Low-Rank Nonnegative Factorizations for Spectral Imaging Applications

Bob Plemmons
Wake Forest University

• Collaborators:
Christos Boutsidis (U. Patras), Misha Kilmer (Tufts), Peter Zhang, Paul Pauca, (WFU)

• Interaction on spectral data with: Kira Abercromby (NASA-Houston)

• Related Papers at: http://www.wfu.edu/~plemmons

• **Project Funded by AFOSR**

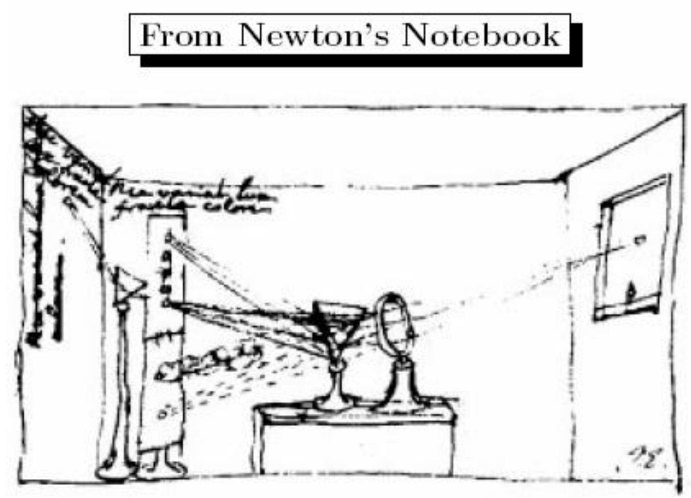
Stanford **Workshop on Algorithms for Modern Massive Data Sets**, June 06

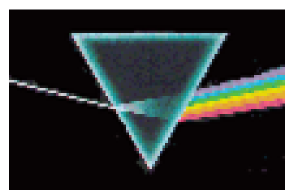
Outline

- Object Identification from Spectral Data
- Features-Based Clustering & Classification
- Nonnegativity Constrained Low-Rank Approximation for Blind Source Separation and Unsupervised Unmixing (Ill-posed, nonlinear inverse problem)
- Nonnegative Matrix Factorization (NMF)
- Results using Air Force data from Maui and data from K. Abercromby at NASA JSC
- Preliminary Results on using Perron-Frobenius Theory to Compress Hyperspectral Sensor Data
- Comments on Nonnegative Tensor Factorization (NTF) for Image data (see poster by Christos Boutsidis)

Simple Analog Illustration

Hidden Components in Light – Separated by a Prism





Our purpose – finding hidden components by <u>data analysis</u>

Blind Source Separation for Finding Hidden Components (Endmembers)

Mixing of Sources

…basic physics often leads to linear mixing…

$X = [X_1, X_2, \dots, X_n]$ –column vectors (1-D spectral scans)

Approximately factor

$$X \quad W H = \quad {}_1^{k} w^{(j)} \pm h^{(j)}$$

$\pm$ denotes outer product

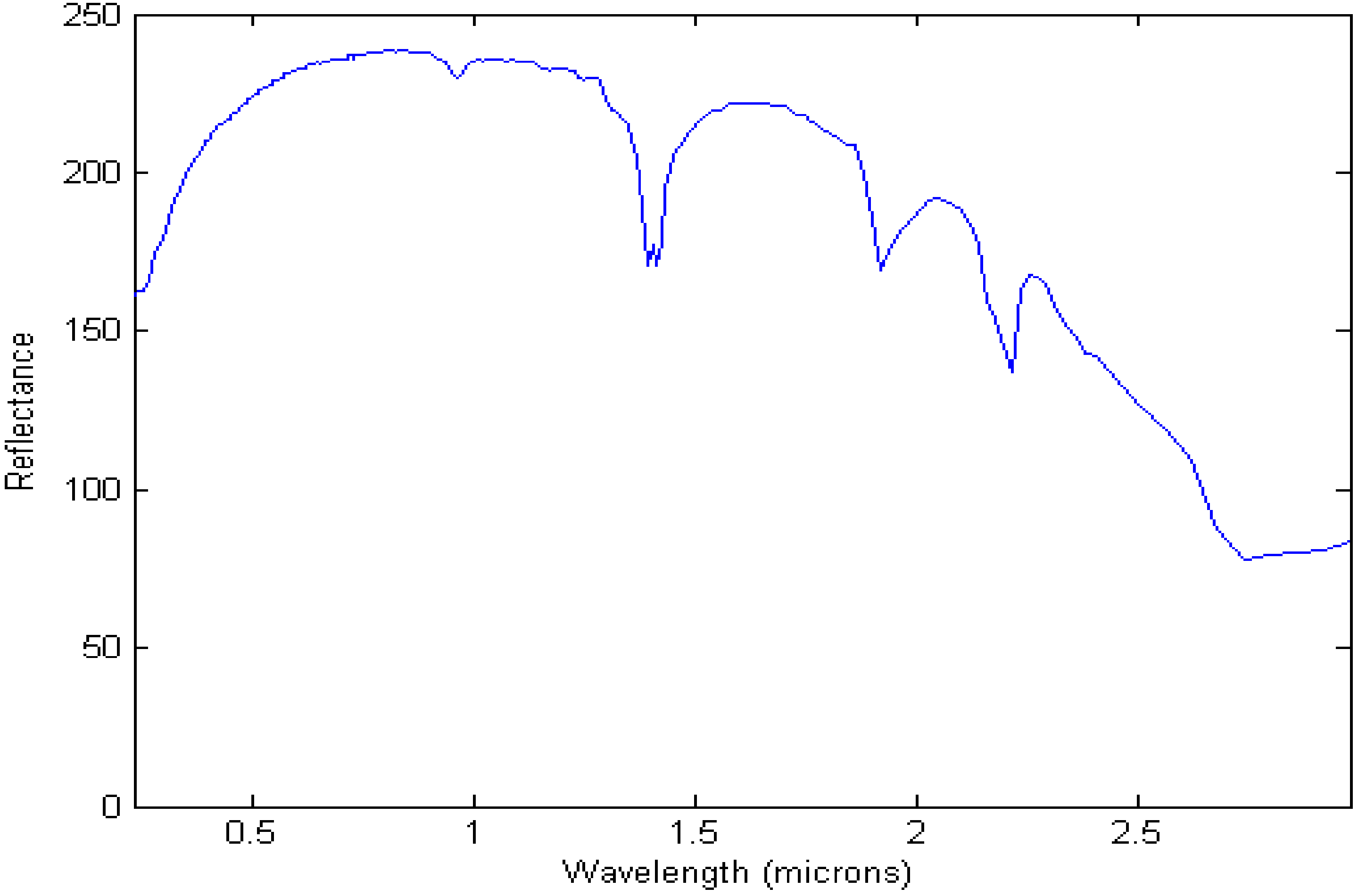
w^j is jth col of W, $\mathbf{h}^j$ is jth col of H^T

X sensor readings (mixed components – observed data)

W separated components (feature basis matrix, unknown, low rank)

H hidden mixing coefficients (unknown), replaced later with abundances of materials that make up the object.

Typical Scan



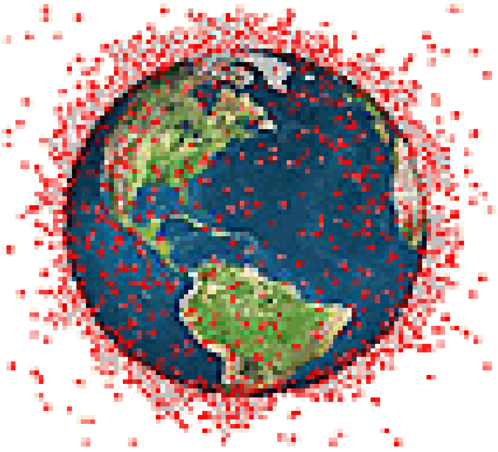
- NMF Allows only additive, not subtractive combinations of the original data, in comparison to orthogonal decomposition methods, e.g. PCA.
- Used by Lee and Seung (MIT) in *Nature*, 1999, in biometrics, preceded and followed by numerous papers related to applications.
- Matlab Toolbox: NMFLAB, http://www.bsp.brain.riken.jp/
- Historical perspective:

  Problem 73-14, Rank Factorization of Nonnegative Matrices, by A. Berman and R.J. Ple., SIAM Review 15 (1973), p. 655: (Also in Berman/Ple. book)

Some General Applications of NMF Techniques

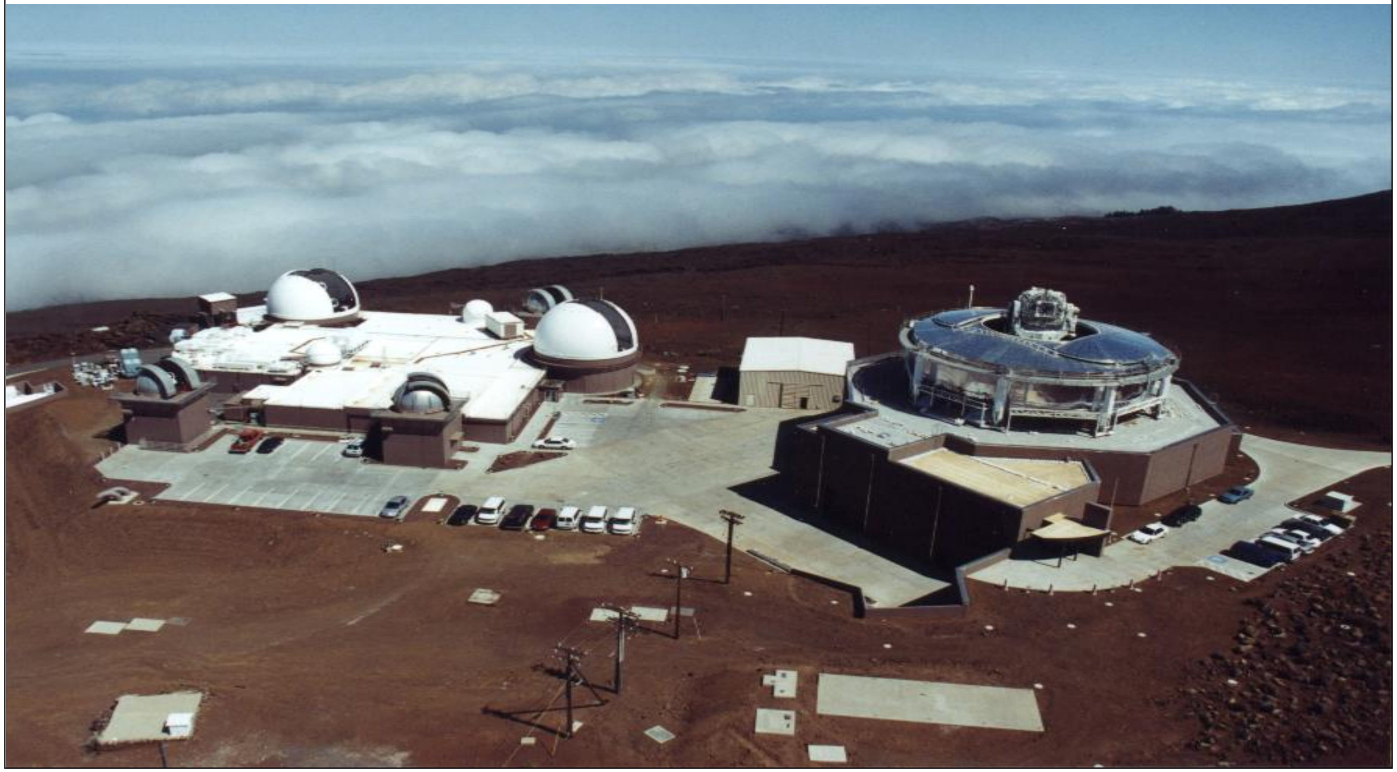
- Source separation in acoustics, speech, video
- EEG in Medicine, electric potentials
- Spectroscopy in chemistry
- Molecular pattern discovery - genomics
- Thermal nondestructive testing - aircraft and missile parts
- Email surveillance
- Document clustering in text data mining
- Atmospheric pollution source identification
- Hyperspectral sensor data compression
- **Spectroscopy for space applications – spectral data mining**
  - **Identifying object surface materials and substances**

Space Object Identification and Characterization from Spectral Reflectance Data

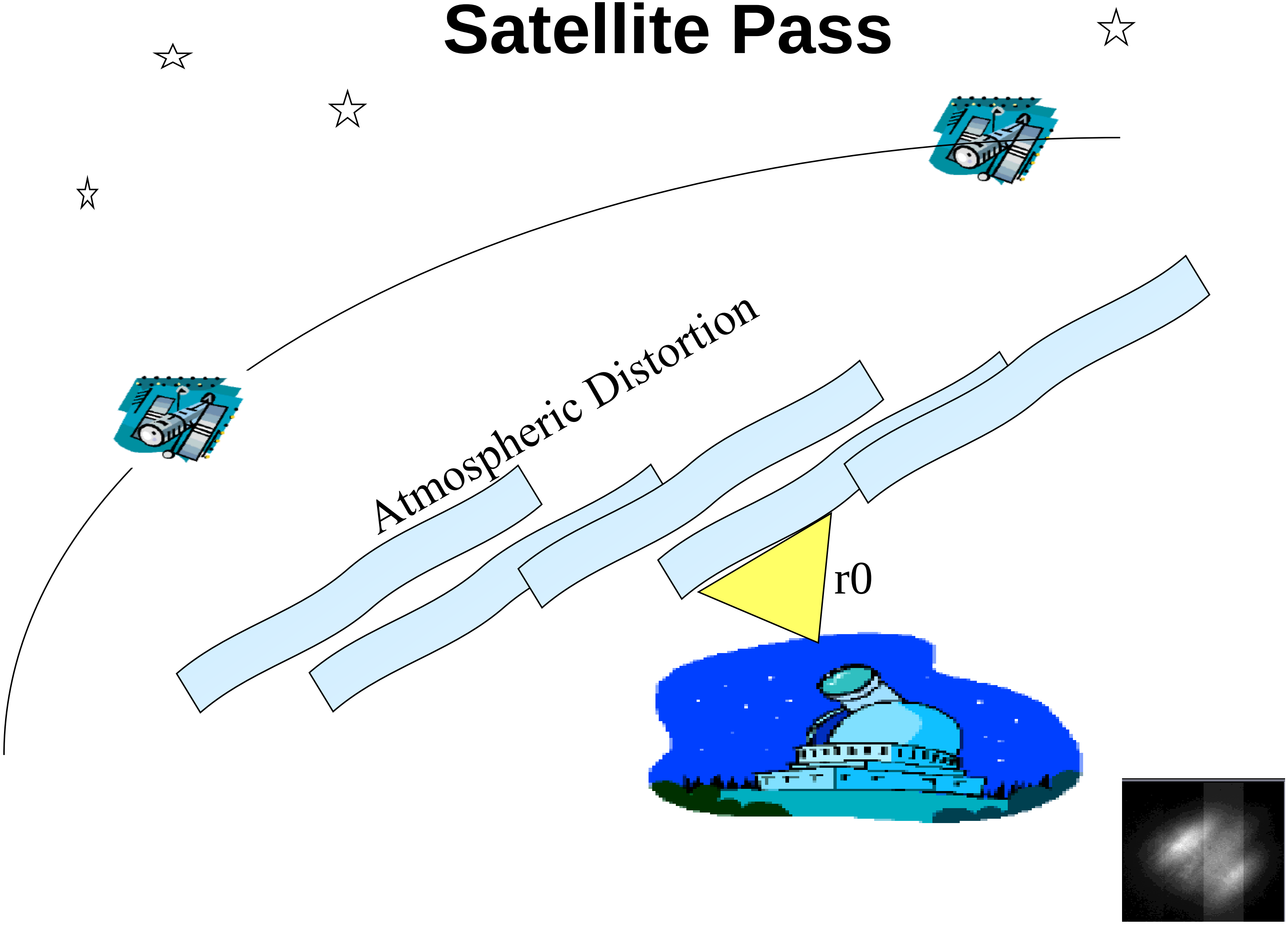


More than 15,000 known objects in orbit: various types of military and commercial satellites, rocket bodies, residual parts, and debris – need for space object database mining, object identification, clustering, classification, etc.

Maui Space Surveillance Site



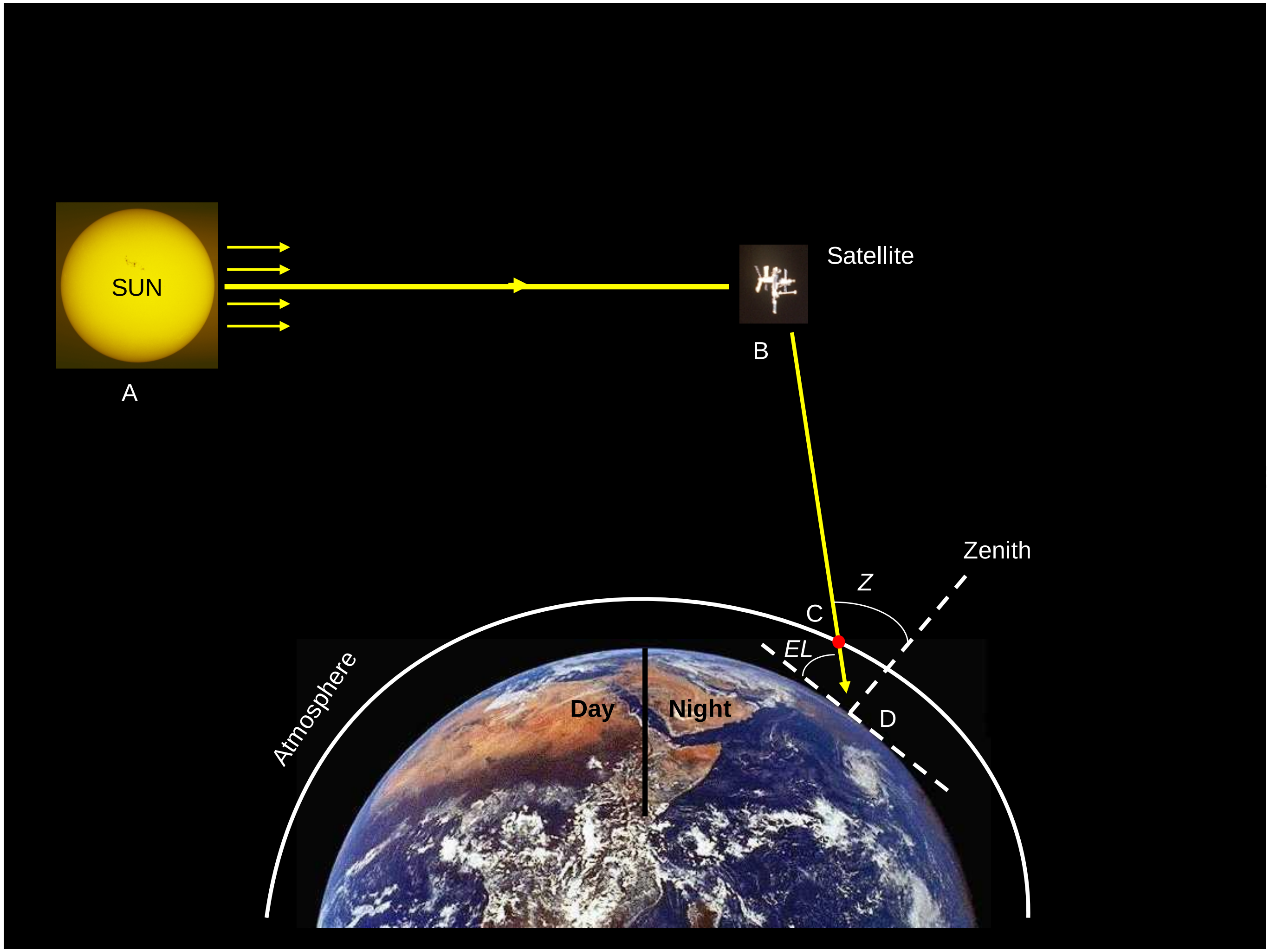
Satellite Pass



Imaging Sciences for Space Situational Awareness by Monitoring Space Satellites (AFOSR)

- 'Listen' (laser enabled vibrometry)
- 'Smell' (chemical sensing with spectrometer)
- 'Touch' (scatterometry/polarimetry for surface texture information)
- 'See' (by sequential speckle <video> imaging)
- 'characterize materials' (spectral imaging)

The creation and observation of a reflectance spectrum

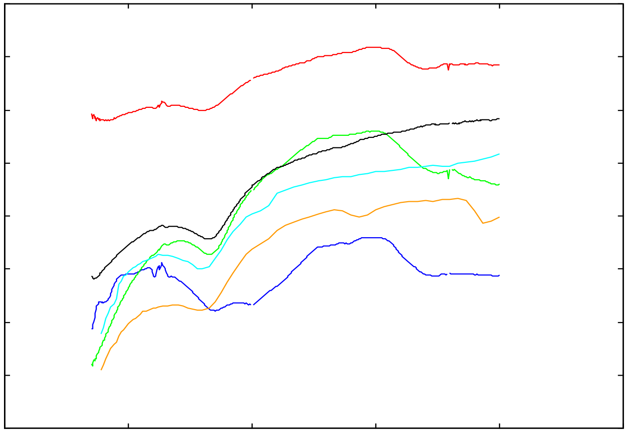


Spectral Imaging of Space Objects

- Current "operational" capability for spectral imaging of space objects

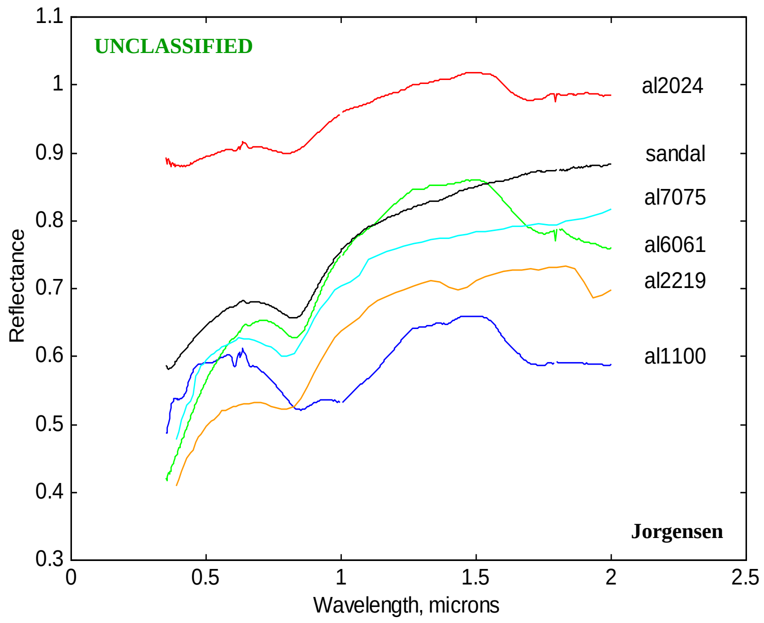
- Panchromatic images

- Non-imaging spectra



Spectral Imaging of Space Objects

- Why look at anything spectrally?
  - Simple answer: Color vs. Black and white
  - More involved answer: Spectral radiometry
- For space objects were looking at being able to:
  - Differentiate between different material classes
  - Material degradation
  - Identify hidden payloads
  - Anomaly resolution



Overview of the SOI Problem

- Space activities require accurate information about orbiting objects for space situational awareness
- Many objects are either in
  - Geosynchronous orbits (about 40,000 KM from earth), or
  - Near-Earth orbits, but too small (e.g., space mines) to be resolved by optical imaging systems
  - Can approximately collect one pixel/object by optical telescope

Overview of the SOI Problem Continued

- Problem solution by <u>learning the parts of objects </u>(hidden components) by low rank nonnegative sparse representation

- Basis representation (dimension reduction) can enable <u>near real-time object (target) recognition, object class **clustering**, and characterization.</u> (ill-posed inverse problem)

- <u>Match recovered **hidden components** with known spectral signatures</u> from substances such as mylar, aluminum, white paint, kapton, and solar panel materials, etc. This is classification.

- Fundamental difficulty: Find from spectral measurements:
  – <u>Endmembers</u>: types of constituent materials
  – <u>Fractional abundances</u>: proportion of materials that comprise the object.

Approximate NMF

- Utilize constraint that sensor data values in X are <u>nonnegative</u>

- Apply non-negativity constrained low rank approximation for blind source separation, dimension reduction and unsupervised unmixing

- Low rank approximation to data matrix ***X*** :

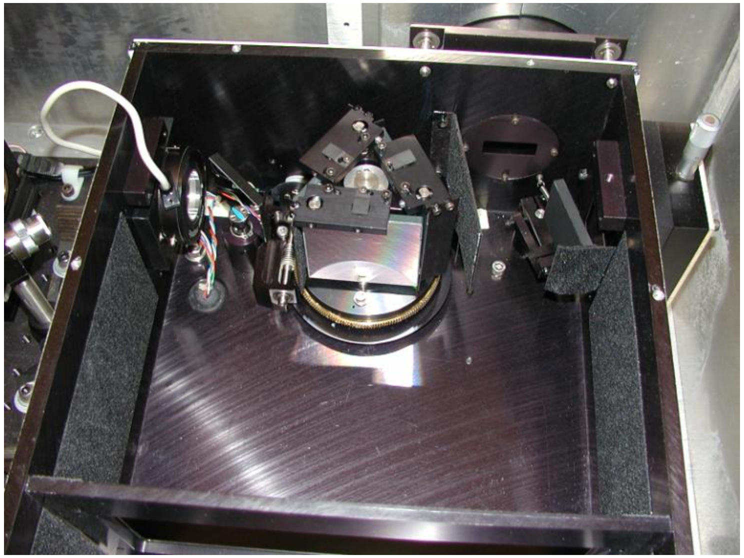
$$\boldsymbol{X} \qquad \boldsymbol{WH},\ \ \boldsymbol{W} \geq 0,\ \ \boldsymbol{H} \geq 0$$

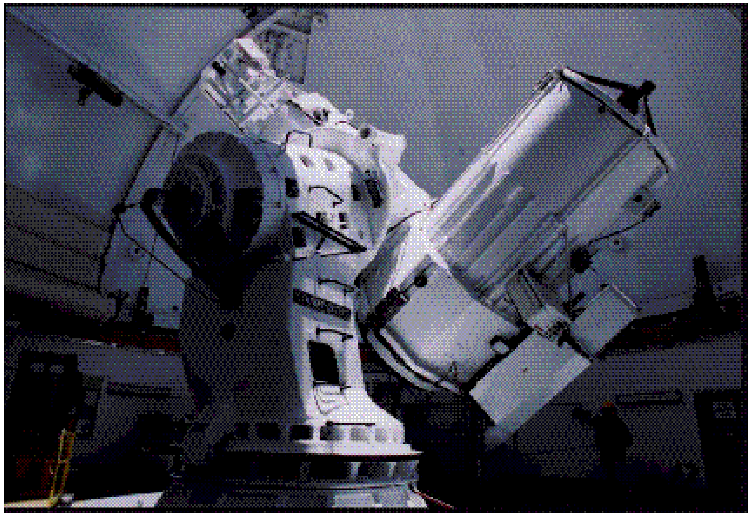
  - Columns of ***W*** are initial basis vectors for spectral trace database, may want smoothness and statistical independence in ***W***.

  - Columns of ***H*** represent mixing coefficients, desire statistical sparsity in ***H*** to force essential uniqueness in *W*. May want <u>sparsity</u> for **H**.

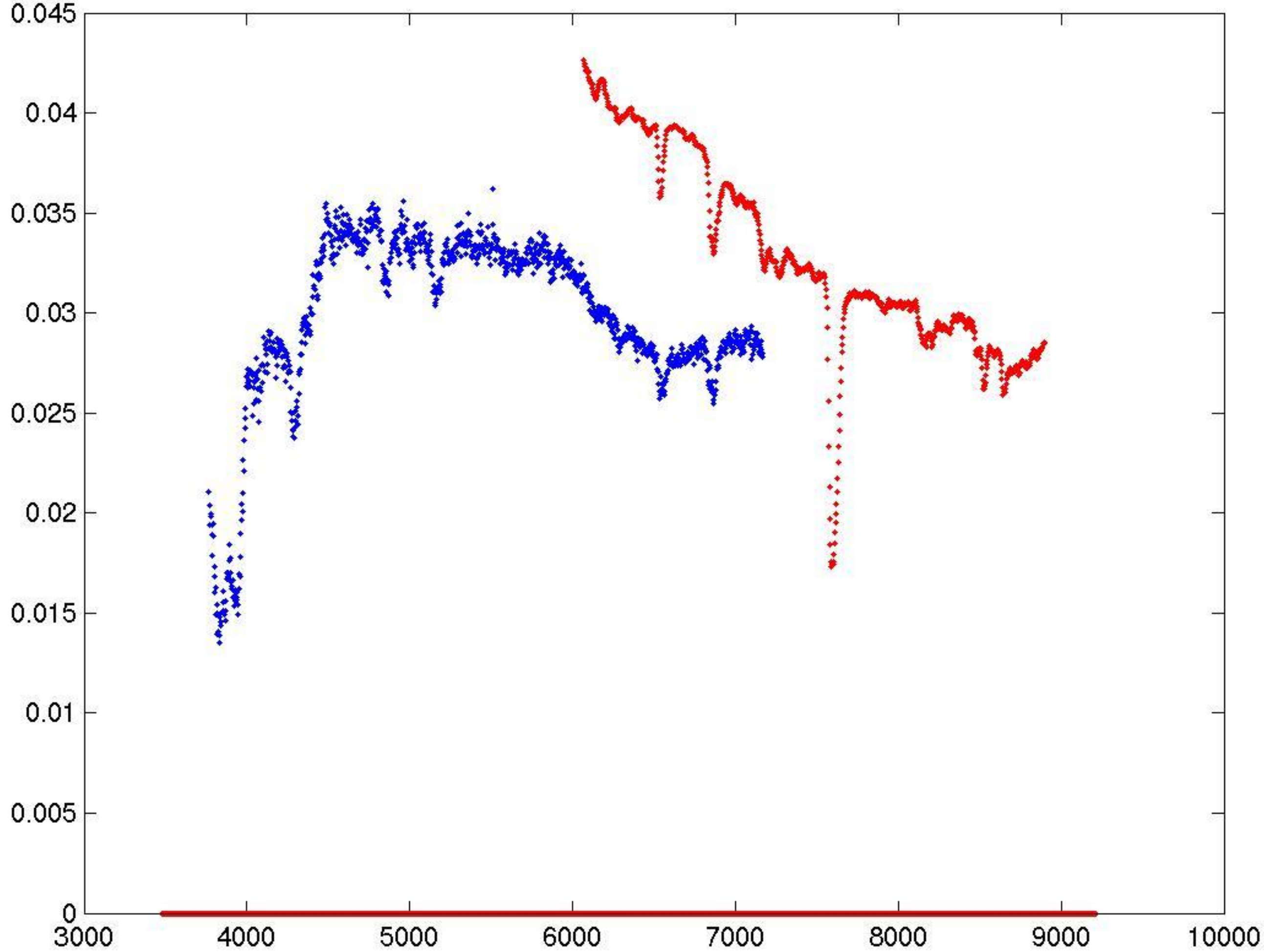
Some of our Data Obtained from a Spica (Space Infrared Telescope for Cosmology and Astrophysics) Spectrometer

- Mission: Support non-imaging SOI with spectroscopic observations
- 3 – 4 angstrom resolution
- Blue mode: 3000 – 6000 angstroms (.3 – .6 mm)
- Red mode: 6000 – 9000 angstroms (.6 – .9 mm)
- Located on Maui





Sample Raw Data Collected in Blue and Red Modes



Electromagnetic Spectrum: Spectral Signatures

- **For any given material, the amount of solar (or other) radiation that it reflects, absorbs, or transmits varies with wavelength.**

- **This property of matter makes it possible to <u>identify different substances</u> out of 300$^+$ and separate them by their spectral signatures (spectral curves) – spectral unmixing, finding fractional abundances.**

An Approach to Finding Endmembers and Fractional Abundances

- Vectorize the spectral scans of space objects into columns of ***X*** (works well for 1-D signals, not for

  2-D images
- Cluster the columns of ***X*** using a NMF scheme

  X   ***WH***, ***W*** ≥ 0 (smooth), ***H*** ≥ 0 (sparse)

(We use a metric by Hoyer to enforce sparsity in **H**.)

$$\text{sparseness}(x) = \frac{\sqrt{n} - \frac{\|x\|_1}{\|x\|_2}}{\sqrt{n} - 1}$$

Parts- Based Clustering & Classification

- Features from hidden components: parts-based learning algorithms from training set data
- Utilize constraint that spectral trace reflectance values are nonnegative
- Arrange the spectral traces into columns of a (nonnegative) database matrix denoted by $\boldsymbol{X}$
- Non-negativity constrained low rank approximation for blind source separation and unsupervised unmixing
- Low rank approximation to data matrix $\boldsymbol{X}: \ \boldsymbol{X} \qquad \boldsymbol{WH}, \ \boldsymbol{W} \geq 0, \ \boldsymbol{H} \geq 0$
    - Columns of $\boldsymbol{W}$ are basis vectors for spectral trace database (endmembers)
    - $\boldsymbol{H}$ eventually discarded and new reduced $\boldsymbol{H}$ computed
    - Alternating iterations used

NMF Problem Formulation

- Given initial database expressed as n x m nonnegative matrix **X**

  find two reduced-dimensional matrices **W (n x r)** and **H (r x m)** to:

$$\min_{W,H} \|X - WH\|_F^2, \quad \text{plus constraints}$$

where $W_{ij} \geq 0$ and $H_{ij} \geq 0$ for each i and j. Choice of r << m is often problem dependent. Can impose other (e.g., smoothness, sparsity) constraints on W and/or H.

NMF - Continued

- Can use convex cone theoretic geometric concepts to determine conditions for uniqueness, up to permutation and scaling of the rows (Donaho and Stodden).
- Constraints on *H* strongly affect uniqueness in *W*.

Lee and Seung (1999) proposed a multiplicative alternating iteration scheme

1. Initialize *W* and *H* with nonnegative values and scale columns of W to unit norm.
2. Iterate for each c, j and i until convergence or stop (eps is a machine dependent small positive pos. no.):

$$\text{(a)}\quad H_{cj} \leftarrow H_{cj}\frac{(W^TX)_{cj}}{(W^TWH)_{cj}+eps}$$

$$\text{(b)}\quad W_{ic} \leftarrow W_{ic}\frac{(XH^T)_{ic}}{(WHH^T)_{ic}+eps}$$

(c) Scale the columns of W to unit norm.

- ❑ Process is essentially a diagonally-scaled gradient descent method of EM (R-L) type.

<u>But</u>, clustering is ill-posed. Regularization may be needed.

New Approach to Selecting Endmembers and Computing Fractional Abundances

- Vectorize the spectral scans of space objects into columns of a matrix ***Y***

- Cluster the columns of ***Y*** using a NMF scheme

$$\boldsymbol{Y} \quad \boldsymbol{WH}, \;\; \boldsymbol{W} \geq 0, \;\; \boldsymbol{H} \geq 0$$

  (Enforce smoothness on *W* and sparsity on *H*.)

- Classify the basis vectors in ***W*** using lab data from Jorgersen and an *information theoretic* scoring method (Kullback-Leibler divergence, i.e., relative entropy). Represent these endmembers by a matrix ***B***.

- ***B*** represents a compressed database for ***Y*** and has a variety of uses, e.g., ...

- Determine the spectral abundances of the space object spectral scans in columns of ***Y*** by iteratively solving nonlinear least squares problems with matrix ***B*** containing the classified endmembers.

  (We use a nonlinear least squares scheme to compute material abundancies.)

- **Minimize a functional *F(W, H)* by solving the following constrained optimization problem. (Here a and b are regularization parameters).**

$$\min_{W,H} \left\{ \|Y - WH\|_F^2 + \alpha J_1(W) + \beta J_2(H) \right\}, \text{ for } W \geq 0 \text{ and } H \geq 0$$

where $\alpha J_1(W)$ and $\beta J_2(H)$ are used to enforce certain application-dependent characteristics on the solution

- **Determine gradients for W and H and set each to zero (alternating iterations).**

Sparse CNMF

We define sparseness of a vector x of length n as

$$\text{sparseness}(x) = \frac{\sqrt{n} - \frac{\|x\|_1}{\|x\|_2}}{\sqrt{n} - 1}$$

Given A, a matrix of arbitrary size, let

$$\bar{A} = \text{vec}(A),$$

denote the vector formed by stacking the columns of A.

Now consider the following objective function:

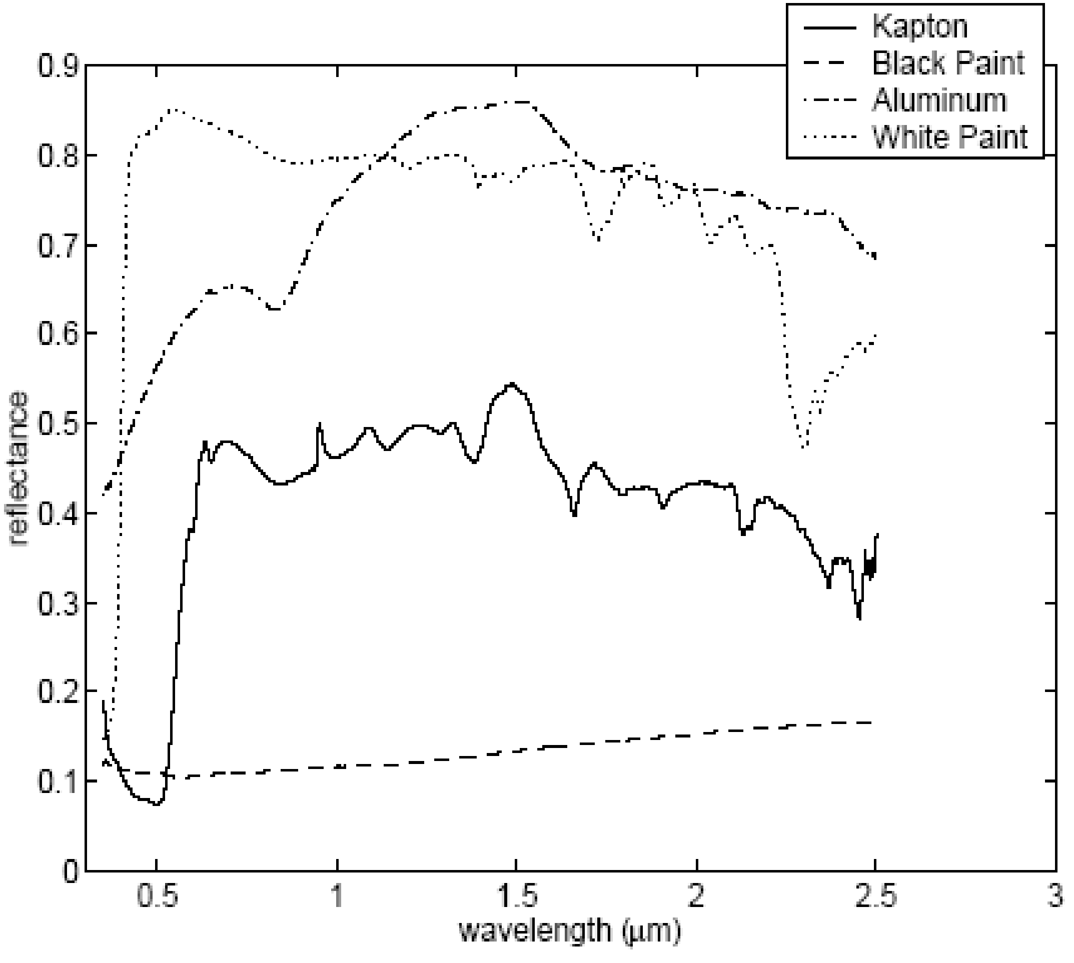
$$F(W, H) = \frac{1}{2}\|Y - WH\|_F^2 + \frac{\beta}{2}(\omega\|\bar{H}\|_2 - \|\bar{H}\|_1)^2,$$

where Y is $m \times n$, W is $m \times k$, and H is $k \times n$, and

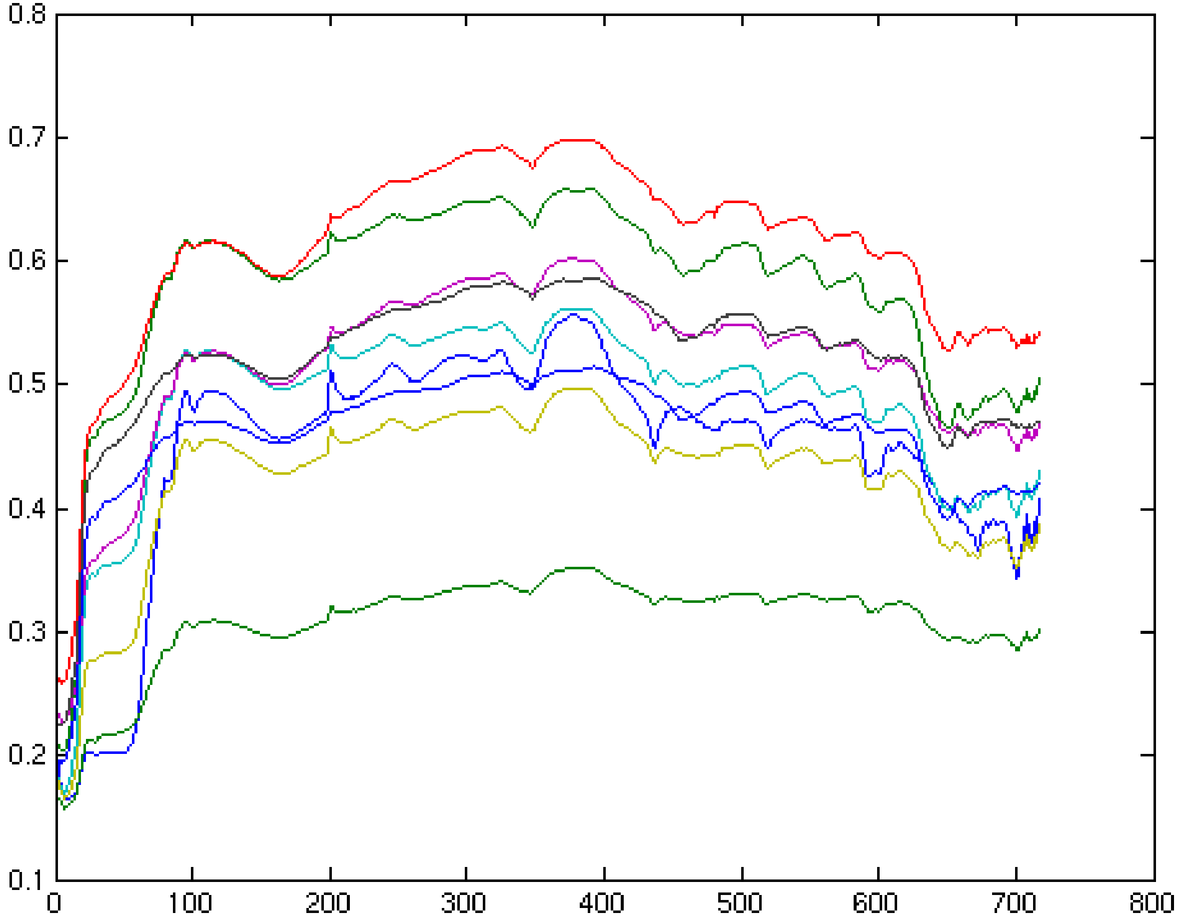
$$\omega = \sqrt{kn} - (\sqrt{kn} - 1)\,\text{sparseness}(H).$$

Compute gradient, insert in basic optimization expression, and apply <u>alternating iterations</u>. Results in basis matrix ***W*** with a sparse mixing matrix ***H***.

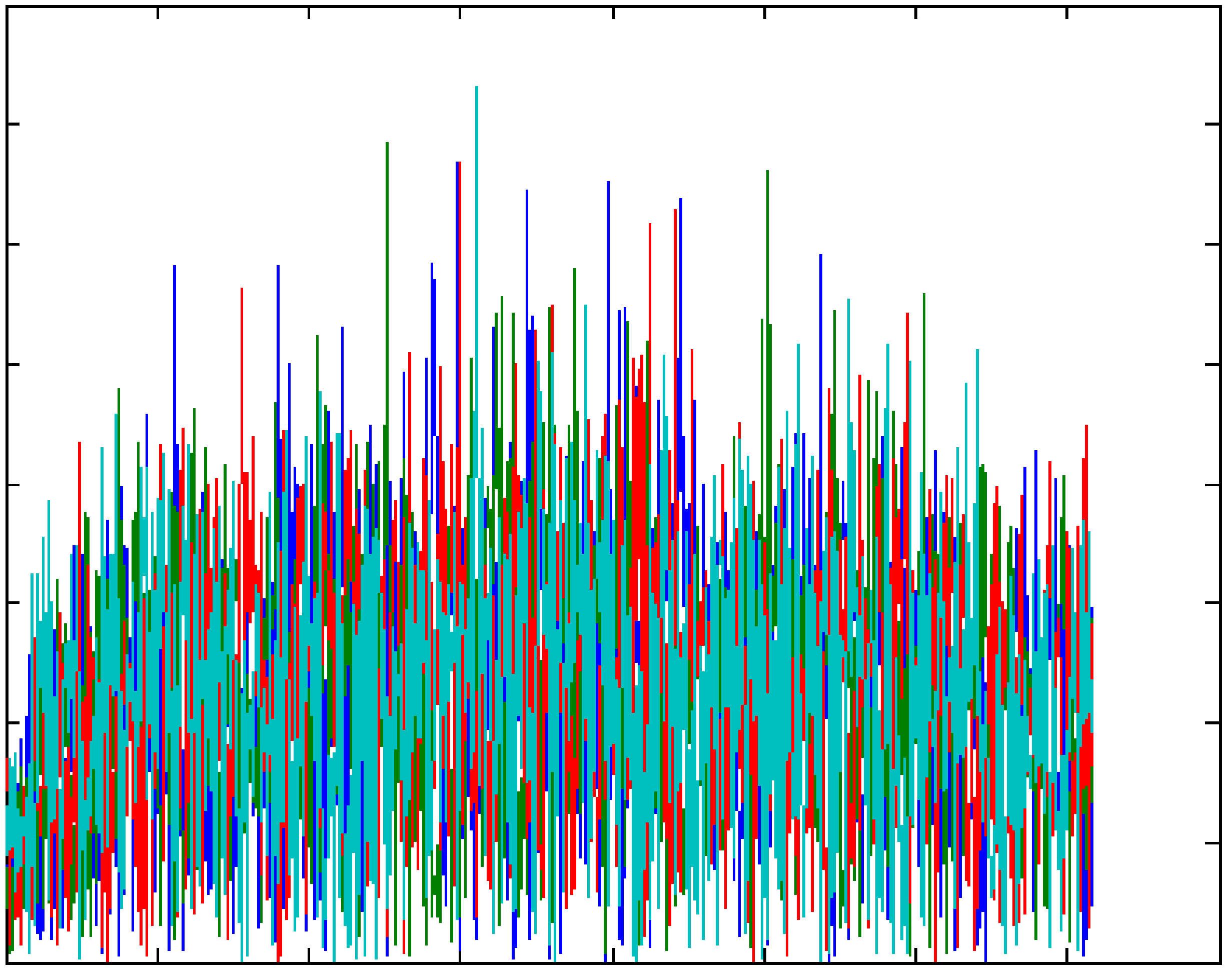
Sample Results – Finding only Endmembers We Form Simulated Satellites from NASA Data



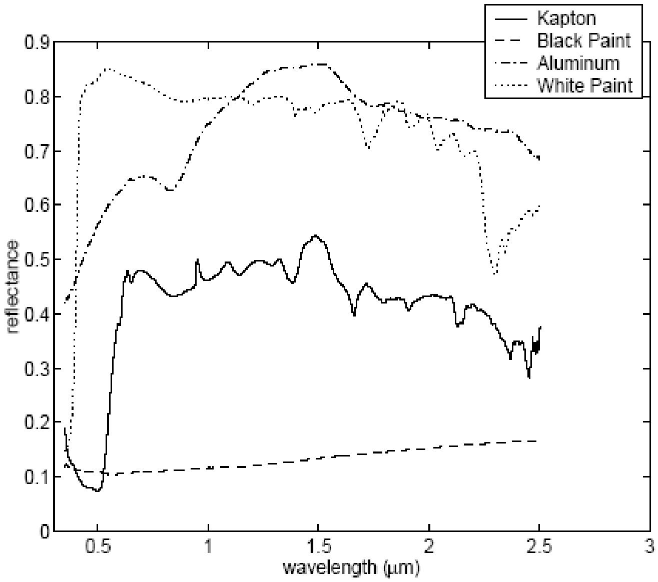
A Few Combined Traces (time varying mixtures)



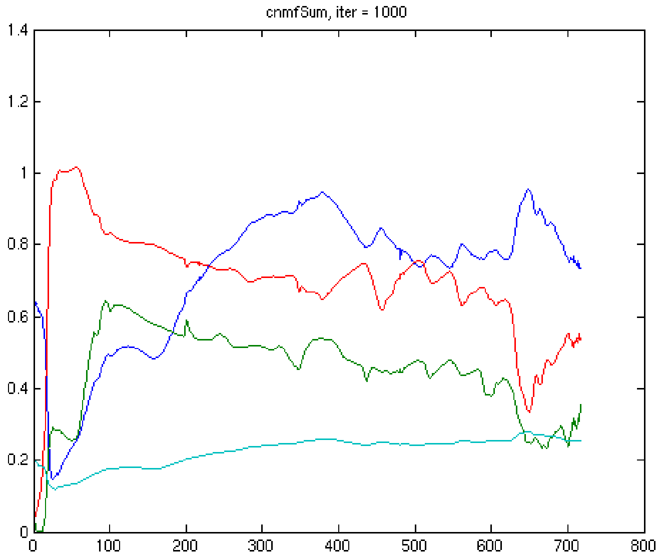
Blind Source Separation Using NMF



Original



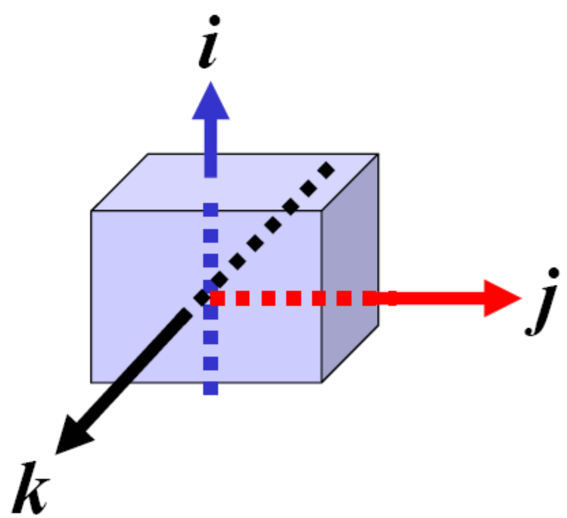
Recovered



Extend NMF: Nonegative Tensor Factorization (NTF)

Joint project with Christos Boutsidis and Peter Zhang

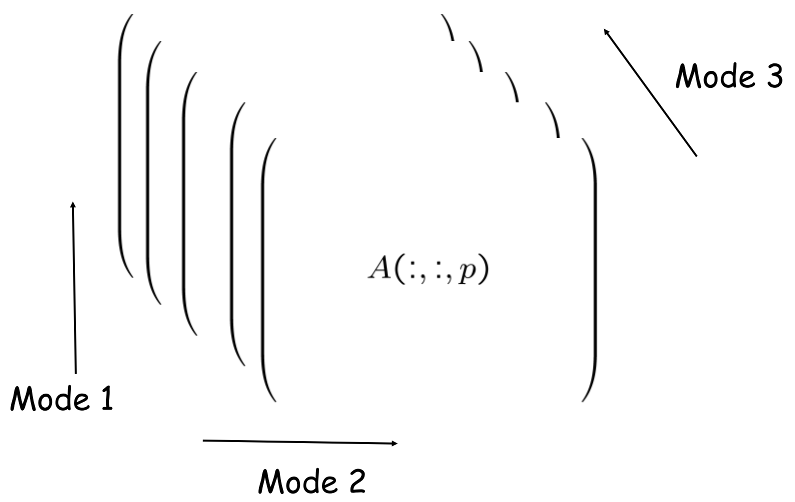
- Our interest: 3-D data. 2-D images stacked into 3-D Array, forming a “box”



Datasets of images modeled as tensors

Goal: Extract features from a tensor dataset (naively, a dataset subscripted by multiple indices). Image samples with diversities, e.g., eigenviews.

m £ n £ p tensor *A*



What is NTF (for 3-D Arrays)?

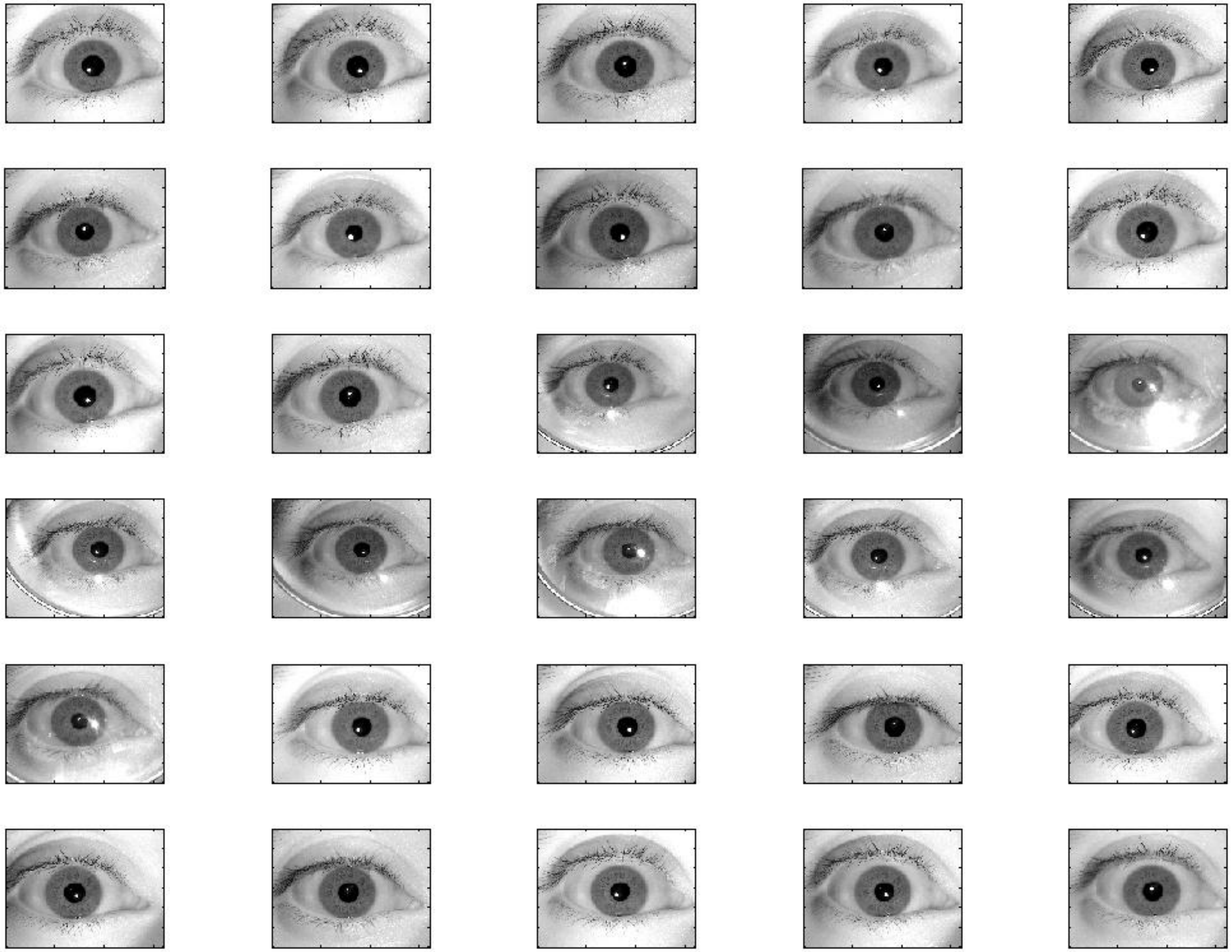
Given a nonnegative tensor $\mathcal{A} \in R^{m\times n\times p}$ *and a positive integer* k*, find nonnegative vectors* $u^{(i)} \in R^{m\times 1}$, $v^{(i)} \in R^{n\times 1}$ *and* $w^{(i)} \in R^{p\times 1}$ *to minimize the functional*

$$\frac{1}{2}\|\mathcal{A} - \sum_{i=1}^{k} u^{(i)} \circ v^{(i)} \circ w^{(i)}\|_F^2.$$

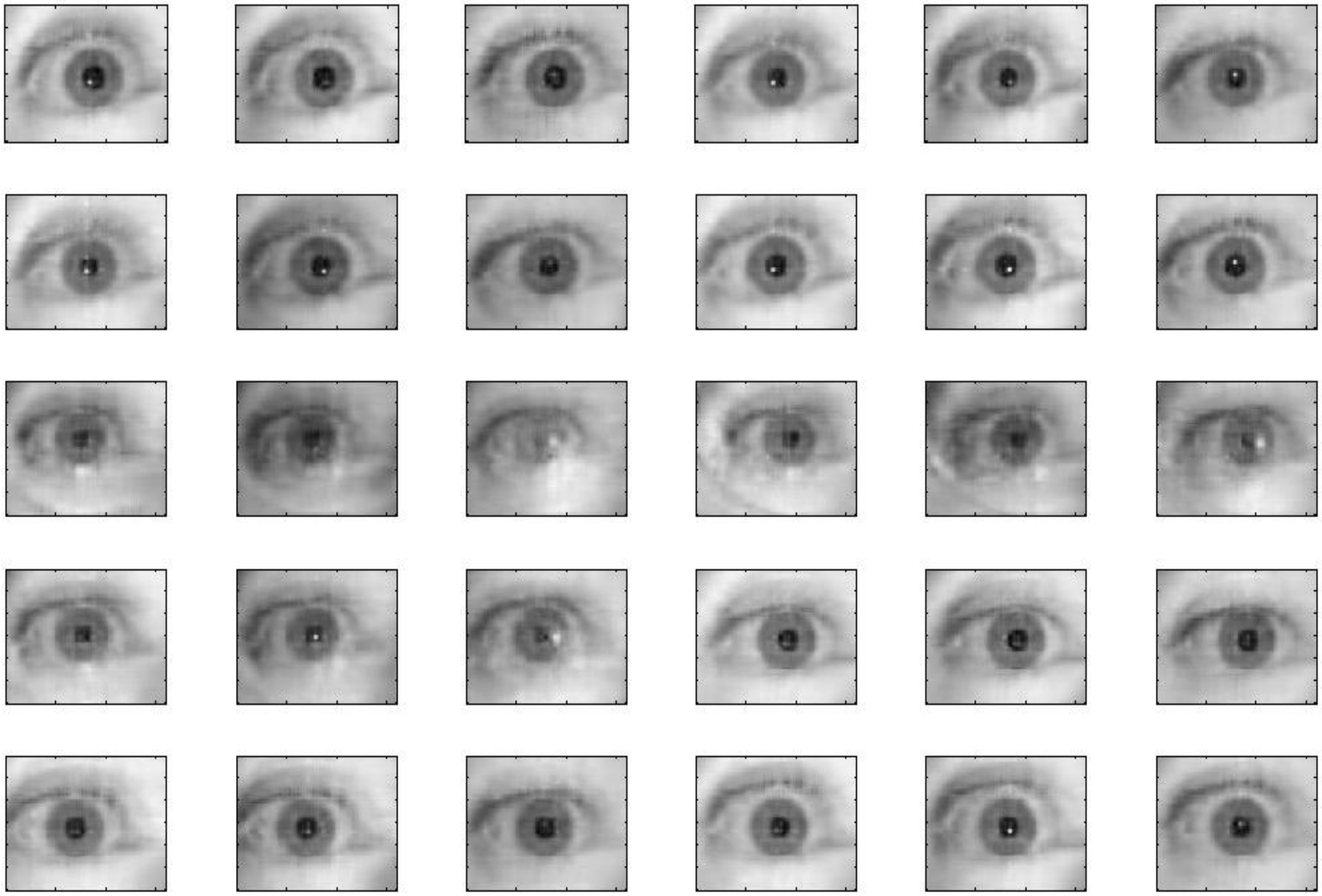
Here $\circ$ *denotes "outer product". The rank-one matrices* $u^{(i)} \circ v^{(i)}$ *are the desired basis components, and* $w^{(i)}$ *the weights.*

- See poster by Christos Boutsidis
- Issues: Uniqueness, Initialization, Efficient optimization algorithms

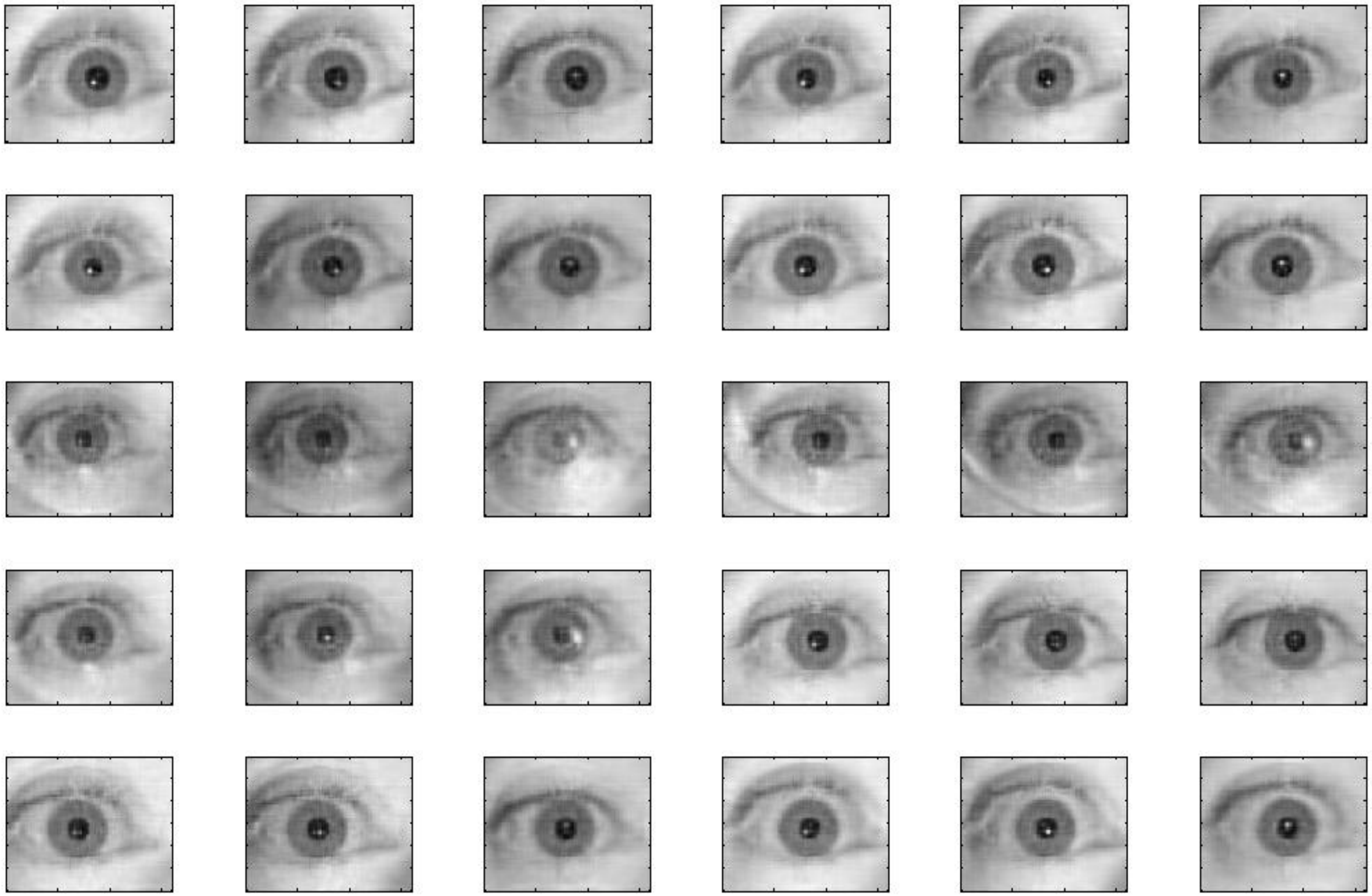
Iris Recognition Images



Recovered Images using PARAFAC ~1 hr



Recovered Images using Boutsidis/Zhang ~ 5 min

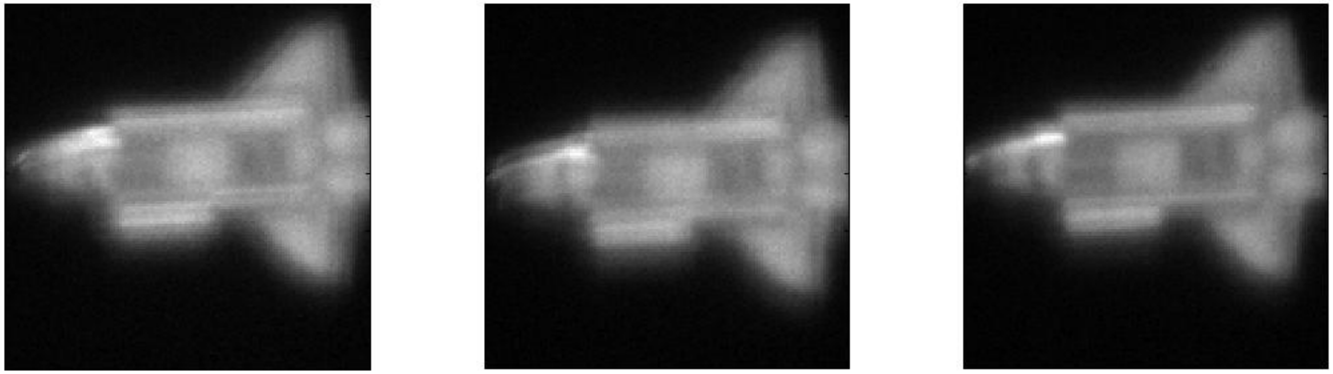


Compression

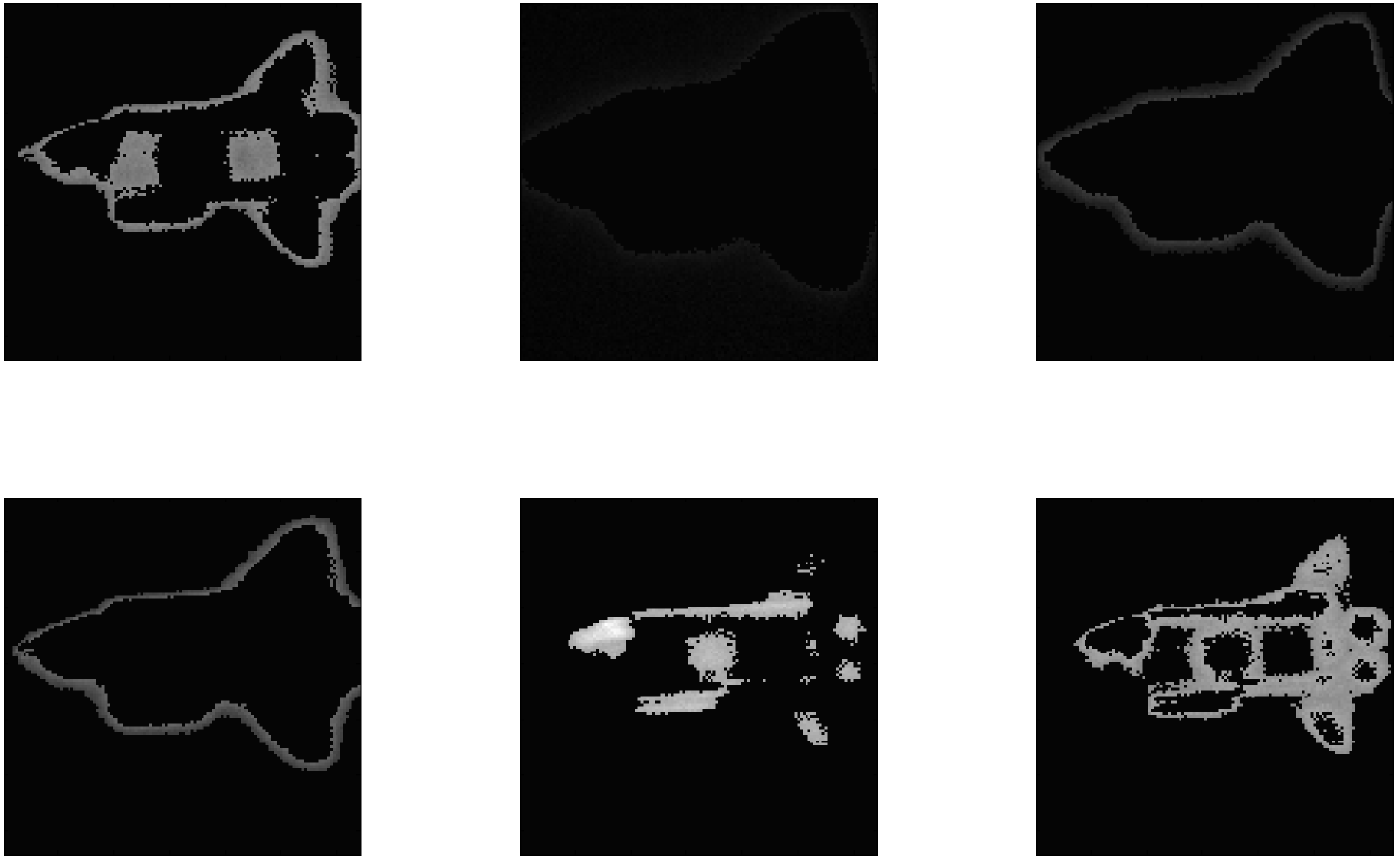
• Original array	120x160x30	4,608,000	bytes
• X	120x50	48,000	bytes
• Y	160x50	64,000	bytes
• Z	30x50	12,000	bytes

- Compression ratio 37 to 1

Columbia in Final Orbit over Maui Space Center



K-means clustering



Summary and NMF Applications for Spectral Data

- Classification of objects in terms of material features and fractional abundances
- Database compression, including hyperspectral data
- Fast determination of whether a new object spectral trace is in the database, using basis matrix approximation
- Multiple observations with object in different orientations can provide object shape information
- Low-rank representation can enable fast object (target) recognition and tracking (Kullback-Leibler matching)
- Enabled in part by modified nonnegative matrix factorization and information theoretic techniques (relative entropy)
- Compression and reconstruction of image arrays data
- Some related papers at: http://www.wfu.edu/~plemmons