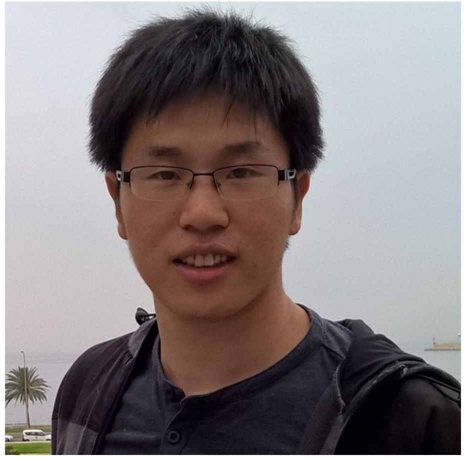
Multi-Modal Image Processing with Applications to Art Investigation and Beyond

Miguel Rodrigues
Dept. Electronic and Electrical Engineering
University College London

Collaborators

Ingrid Daubechies
Duke U.

Bruno Cornellis
VUB



Pingfan Song
UCL

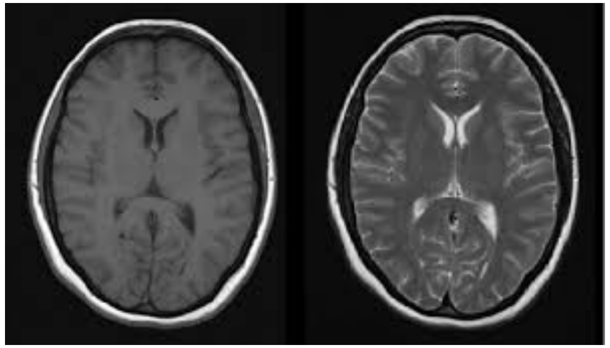
Joao Mota
Heriot Watt U.

Nikos Deligiannis
VUB

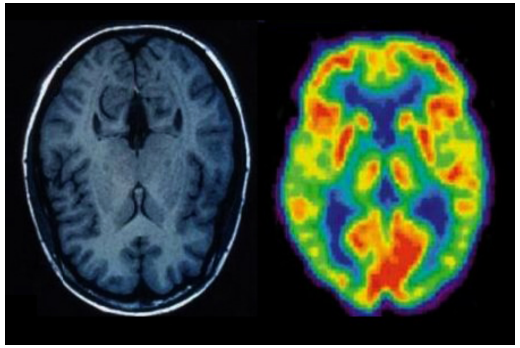
Multi-Modal Data Processing in Healthcare

Medical Imaging

T1 and T2



MRI and PET



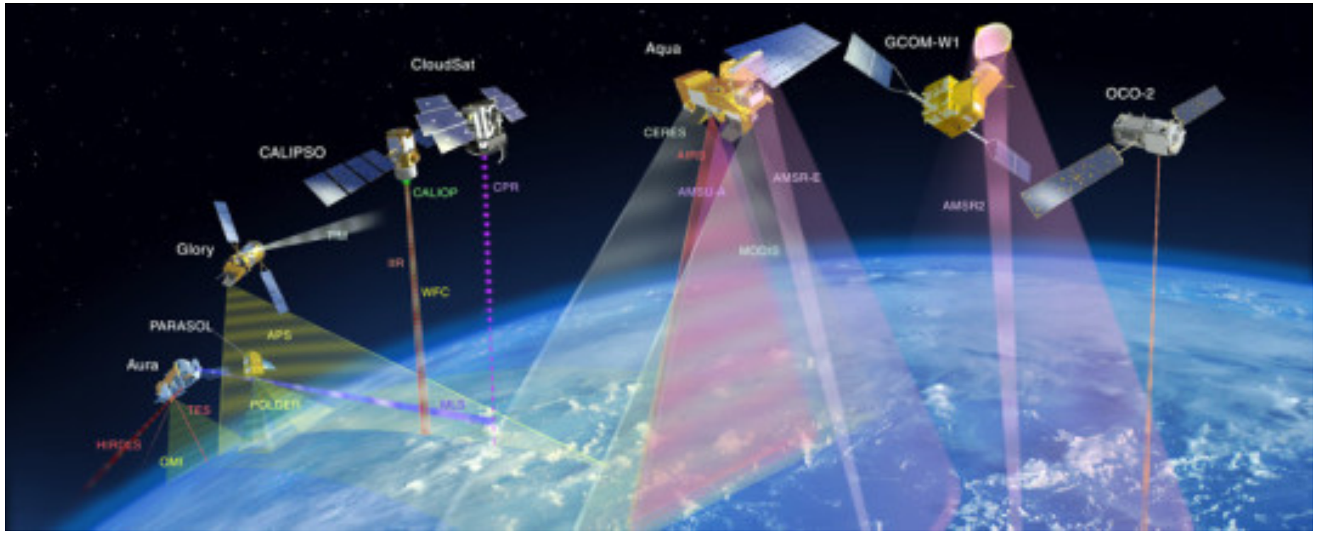
Emerging questions

The questions that arise in medical imaging include:

- How to trade-off acquisition resolution across the various imaging modalities?
- How to analyse multiple complementary image modalities?

Multi-Modal Data Processing in Engineering

Remote Sensing

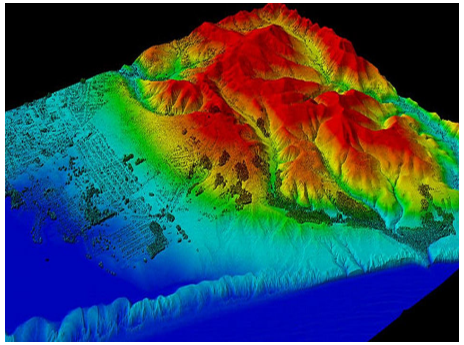


Emerging questions

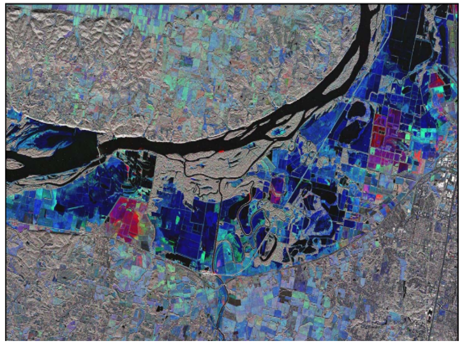
The questions that arise in remote sensing also include:

- How to trade-off acquisition resolution across the various imaging modalities?
- How to analyse multiple complementary image modalities?

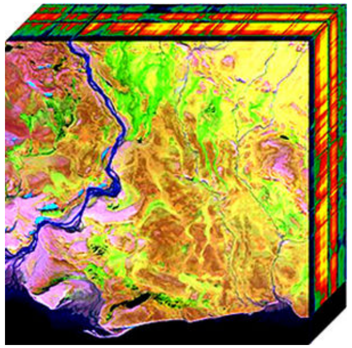
LIDAR Data



SAR Data

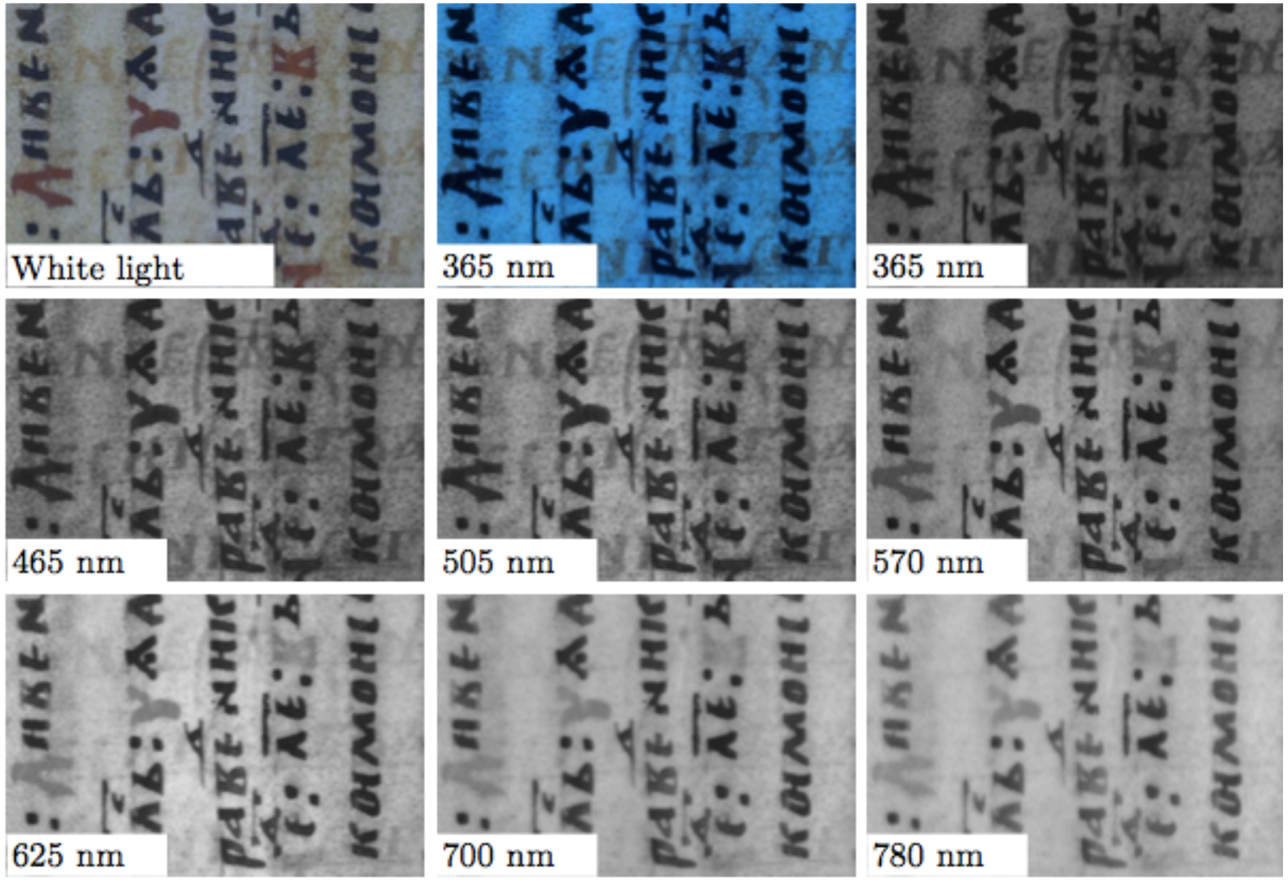


Hyper-Spectral Data



Multi-Modal Data Processing in Arts and Humanities

Palimpsests in Cultural Heritage and Archeology



Palimpsest contains a Cyrillic overwriting and partly Greek, partly Cyrillic underwritings, which have been washed off

Emerging questions

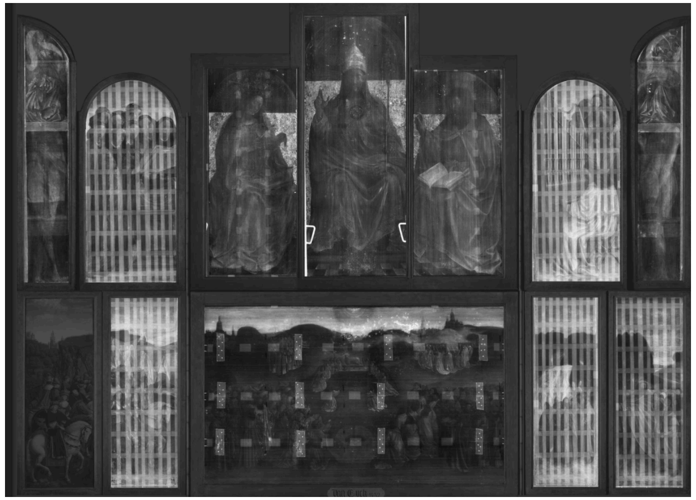
- Common practice in medieval ecclesiastical circles to rub out an earlier piece of writing by means of washing or scraping the manuscript, in order to prepare it for a new text.
- Modern historians are usually more interested in older writings, so multi-modal data processing technology is needed to attempt to recover erased old texts.

Multi-Modal Data Processing in Arts and Humanities

Art Investigation, Preservation and Restoration

The Ghent Alterpiece - Visuals

The Ghent Alterpiece – X-Rays



Emerging questions

Some tasks that arise in art investigation, restoration and preservation include:

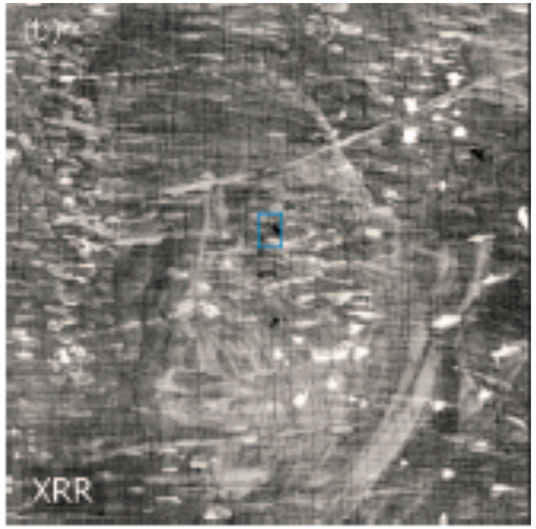
- The separation of paintings onto different layers for technical study purposes.
- The identification of areas associated with degradation / restoration.

The imaging modalities used in art investigation include macrophotography, X-radiography, hyperspectral imaging, infrared imaging, X-ray fluorescence (XRF) mapping

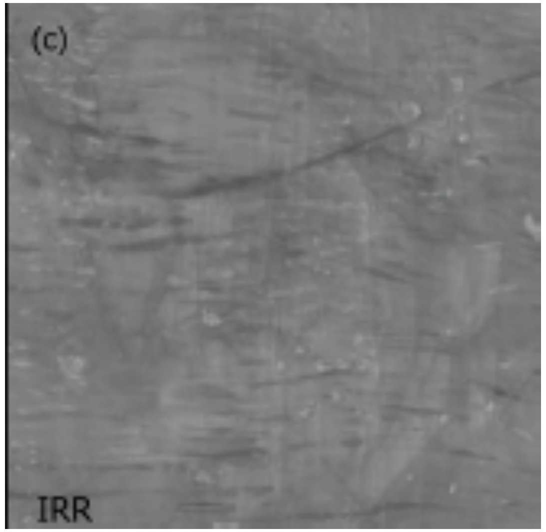
Multi-Modal Data Processing in Arts and Humanities

Vincent van Gogh
Patch of Grass, Paris, Apr-June 1887

X-ray radiation transmission radiograph (XRR)



Infrared reflectograph (IRR)



Dik et al. Visualization of a Lost Painting by Vincent van Gogh Using Synchrotron Radiation Based X-ray Fluorescence Elemental Mapping. Anal. Chem. 2008, 80, 6436–6442

Outline

i. Parsimonious Representations for Unimodal Data Processing

ii. Joint Parsimonious Representations for Multimodal Data Processing

iii. Multimodal Data Aided Processing

   a. Image separation aided by multimodal data

   b. Image super-resolution aided by multimodal data

iv. Concluding Remarks and Directions

Outline

i. **Parsimonious Representations for Unimodal Data Processing**

ii. Joint Parsimonious Representations for Multimodal Data Processing

iii. Multimodal Data Aided Processing

   a. Image separation aided by multimodal data

   b. Image super-resolution aided by multimodal data

iv. Concluding Remarks and Directions

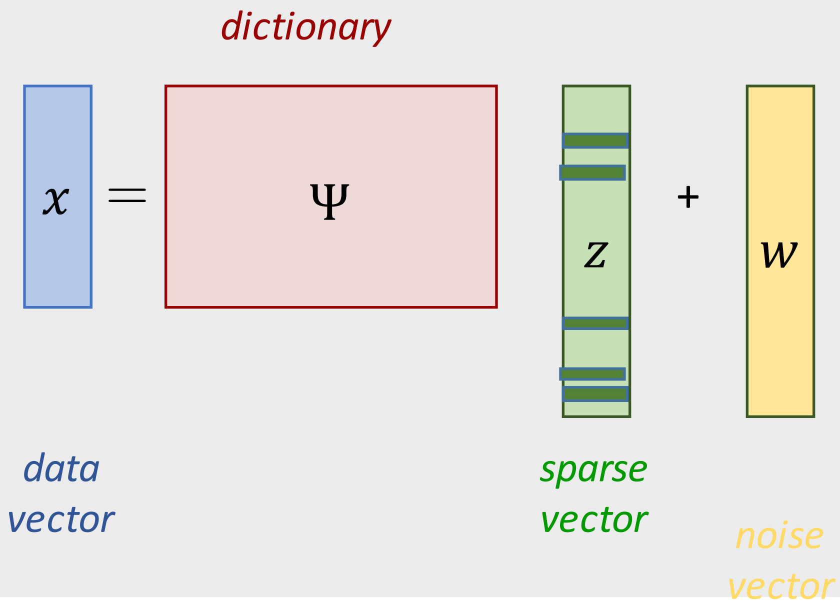
Sparse Representations for Data Processing

Parsimonious representations

The data vector $x \in \mathbb{R}^n$ can be represented in terms of a sparse vector $z \in \mathbb{R}^m$ as follows:

$$x = \Psi z + w$$

where $\Psi \in \mathbb{R}^{n \times m}$ is a dictionary such as a wavelet basis or a learnt one.



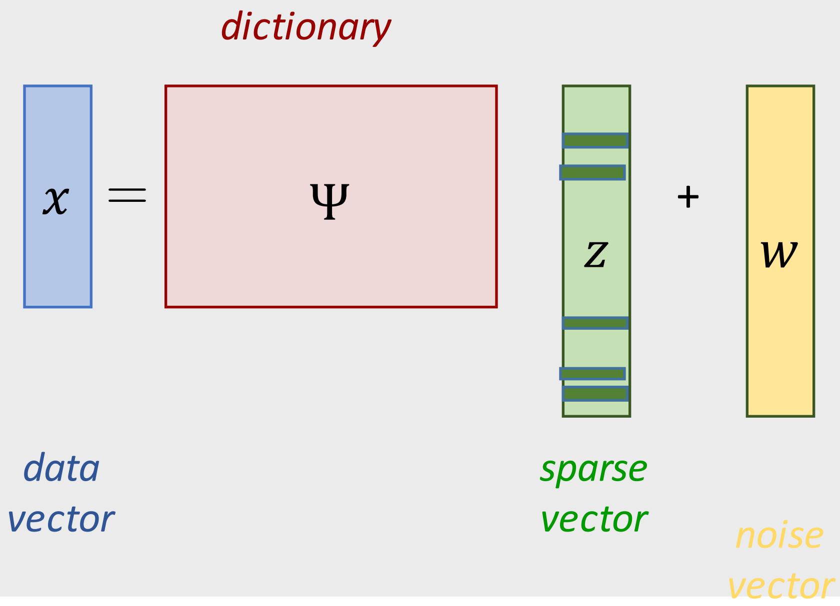
Sparse Representations for Data Processing

Parsimonious representations

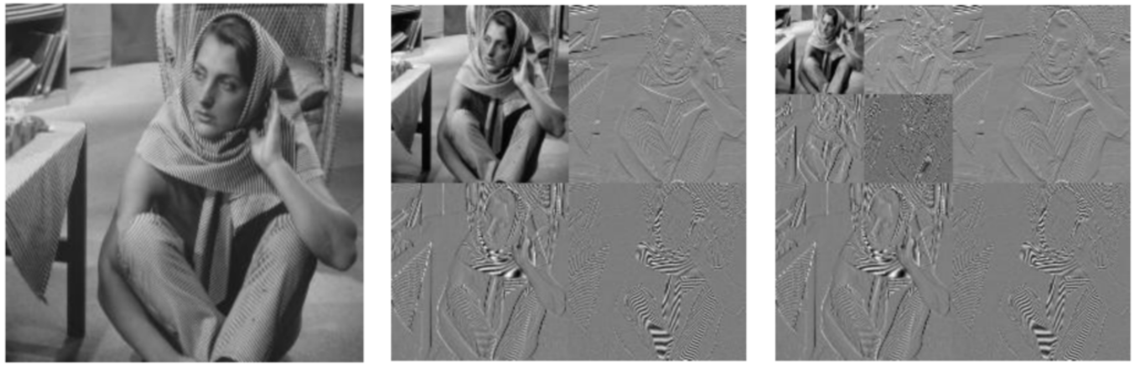
The data vector $x \in \mathbb{R}^n$ can be represented in terms of a sparse vector $z \in \mathbb{R}^m$ as follows:

$$x = \Psi z + w$$

where $\Psi \in \mathbb{R}^{n \times m}$ is a dictionary such as a wavelet basis or a learnt one.



Wavelet representations



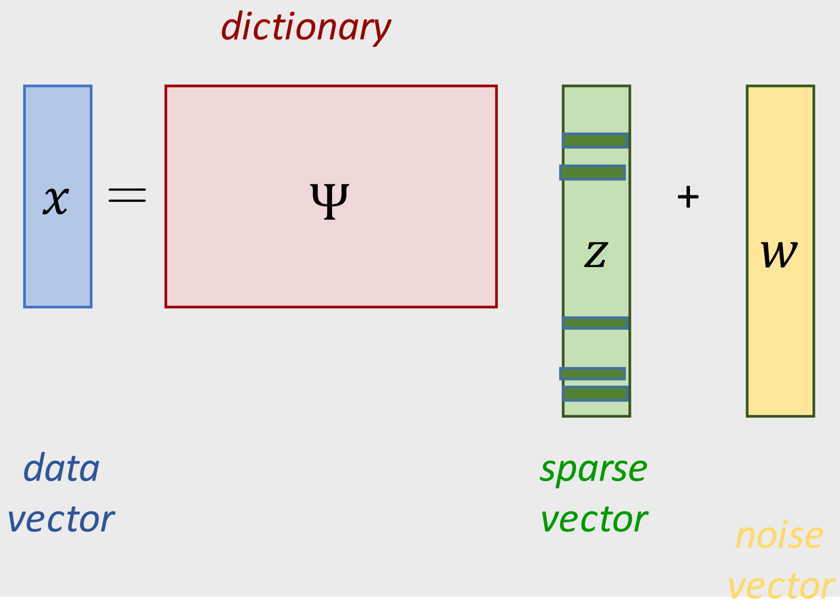
Sparse Representations for Data Processing

Parsimonious representations

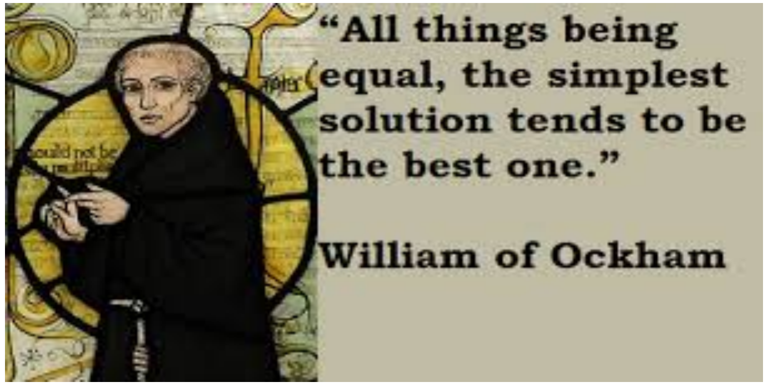
The data vector $x \in \mathbb{R}^n$ can be represented in terms of a sparse vector $z \in \mathbb{R}^m$ as follows:

$$x = \Psi z + w$$

where $\Psi \in \mathbb{R}^{n \times m}$ is a dictionary such as a wavelet basis or a learnt one.



Occam's Razor



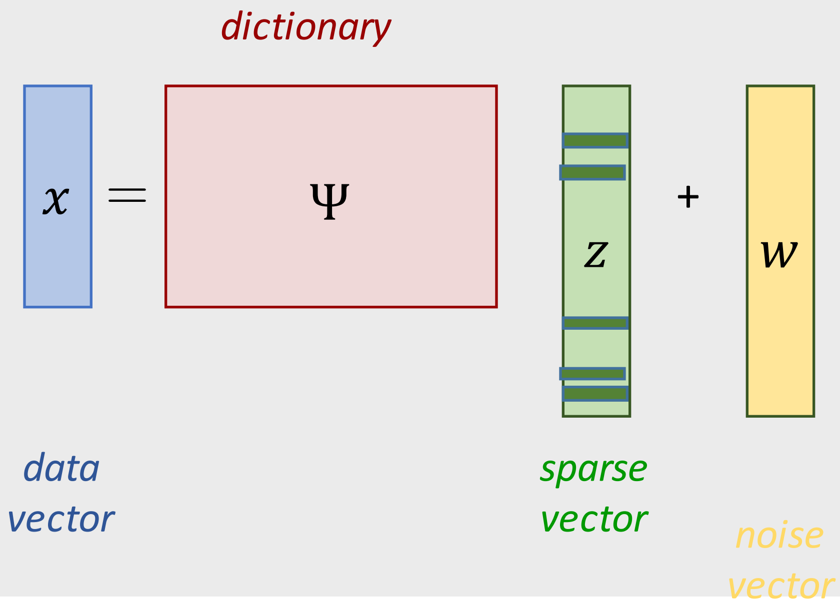
Sparse Representations for Data Processing

Parsimonious representations

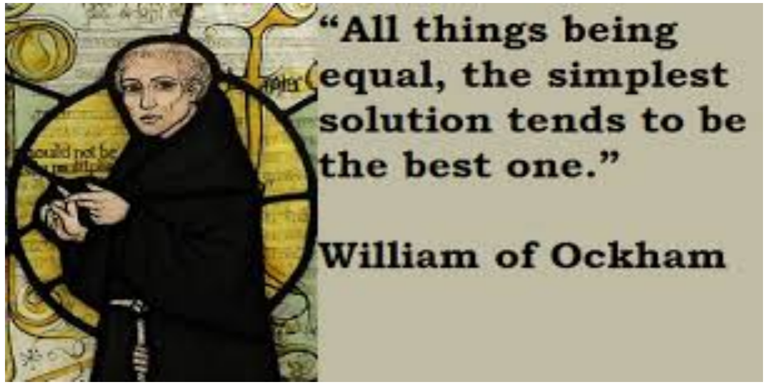
The data vector $x \in \mathbb{R}^n$ can be represented in terms of a sparse vector $z \in \mathbb{R}^m$ as follows:

$$x = \Psi z + w$$

where $\Psi \in \mathbb{R}^{n \times m}$ is a dictionary such as a wavelet basis or a learnt one.



Occam's Razor



Applications

Sparse representations have had implications in various problems such as:

1. Compressive sensing
2. Image in-painting, denoising, debluring
3. Image super-resolution
4. Source separation/de-mixing

The Compressive Sensing Problem

Signal Sensing

The measurement vector is generated from the signal vector as follows:

$$y = \Phi x = \Phi \Psi z$$

where Φ is a "wide" measurement matrix.

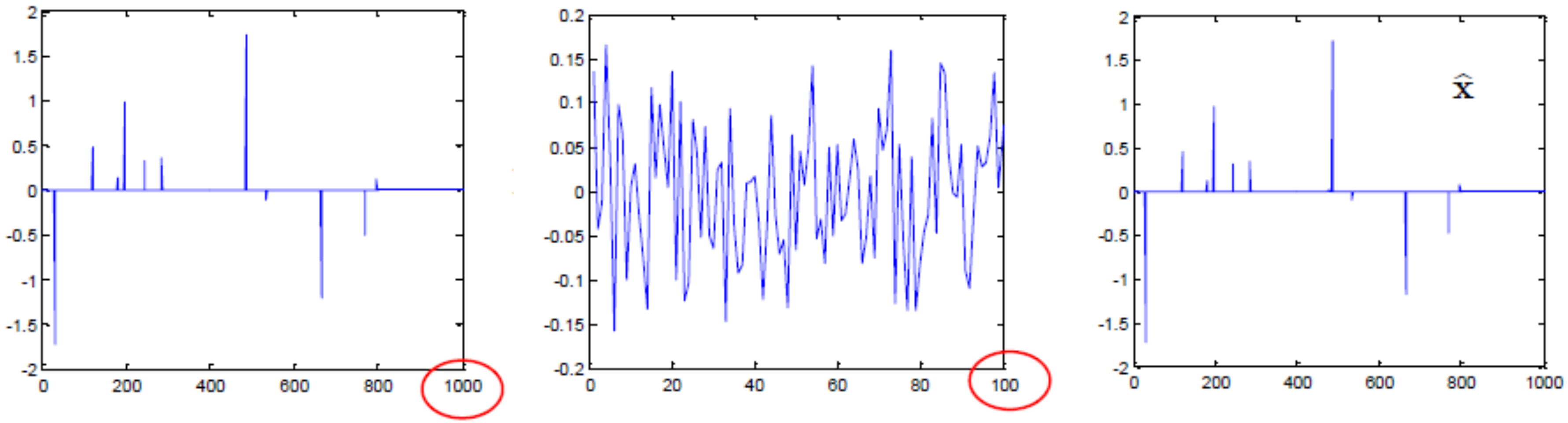
Signal Reconstruction

The signal sparse representation vector can be recovered from the measurement vector as follows:

$$\hat{z} = \arg\min_{z} \|z\|_1 \text{ subject to } y = \Phi \Psi z$$

Optimization- and greedy-based algorithms can be used to reconstruct the signal vector from the measurement vector.

Sparse Vector (z) **Measured Vector (y)** **Recovered Sparse Vector**



The Compressive Sensing Problem:
The Single-Pixel Camera

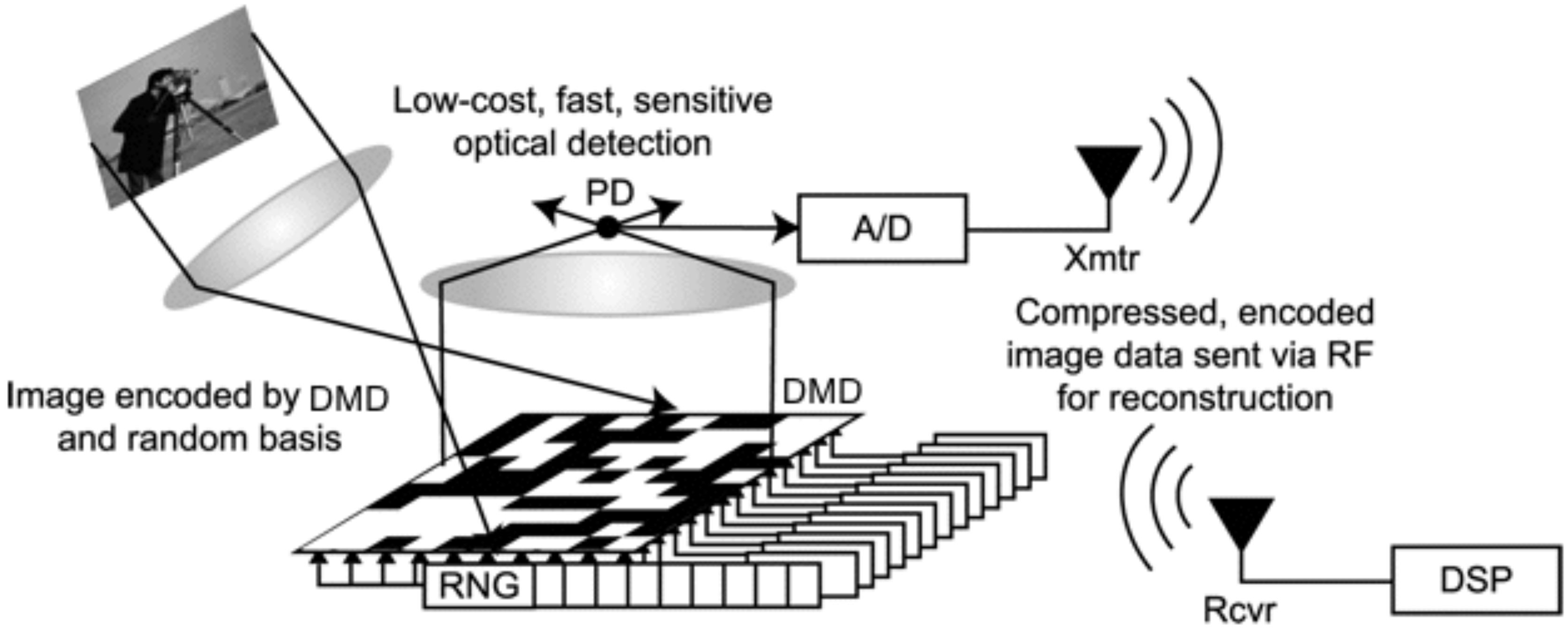


Image De-Noising, De-Blurring and In-Painting

De-Noising

Noisy Image

De-Noised Image

De-Blurring

Blurred Image

De-Blurred Image

In-Painting

Original Image

New Image

Angle-of-Attack

Model

One postulates that the true image admits a sparse representation in some dictionary.

Algorithm

One then obtains the sparse represent. associated with the image as well as the dictionary given the noisy / blurred / in-painted image.

Image De-Noising

De-noising model

One observes a noisy version y_i of image (patches) x_i:

$$y_i = x_i + w_i, \ \forall i$$

The image (patches) x_i obey a sparse representation z_i in a dictionary D:

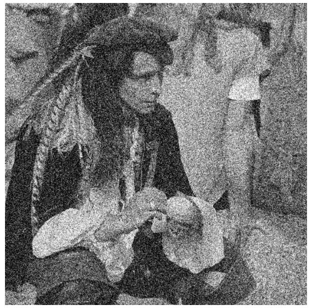
$$x_i = Dz_i, \qquad \forall i$$

Sparse representations based de-noising

This problem can be addressed using sparse representations whereby the de-noised image is generated from the noisy image as follows:

$$\min_{D,z_i} \sum_i \|y_i - Dz_i\|_2^2 + \|z_i\|_1 \quad \Rightarrow \quad \hat{x}_i = D\hat{z}_i, \qquad \forall i$$

Original Noisy Image

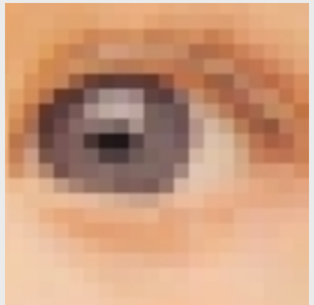



De-noised Image

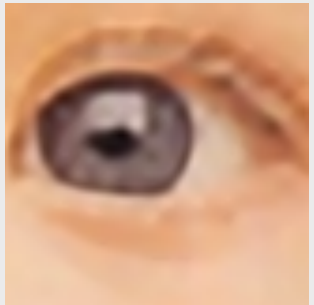

Image Super-Resolution (SR)

Super-Resolution Problem

Low-resolution Image



High-resolution Image



Angle-of-Attack

Model

One postulates that both the HR and the LR images admit a sparse representation in HR and LR dictionaries.

Algorithm

One then obtains the HR image from the LR image by determining the sparse representation associated with the images as well as the HR and LR dictionaries.

Image Super-Resolution

Super-resolution model

One postulates that HR patches x_i^{HR} and LR patches x_i^{LR} admit a common sparse representation z_i in HR and LR dictionaries D^{HR} and D^{LR}:

$$x_i^{HR} = D^{HR} z_i,\ \forall i$$

$$x_i^{LR} = D^{LR} z_i,\ \forall i$$

Sparse representations based super-resolution

This problem can be addressed using sparse representations whereby the HR image is generated from the LR image as follows:

Training:

$$\min_{D^{HR}, D^{LR}, z_i} \sum_i \left\| x_i^{HR} - D^{HR} z_i \right\|_2^2 + \left\| x_i^{LR} - D^{LR} z_i \right\|_2^2 + \lambda \cdot \| z_i \|_1$$

Testing:

$$\hat{z}_i = \underset{z_i}{\operatorname{argmin}} \left\| x_i^{LR} - D^{LR} z_i \right\|_2^2 + \lambda \cdot \| z_i \|_1 \Rightarrow \hat{x}_i^{HR} = D^{HR} \hat{z}_i$$

Low-resolution Image

High-resolution Image

Outline

i. Parsimonious Representations for Unimodal Data Processing

ii. Joint Parsimonious Representations for Multimodal Data Processing

iii. Multimodal Data Aided Processing

   a. Image separation aided by multimodal data

   b. Image super-resolution aided by multimodal data

iv. Concluding Remarks and Directions

Joint Sparse Representations for Multi-Modal Data

Wishlist

1. Model to represent accurately each individual image modality;
2. Model to connect the various image modalities;
3. Model to be readily learnt from data using simple algorithms;
4. Model to lead to simple multi-modal processing algorithms.

Joint Parsimonious Representations

Each individual image modalities admit sparse representations in a dictionary.

The various image modalities are connected via sparse representations.

$$x_1 = \Phi^c \boxed{z^c} + \Phi \boxed{z_1} \quad \textit{data modality 1}$$

$$x_2 = \Psi^c \boxed{z^c} + \Psi \boxed{z_2} \quad \textit{data modality 2}$$

Common Components *Innovation Components*

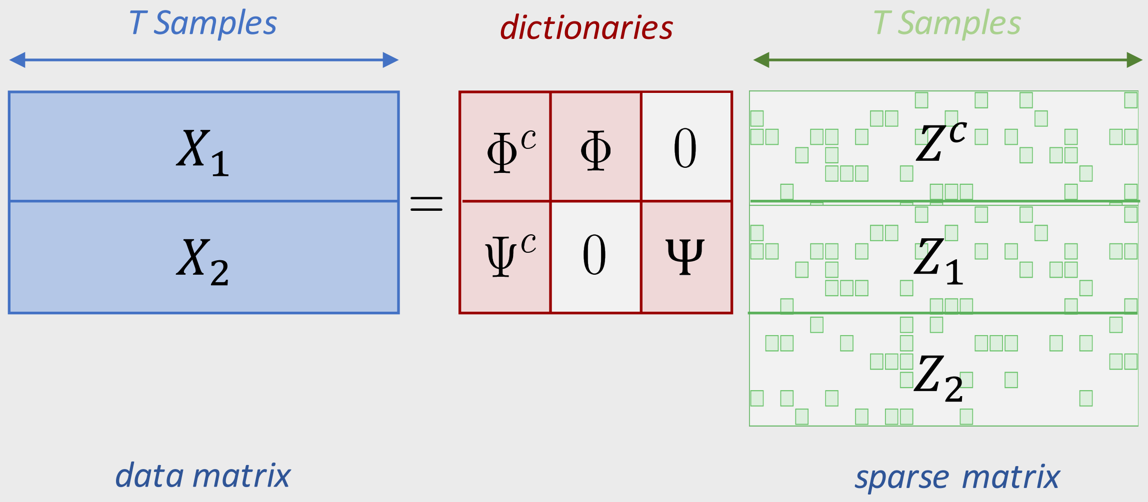
Joint Sparse Representations for Multi-Modal Data

Wishlist

1. Model to represent accurately each individual image modality;
2. Model to connect the various image modalities;
3. Model to be readily learnt from data using simple algorithms;
4. Model to lead to simple multi-modal processing algorithms.

Learning, Analysis and Processing Algorithms

Our model can also be readily learnt using matrix factorization techniques.



Our model also leads to simple multi-modal image processing algorithms that exploit the joint sparse representations.

Joint Sparse Representations for Multi-Modal Data

Wishlist

1. Model to represent accurately each individual image modality;
2. Model to connect the various image modalities;
3. Model to be readily learnt from data using simple algorithms;
4. Model to lead to simple multi-modal processing algorithms.

Coupled Dictionary Learning Algorithm

$$\min_{\substack{\Phi^c,\Phi,\Psi^c,\Psi\\ Z^c,Z_1,Z_2}} \|X_1 - \Phi^c Z^c - \Phi Z_1\|_F^2 + \|X_2 - \Psi^c Z^c - \Psi Z_2\|_F^2$$

$$\text{s.t.}\ card\big(Z^c(i)\big) \le s_c, \mathrm{i} = 1,\dots,T$$

$$card\big(Z_1(i)\big) \le s_1, \mathrm{i} = 1,\dots,T$$

$$card\big(Z_2(i)\big) \le s_2, \mathrm{i} = 1,\dots,T$$

Learn dictionaries by alternating between:

1. Learning the sparse representations given the dictionaries (sparse coding step)
2. Learning the dictionaries given the sparse representations (dictionary update step)

Outline

i. Parsimonious Representations for Unimodal Data Processing

ii. Joint Parsimonious Representations for Multimodal Data Processing

iii. Multimodal Data Aided Processing

  a. Image separation aided by multimodal data

  b. Image super-resolution aided by multimodal data

iv. Concluding Remarks and Directions

Multi-Modal Data Aided Image Separation

Problem

This problem involves separating the super-position of the x-rays given the visuals.

Model$_{coupling}$

$$y = \Psi^c \boxed{z} \qquad \textit{Visual}$$

$$x = \Phi^c \boxed{z} + \Phi v \qquad \textit{X-Ray}$$

Visual Rear Panel

Visual Front Panel

Mixed X-Ray

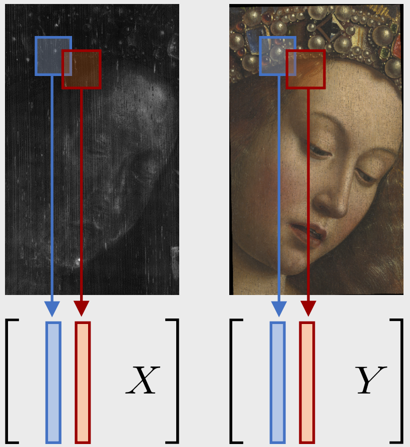
Multi-Modal Data Aided Image Separation

Learning Phase

The goal is to learn the joint parsimonious model from available data.

Algorithm

X-Ray



Visible

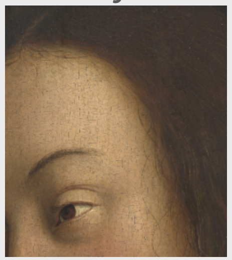
$$
\begin{aligned}
\underset{\substack{\Psi^c,\Phi^c,\Phi\\ Z,V}}{\text{minimize}} \quad & \left\|Y-\Psi^c Z\right\|_F^2+\left\|X-\Phi^c Z-\Phi V\right\|_F^2 \\
\text{subject to} \quad & \operatorname{card}(Z_i)\leq s_z\,, \quad i=1,\ldots,T \\
& \operatorname{card}(V_i)\leq s_v\,, \quad i=1,\ldots,T
\end{aligned}
$$

Processing Phase

The goal is to unmix the x-rays given the x-ray mixture and the visuals.

Algorithm

visual front



$y_1 = \Psi^c z_1$

visual back

$y_2 = \Psi^c z_2$

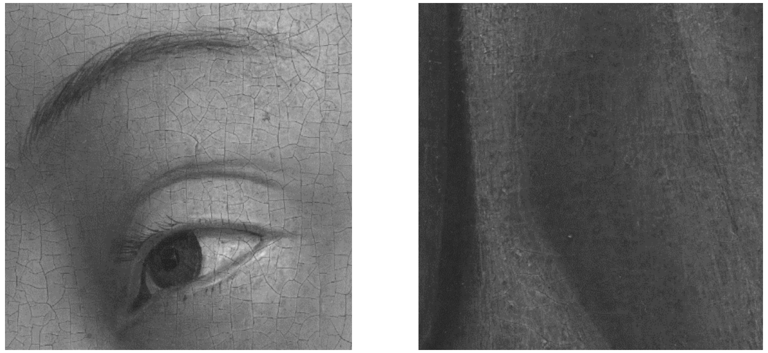
mixed x-ray

$x = \Phi^c(z_1+z_2) + 2\Phi v$

$$
\begin{aligned}
\underset{z_1,z_1,v}{\text{minimize}} \quad & \|z_1\|_1+\|z_2\|_1+\|v\|_1 \\
\text{subject to} \quad & x=\Phi^c(z_1+z_2)+2\Phi v \\
& y_1=\Psi^c z_1 \\
& y_2=\Psi^c z_2
\end{aligned}
$$

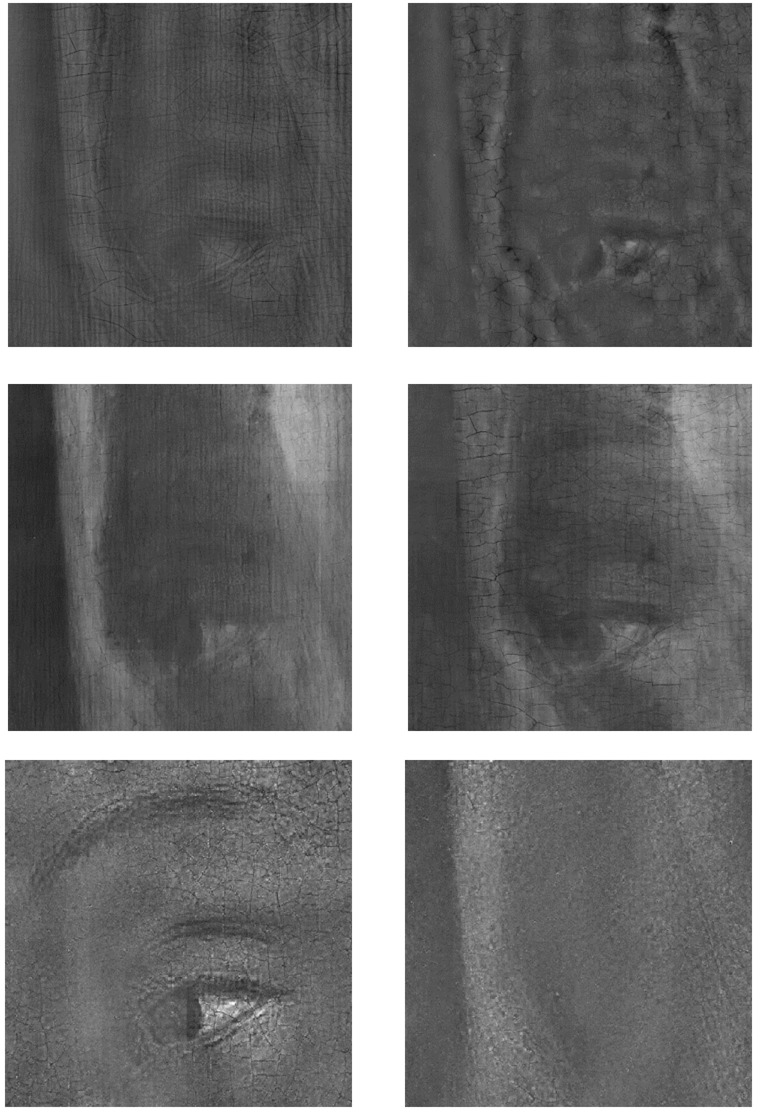
Multi-Modal Data Aided Image Separation

visuals in grayscale



mixed x-ray

reconstructed x-rays

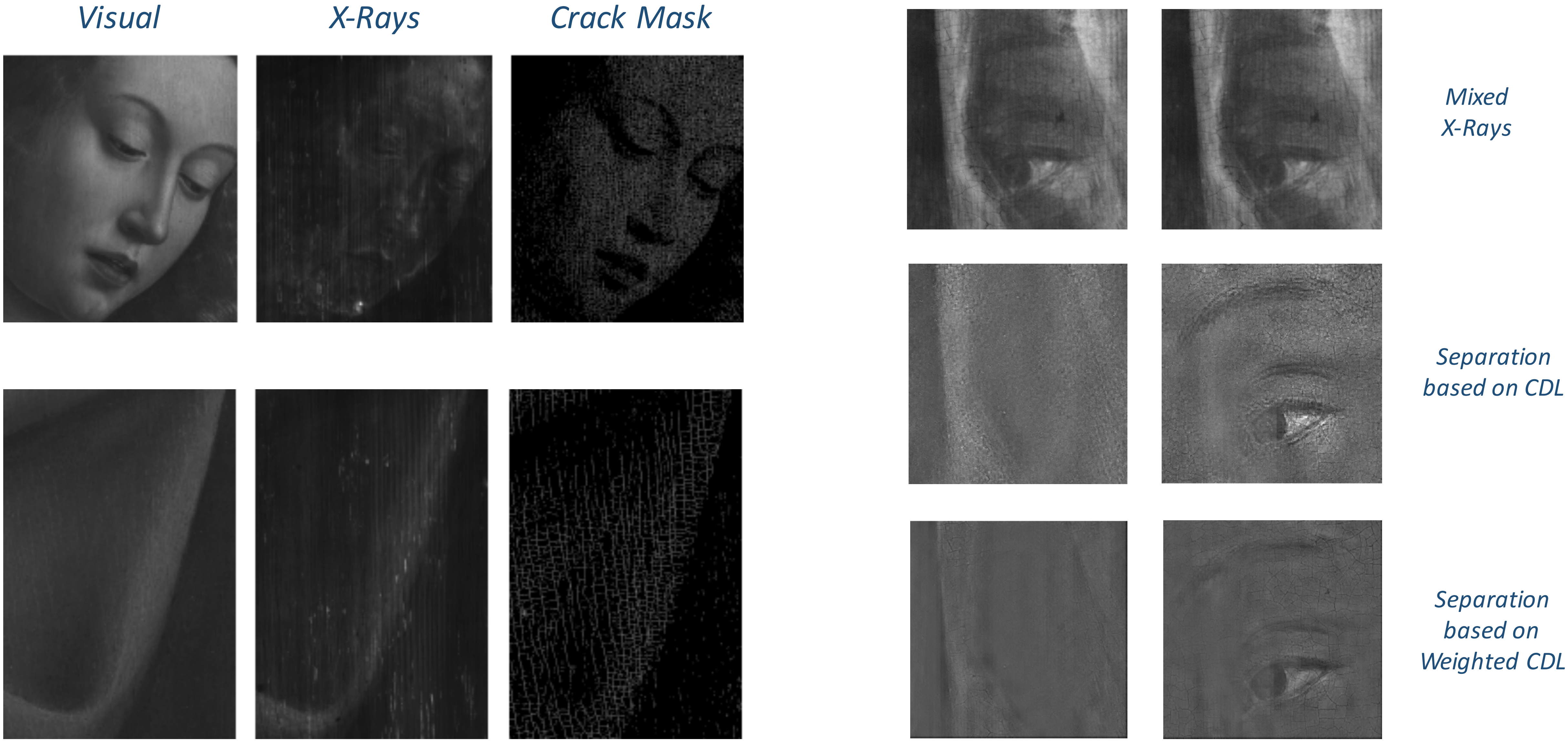


MCA

multiscale MCA w/KSVD

Ours

Multi-Modal Data Aided Image Separation



Outline

i. Parsimonious Representations for Unimodal Data Processing

ii. Joint Parsimonious Representations for Multimodal Data Processing

iii. Multimodal Data Aided Processing

  a. Image separation aided by multimodal data

  b. Image super-resolution aided by multimodal data

iv. Concluding Remarks and Directions

Multi-Modal Data Aided Super-Resolution

Problem

This problem involves producing a high-resolution (HR) image from a low-resolution (LR) one of the same scene, by leveraging the presence of other images associated with the scene.

Model

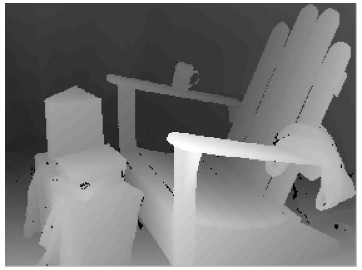
$$x^{hr} = \Psi_c^{hr} z_c + \Psi^{hr} u$$ *HR image of interest*

$$x^{lr} = \Psi_c^{lr} z_c + \Psi^{lr} u$$ *LR image of interest*

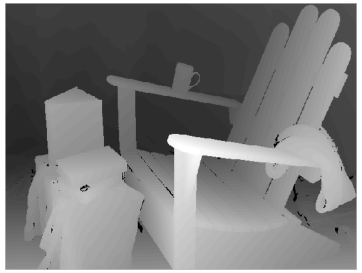
$$y^{hr} = \Phi_c^{hr} z_c + \Phi^{hr} v$$ *another HR image*

coupling between HR and LR image

LR Image



HR Image



HR Side Information

Multi-Modal Data Aided Super-Resolution

Problem

This problem involves producing a high-resolution (HR) image from a low-resolution (LR) one of the same scene, by leveraging the presence of other images associated with the scene.

Model

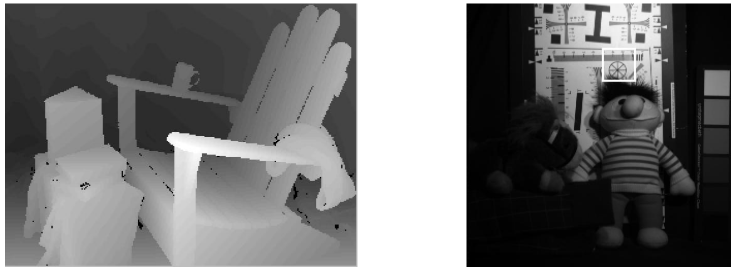
$$x^{hr} = \Psi_c^{hr}\, z_c + \Psi^{hr}\, u$$ *HR image of interest*

$$x^{lr} = \Psi_c^{lr}\, z_c + \Psi^{lr}\, u$$ *LR image of interest*

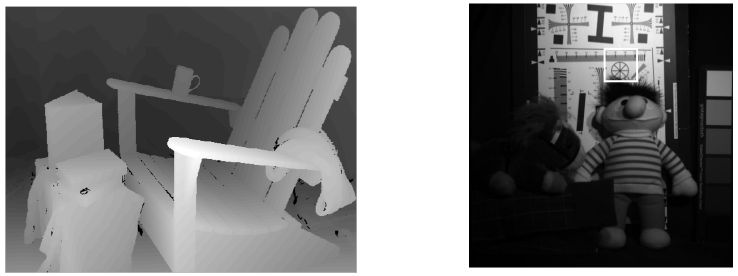
$$y^{hr} = \Phi_c^{hr}\, z_c + \Phi^{hr}\, v$$ *another HR image*

coupling between modalities

LR Image



HR Image



HR Side Information

Multi-Modal Data Aided Super-Resolution

Problem

This problem involves producing a high-resolution (HR) image from a low-resolution (LR) one of the same scene, by leveraging the presence of other images associated with the scene.

Model

$x^{hr} = \Psi_c^{hr} z_c + \Psi^{hr} u$ — *HR image of interest*

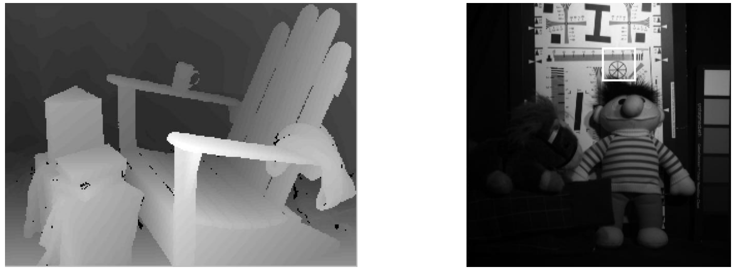
$x^{lr} = \Psi_c^{lr} z_c + \Psi^{lr} u$ — *LR image of interest*

$y^{hr} = \Phi_c^{hr} z_c + \Phi^{hr} v$ — *another HR image*

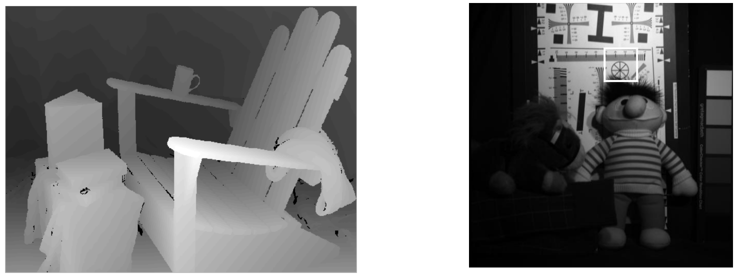
coupling between modalities

Training Phase → **Processing Phase**

LR Image



HR Image

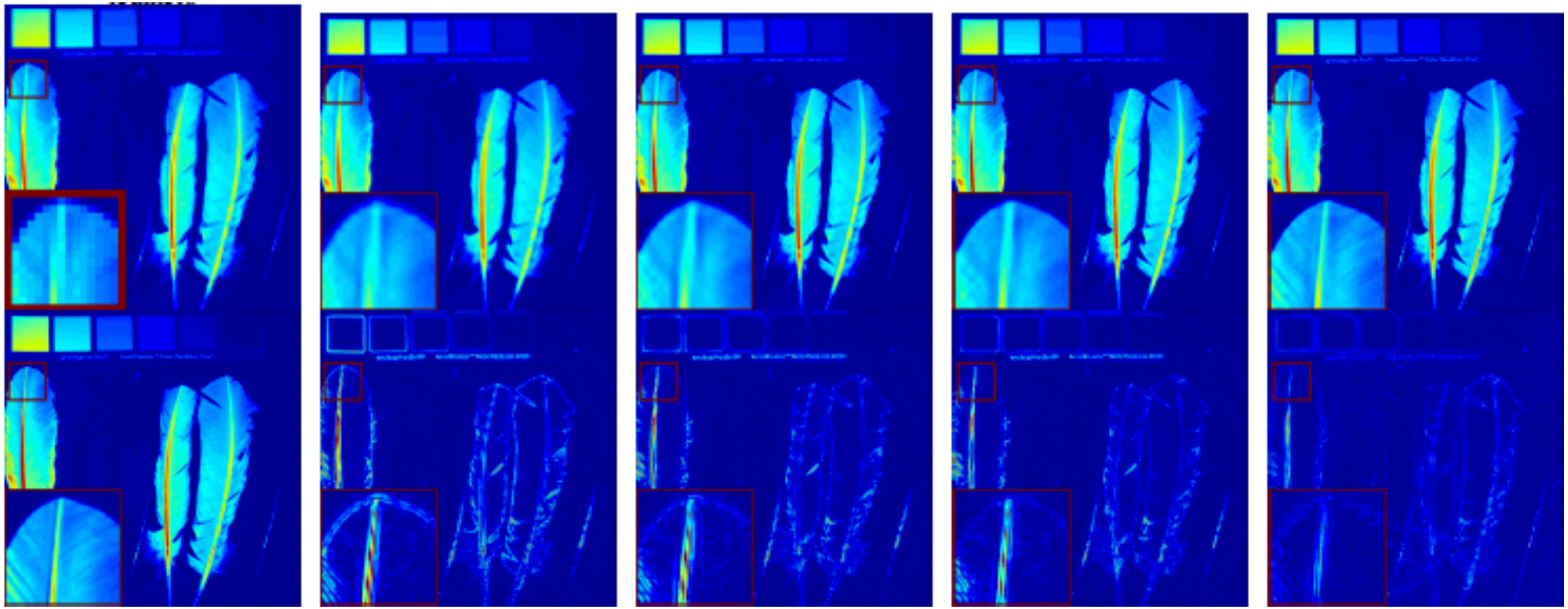


HR Side Information

Multi-Modal Data Aided Super-Resolution

Super-resolving hyper-spectral images with the aid of RGB images

LR-Image | HR-Image - Bicubic | HR-Image – Zeyde et al. | HR-Image – A+ | HR-Image – Ours

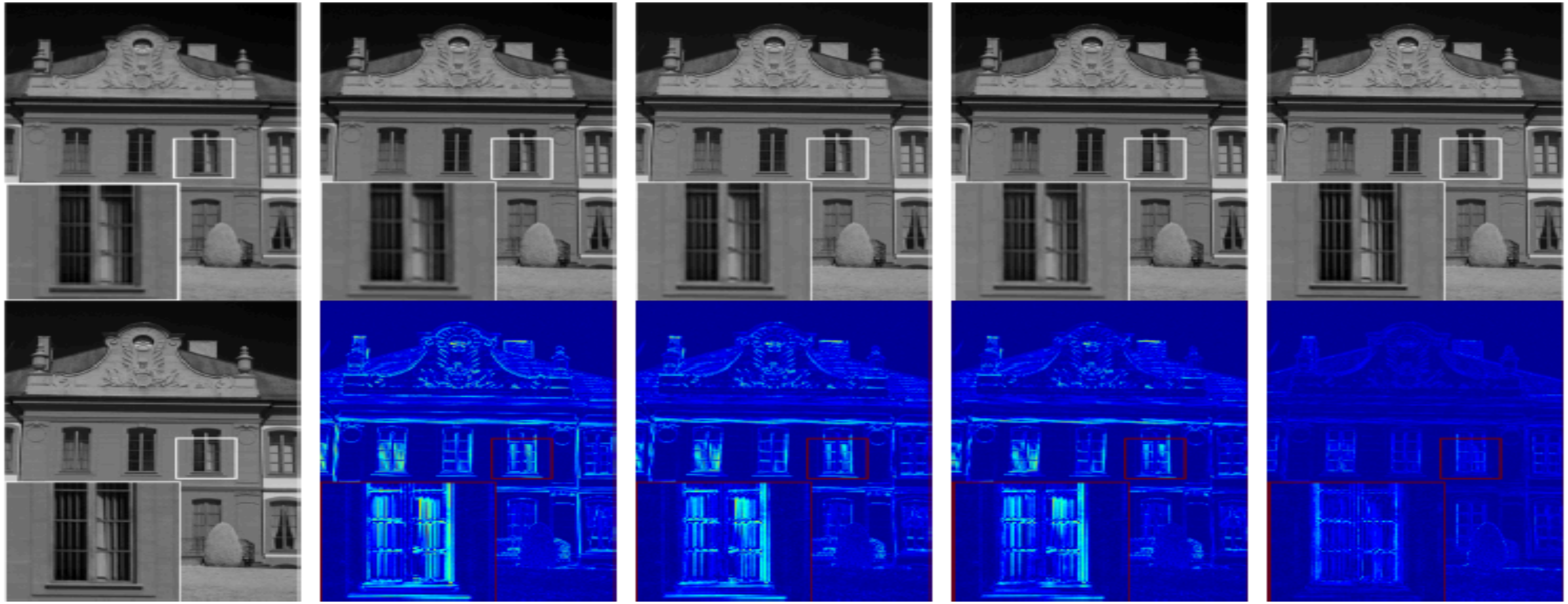


Ground Truth | Error - Bicubic | Error – Zeyde et al. | Error – A+ | Error – Ours

Multi-Modal Data Aided Super-Resolution

Super-resolving infrared images with the aid of RGB images

LR-Image | HR-Image - Bicubic | HR-Image – Zeyde et al. | HR-Image – A+ | HR-Image – Ours



Ground Truth | Error - Bicubic | Error – Zeyde et al. | Error – A+ | Error – Ours

Outline

i. Parsimonious Representations for Unimodal Data Processing

ii. Joint Parsimonious Representations for Multimodal Data Processing

iii. Multimodal Data Aided Processing

   a. Image separation aided by multimodal data

   b. Image super-resolution aided by multimodal data

iv. Concluding Remarks and Directions

Concluding Remarks and Directions

i. Joint sparse representations induced by coupled dictionaries can also address emerging multi-modal data processing problems.

ii. A number of applications have been demonstrated in the context of art-investigation and beyond.

iii. The techniques can be used to address various other multi-modal imaging processing tasks and applications.