019	0.016	0.0142	0.0131	0.0123	0.0116

Algorithm

1. make use of the matricizations:

$$\boldsymbol{A} \leftarrow \arg\min_{\boldsymbol{A}} \| \boldsymbol{Y}_H^{(1)} - (\mathbf{C} \odot \boldsymbol{P}_2 \boldsymbol{B}) \boldsymbol{A}^\top \boldsymbol{P}_1^\top \|_F^2 + \lambda \| \boldsymbol{Y}_M^{(1)} - (\boldsymbol{P}_M \boldsymbol{C} \odot \boldsymbol{B}) \boldsymbol{A}^\top \|_F^2$$

$$\boldsymbol{B} \leftarrow \arg\min_{\boldsymbol{B}} \| \boldsymbol{Y}_H^{(2)} - (\boldsymbol{C} \odot \boldsymbol{P}_1 \boldsymbol{A}) \boldsymbol{B}^\top \boldsymbol{P}_2^T \|_F^2 + \lambda \| \boldsymbol{Y}_H^{(2)} - (\boldsymbol{P}_M \boldsymbol{C} \odot \boldsymbol{A}) \boldsymbol{B}^\top \|_F^2$$

$$\boldsymbol{C} \leftarrow \arg\min_{\boldsymbol{C}} \| \boldsymbol{Y}_H^{(3)} - (\boldsymbol{P}_2 \boldsymbol{B} \odot \boldsymbol{P}_1 \boldsymbol{A}) \boldsymbol{C}^\top \|_F^2 + \lambda \| \boldsymbol{Y}_M^{(3)} - (\boldsymbol{B} \odot \boldsymbol{A}) \boldsymbol{C}^\top \boldsymbol{P}_M^\top \|_F^2;$$

Algorithm

1. make use of the matricizations:

$$\boldsymbol{A} \leftarrow \arg\min_{\boldsymbol{A}} \| \boldsymbol{Y}_H^{(1)} - (\mathbf{C} \odot \boldsymbol{P}_2 \boldsymbol{B}) \boldsymbol{A}^\top \boldsymbol{P}_1^\top \|_F^2 + \lambda \| \boldsymbol{Y}_M^{(1)} - (\boldsymbol{P}_M \boldsymbol{C} \odot \boldsymbol{B}) \boldsymbol{A}^\top \|_F^2$$

$$\boldsymbol{B} \leftarrow \arg\min_{\boldsymbol{B}} \| \boldsymbol{Y}_H^{(2)} - (\boldsymbol{C} \odot \boldsymbol{P}_1 \boldsymbol{A}) \boldsymbol{B}^\top \boldsymbol{P}_2^T \|_F^2 + \lambda \| \boldsymbol{Y}_H^{(2)} - (\boldsymbol{P}_M \boldsymbol{C} \odot \boldsymbol{A}) \boldsymbol{B}^\top \|_F^2$$

$$\boldsymbol{C} \leftarrow \arg\min_{\boldsymbol{C}} \| \boldsymbol{Y}_H^{(3)} - (\boldsymbol{P}_2 \boldsymbol{B} \odot \boldsymbol{P}_1 \boldsymbol{A}) \boldsymbol{C}^\top \|_F^2 + \lambda \| \boldsymbol{Y}_M^{(3)} - (\boldsymbol{B} \odot \boldsymbol{A}) \boldsymbol{C}^\top \boldsymbol{P}_M^\top \|_F^2;$$

2. cyclically update until convergence.

Algorithm

1. make use of the matricizations:

$$\boldsymbol{A} \leftarrow \arg\min_{\boldsymbol{A}} \|\boldsymbol{Y}_H^{(1)} - (\mathbf{C} \odot \boldsymbol{P}_2\boldsymbol{B})\boldsymbol{A}^\top \boldsymbol{P}_1^\top\|_F^2 + \lambda \|\boldsymbol{Y}_M^{(1)} - (\boldsymbol{P}_M\boldsymbol{C} \odot \boldsymbol{B})\boldsymbol{A}^\top\|_F^2$$

$$\boldsymbol{B} \leftarrow \arg\min_{\boldsymbol{B}} \|\boldsymbol{Y}_H^{(2)} - (\boldsymbol{C} \odot \boldsymbol{P}_1\boldsymbol{A})\boldsymbol{B}^\top \boldsymbol{P}_2^T\|_F^2 + \lambda \|\boldsymbol{Y}_H^{(2)} - (\boldsymbol{P}_M\boldsymbol{C} \odot \boldsymbol{A})\boldsymbol{B}^\top\|_F^2$$

$$\boldsymbol{C} \leftarrow \arg\min_{\boldsymbol{C}} \|\boldsymbol{Y}_H^{(3)} - (\boldsymbol{P}_2\boldsymbol{B} \odot \boldsymbol{P}_1\boldsymbol{A})\boldsymbol{C}^\top\|_F^2 + \lambda \|\boldsymbol{Y}_M^{(3)} - (\boldsymbol{B} \odot \boldsymbol{A})\boldsymbol{C}^\top \boldsymbol{P}_M^\top\|_F^2;$$

2. cyclically update until convergence.
3. every subproblem is an unconstrained quadratic program (least squares) and can be solved in closed-form.

Algorithm

1. make use of the matricizations:

$$\boldsymbol{A} \leftarrow \arg\min_{\boldsymbol{A}} \|\boldsymbol{Y}_H^{(1)} - (\mathbf{C} \odot \boldsymbol{P}_2\boldsymbol{B})\boldsymbol{A}^\top \boldsymbol{P}_1^\top\|_F^2 + \lambda\|\boldsymbol{Y}_M^{(1)} - (\boldsymbol{P}_M\boldsymbol{C} \odot \boldsymbol{B})\boldsymbol{A}^\top\|_F^2$$

$$\boldsymbol{B} \leftarrow \arg\min_{\boldsymbol{B}} \|\boldsymbol{Y}_H^{(2)} - (\boldsymbol{C} \odot \boldsymbol{P}_1\boldsymbol{A})\boldsymbol{B}^\top \boldsymbol{P}_2^T\|_F^2 + \lambda\|\boldsymbol{Y}_H^{(2)} - (\boldsymbol{P}_M\boldsymbol{C} \odot \boldsymbol{A})\boldsymbol{B}^\top\|_F^2$$

$$\boldsymbol{C} \leftarrow \arg\min_{\boldsymbol{C}} \|\boldsymbol{Y}_H^{(3)} - (\boldsymbol{P}_2\boldsymbol{B} \odot \boldsymbol{P}_1\boldsymbol{A})\boldsymbol{C}^\top\|_F^2 + \lambda\|\boldsymbol{Y}_M^{(3)} - (\boldsymbol{B} \odot \boldsymbol{A})\boldsymbol{C}^\top \boldsymbol{P}_M^\top\|_F^2;$$

2. cyclically update until convergence.
3. every subproblem is an unconstrained quadratic program (least squares) and can be solved in closed-form.
4. the '(semi-)blind case' can be solved in a similar fashion, with two more blocks (w.r.t. $\tilde{\boldsymbol{A}}$ and $\tilde{\boldsymbol{B}}$).

Remarks

1. the first identifiability-guaranteed HSR approach.
2. identifiability holds under mild conditions.
3. does not need the knowledge of the spatial degradation operator.
4. **challenges**:
   - nonconvexity
   - determining the tensor rank (NP-hard in theory)

Experiments - Semi-Real Data

1. Salinas scene was downloaded from the AVIRIS platform. It represents a field, consisting of 6 different agricultural products.
2. $\underline{\boldsymbol{Y}}_S \in \mathbb{R}^{80\times 84\times 204}$ is a subimage of the Salinas scene.
3. The HSI $\underline{\boldsymbol{Y}}_H \in \mathbb{R}^{20\times 21\times 204}$ is generated after blurring by a 9×9 Gaussian Kernel and downsampling by a factor of 4 (choose 1 out of 16 pixels).
4. the MSI $\underline{\boldsymbol{Y}}_M \in \mathbb{R}^{80\times 84\times 6}$ is generated according to specs of the LANDSAT multispectral sensor.
5. $R = 100$.
6. State of Art Baselines: `FUMI` [Wei *et al.*, '17], `HySure` [Simoes *et al.*, '15], `CNMF` [Yokoya *at al*, '12], `FUSE`, `FUSE-Sparse` [Wei *et al.* '15].

Numerical Results

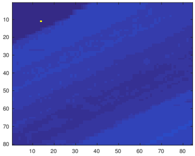
Evaluation metrics: Reconstruction Signal-to-Noise ratio **(R-SNR)**, Cross Correlation **(CC)**, **runtime**.

- **RSNR** $= 10 \log \left(\frac{\sum_{k=1}^{K} \| \underline{\boldsymbol{Y}_S}(:,:,k) \|_F^2}{\sum_{k=1}^{K} \| \underline{\hat{\boldsymbol{Y}}_S}(:,:,k) - \underline{\boldsymbol{Y}_S}(:,:,k) \|_F^2} \right)$
- **CC** $= \sum_{k=1}^{K} \rho \left(\underline{\boldsymbol{Y}_S}(:,:,k), \underline{\hat{\boldsymbol{Y}}_S}(:,:,k) \right)$

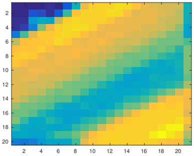
Table: SALINAS scene

Algorithm	**RSNR**	**CC**	**runtime** (sec)
STEREO	**39.39**	**0.9864**	1.5
FUSE	28.71	0.9174	**0.07**
FUSE-Sparse	28.71	0.9173	69.7
FUMI	29.40	0.9126	1.56
HySure	26.86	0.8981	1.6
CNMF	25.48	0.9013	1.7

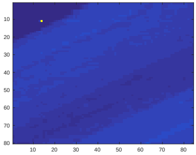
Illustration



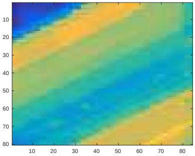
(a) SRI



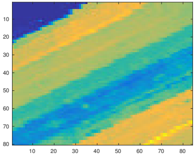
(b) HSI



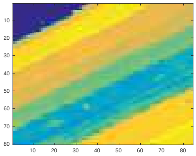
(c) STEREO



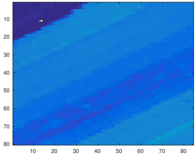
(d) HySure



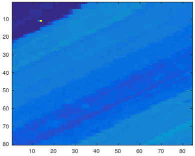
(e) CNMF



(f) FUMI

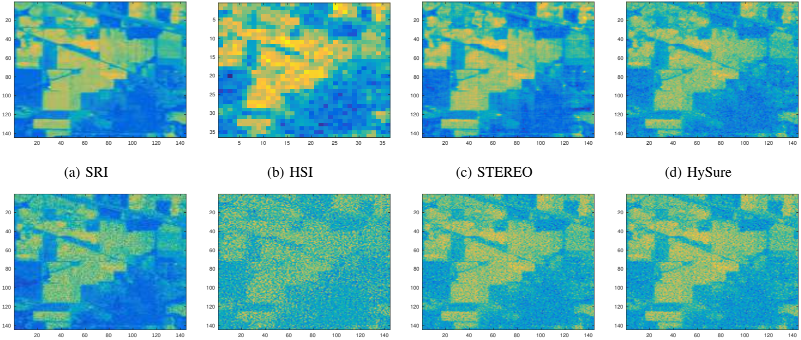


(g) FUSE



(h) FUSE-SPARSE

Blind Reconstruction

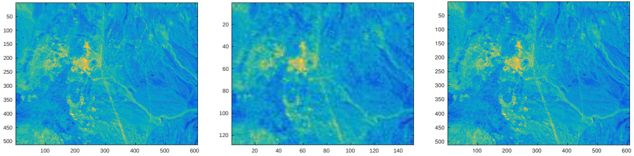


(a) SRI (b) HSI (c) STEREO (d) HySure

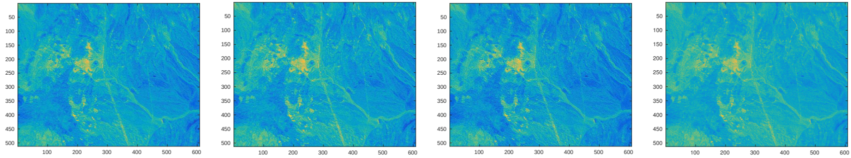
(e) CNMF (f) FUMI (g) FUSE (h) FUSE-SPARSE

† STEREO does not assume knowledge of $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$.

Blind Reconstruction



(a) SRI (b) HSI (c) STEREO



(d) HySure (e) CNMF (f) FUMI (g) FUSE

† STEREO does not assume knowledge of $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$.

Conclusion

1. The problem of hyperspectral super-resolution has been introduced.
2. The idea behind state-of-the-art methods has been introduced.
3. A tensor based approach (and tensor preliminaries) has been introduced
   - which provides the only theoretical identifiability support among existing methods and
   - can work without the spatial degradation info.
4. Take-home: tensor is a powerful tool that has many good properties that can be leveraged in signal processing and machine learning.

1. [Yokoya *et al.*, '17] Naoto Yokoya, Claas Grohnfeldt, and Jocelyn Chanussot, "Hyperspectral and Multispectral Data Fusion: A comparative review of the recent literature" IEEE Geoscience and Remote Sensing Magazine 5.2 (2017): 29-56.
2. [Kanatsoulis, Fu, Sidiropoulos and Ma '18] Charilaos Kanatsoulis, Xiao Fu, Nicholas D. Sidiropoulos, and Wing-Kin Ma, "Tensor-Based Hyperspectral Super Resolution: Identifiability and Algorithm", to appear at IEEE ICASSP 2018, Calgary, Canada, April 2018.
3. [Sidiropoulos, De Lathauwer, Fu *et al.* '17] Nicholas D. Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E. Papalexakis, and Christos Faloutsos. "Tensor decomposition for signal processing and machine learning." IEEE Transactions on Signal Processing 65, no. 13 (2017): 3551-3582.
4. [Wei *et al.* '17] Qi Wei, José Bioucas-Dias, Nicolas Dobigeon, Jean-Yves Tourneret, Marcus Chen, and Simon Godsill, "Multiband image fusion based on spectral unmixing," IEEE Transactions on Geoscience and Remote Sensing 54, no. 12 (2016): 7236-7249.
5. [Wei *et al.* '15] Qi Wei, José Bioucas-Dias, Nicolas Dobigeon, Jean-Yves Tourneret "Hyperspectral and multispectral image fusion based on a sparse representation." IEEE Transactions on Geoscience and Remote Sensing 53, no. 7 (2015): 3658-3668.
6. [Yokoya *et al.*, '12] Naoto, Yokoya, Takehisa Yairi, and Akira Iwasaki. "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion." IEEE Transactions on Geoscience and Remote Sensing 50.2 (2012): 528-537.