

Robust Multivariate Analysis for Problem Images

Jeremy M. Shaver

Eigenvector Research, Inc., Wenatchee, WA

Lei Pei

Guilin Jiang

Matthew C. Asplund

Matthew R. Linford

Department of Chemistry and Biochemistry,
Brigham Young University, Provo, UT 84602



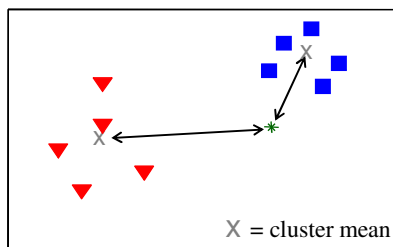
Robust Methods

- Multivariate Curve Resolution (MCR)
 - Quantitative/Qualitative analysis method of choice in many situations
 - Highly influenced by outliers!
- Statistically proven robust methods exist for performing standard analyses
 - Minimum Covariance Determinant \Rightarrow Robust PCA
 - Trimmed Least Squares \Rightarrow Robust MCR
 - Methods require several “phases” and can be slow for images (10k+ samples per image!)
- Exploring various methods for quick, qualitative and quantitative survey of images.
- Go “old school”...



k-Means Clustering

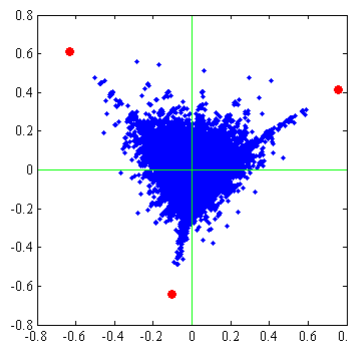
- Find natural clusters in data. Samples grouped by proximity to a class mean
- Gives average spectrum of each cluster (akin to loadings).
- Exceptionally fast relative to Multivariate Curve Resolution (MCR).
- Can handle rank-deficient situations.
- Does NOT handle mixture data well (e.g. when image spatial resolution is poor).
- Sensitive to initial guess.
- Various modifications exist in literature to temper algorithm.



EIGENVECTOR
RESEARCH INCORPORATED

“Pure Pixels” *(a.k.a. SIMPLISMA, purity)*

- MCR & Cluster have as goal to locate Pure Component Spectra
- Samples which are farthest out in multivariate space provide prototypical “target spectra”
- Pure pixel as k-Means initial guess provides method for stabilization! (Deterministic)

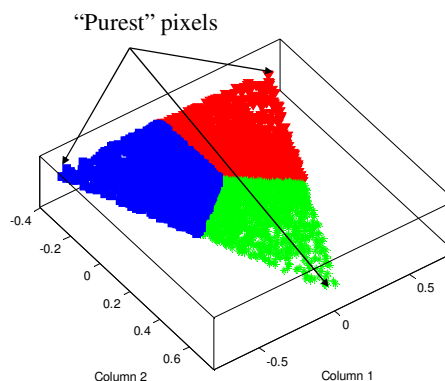


EIGENVECTOR
RESEARCH INCORPORATED

“Pure Pixel” k-Means Clustering

k-Means clustering with pure pixels:

1. Identify “n” most unusual pixels
2. Classify all other pixels by their closest pure pixel (projection)
3. Iterate with class mean as new “pure pixel” basis



EIGENVECTOR
RESEARCH INCORPORATED

TOF-SIMS of Time Released Drug Delivery System

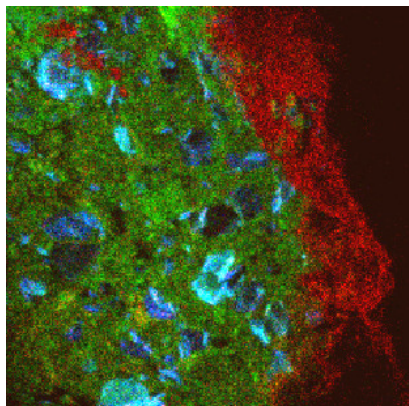
- ◆ Multilayer drug beads serve as controlled-release delivery system
- ◆ TOF-SIMS taken of cross section of bead
- ◆ Evaluate integrity of layers, distribution of ingredients
- ◆ Thanks again to Physical Electronics and Anna Belu for the data!

Reference: A.M. Belu, M.C. Davies, J.M. Newton and N. Patel, "TOF-SIMS Characterization and Imaging of Controlled-Release Drug Delivery Systems, Anal. Chem., 72(22), pps 5625-5638, 2000

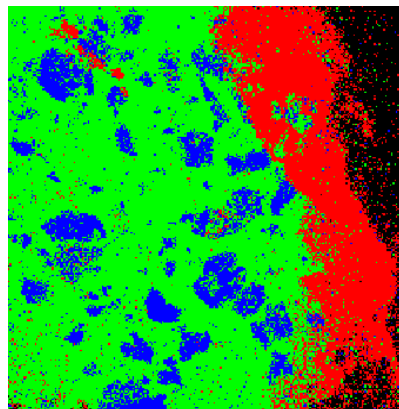
EIGENVECTOR
RESEARCH INCORPORATED

Avicel Pure-Pixel k-Means Clustering

False-color MCR Results



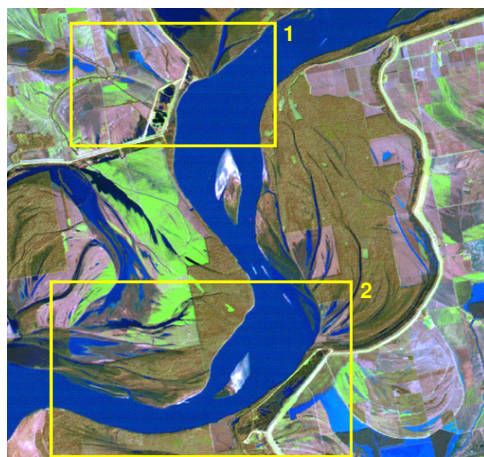
Pure Pixel Clusters



(3 clusters)



Satellite Image of Mississippi River



False-Color Scale
Blue = Short Vis. (red)
Green = NIR
Red = SWIR-1

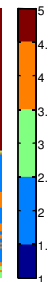
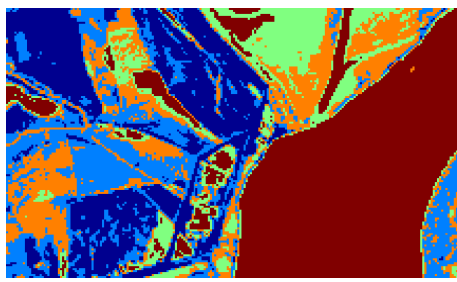
Seven-Slab Image: Red (vis), Green (vis),
Blue (vis), NIR, SWIR-1, SWIR-2, Thermal



“Pure Pixel” k-Means Clustering (region 1)



False-color image
Blue = Short Vis. (red)
Green = NIR
Red = SWIR-1



Pure pixel clustering with 5 classes identified.

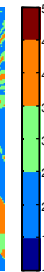
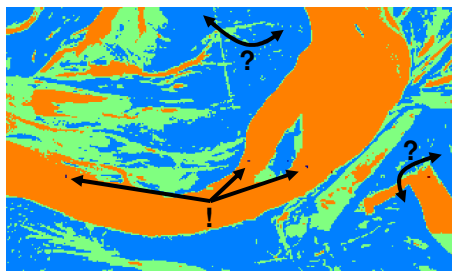
Can identify different land feature groups fairly well



Pure-Pixel K-Means Clustering (region 2)



False-color image



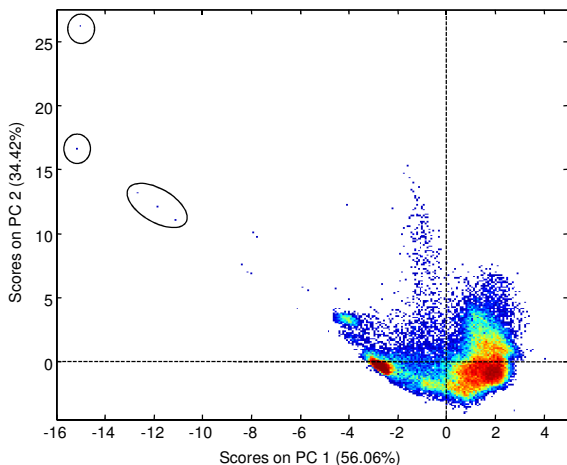
5-class k-means cluster

WHAT HAPPENED?

Obvious class differences are missed because classes 1 and 5 are wasted on bad pixels (in orange region)



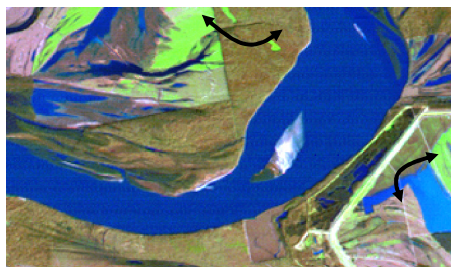
PCA Scores (PCs 1 and 2)



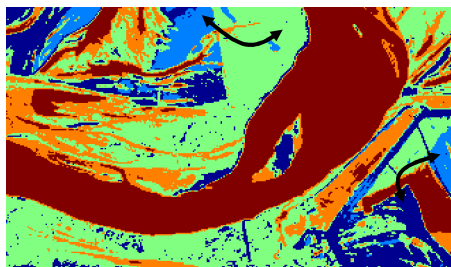
- Because outliers fall on the outside edges of multivariate space, they are almost always selected early by Pure-Pixel selection algorithm
- They are not similar to many other pixels
- THEREFORE: Exclude a pure point (and all like it) if it classifies < specified % of pixels... then select new pure points.
- NOTE: can save these for analysis of outliers! Might be what is of interest!



Robust Pure-Pixel k-Means Clustering



False-color image

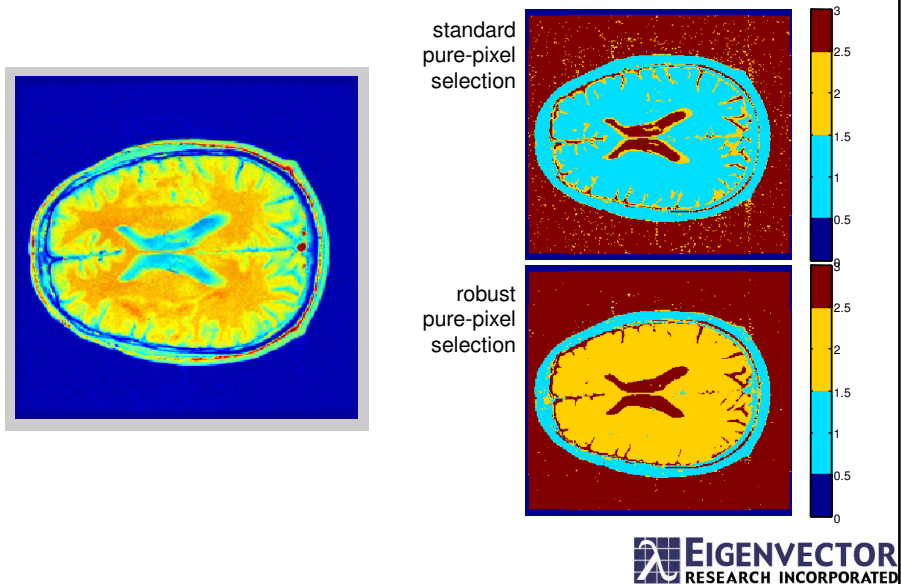


5-class k-means cluster WITH small-class exclusion

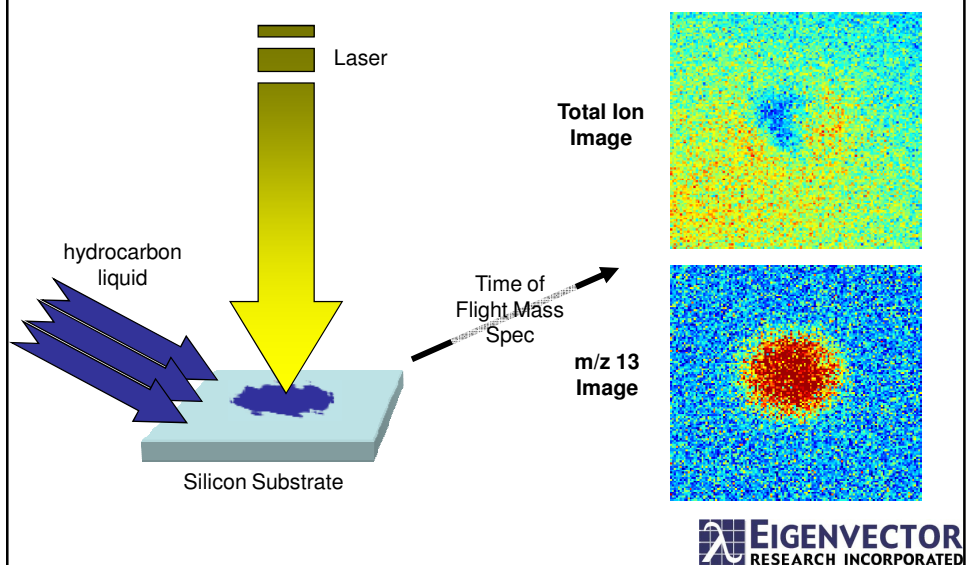
Bad pixels are discarded from "pure spectrum" selection when classes from them do not capture > 1% of pixels. Result is better classification of variations.

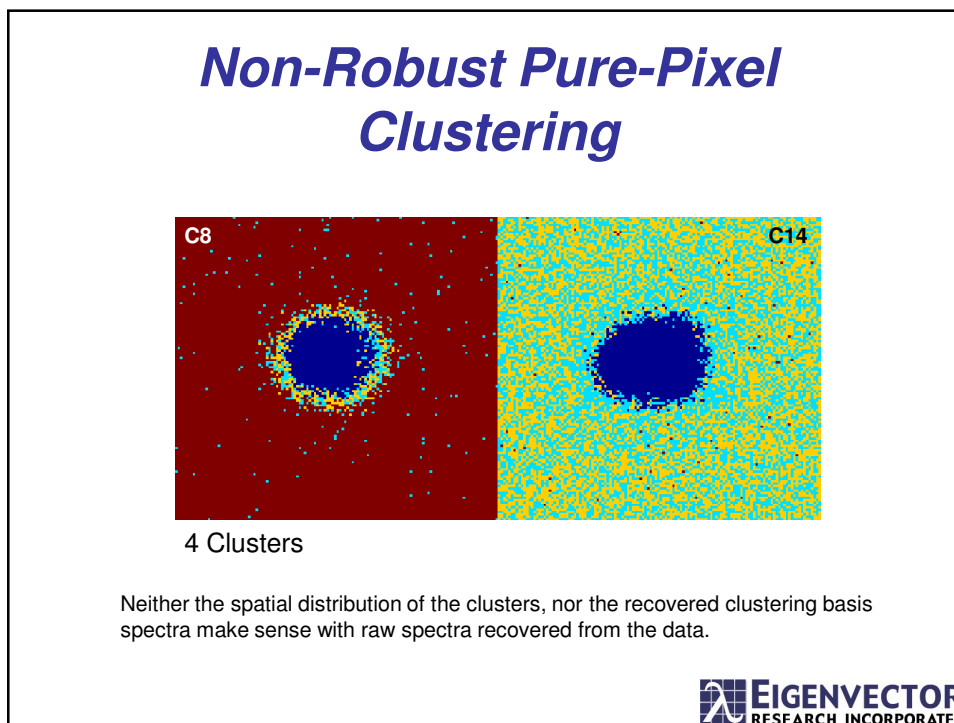
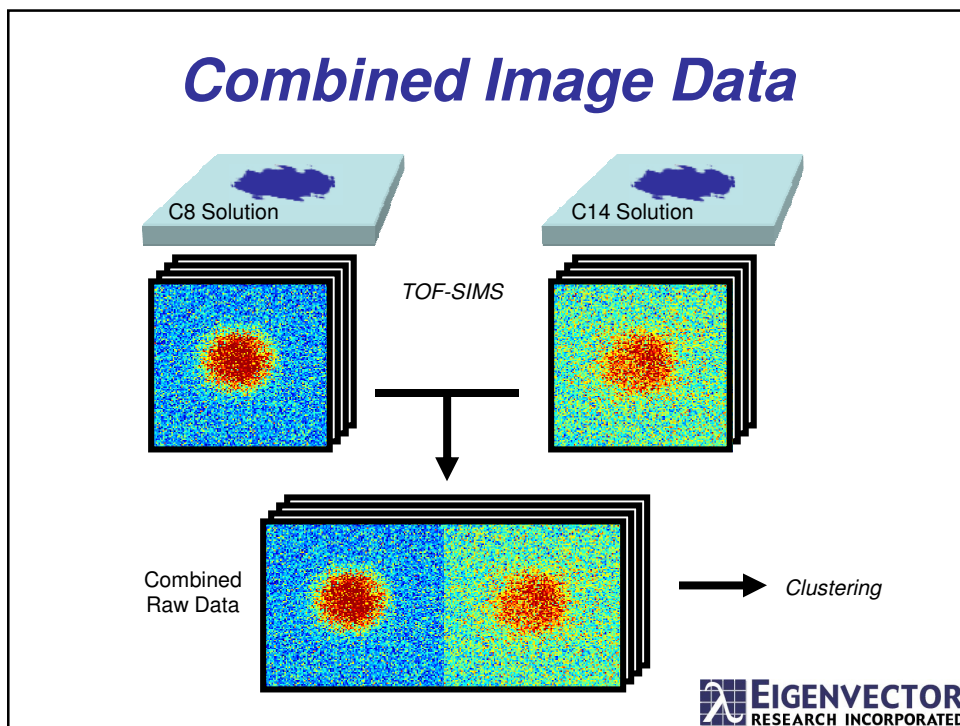


NMR Image of Brain

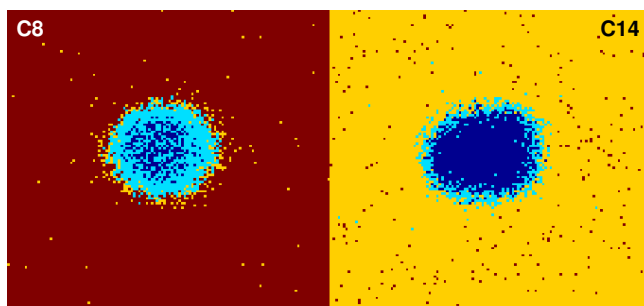


Laser-Activation Modification of Semiconductor Surfaces (LAMSS)





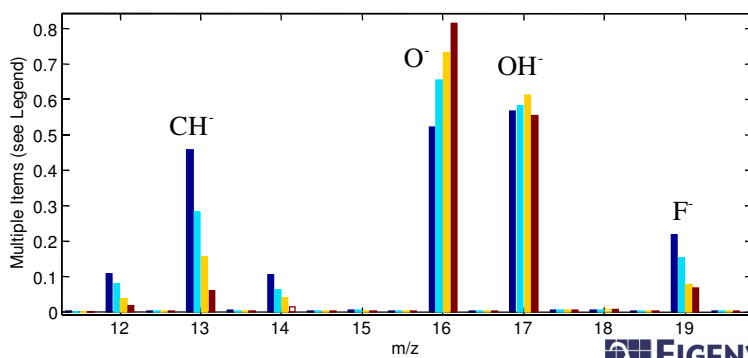
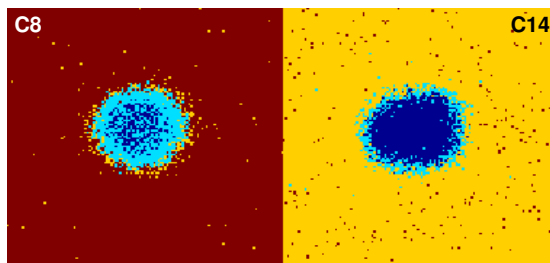
Robust Pure-Pixel Clustering



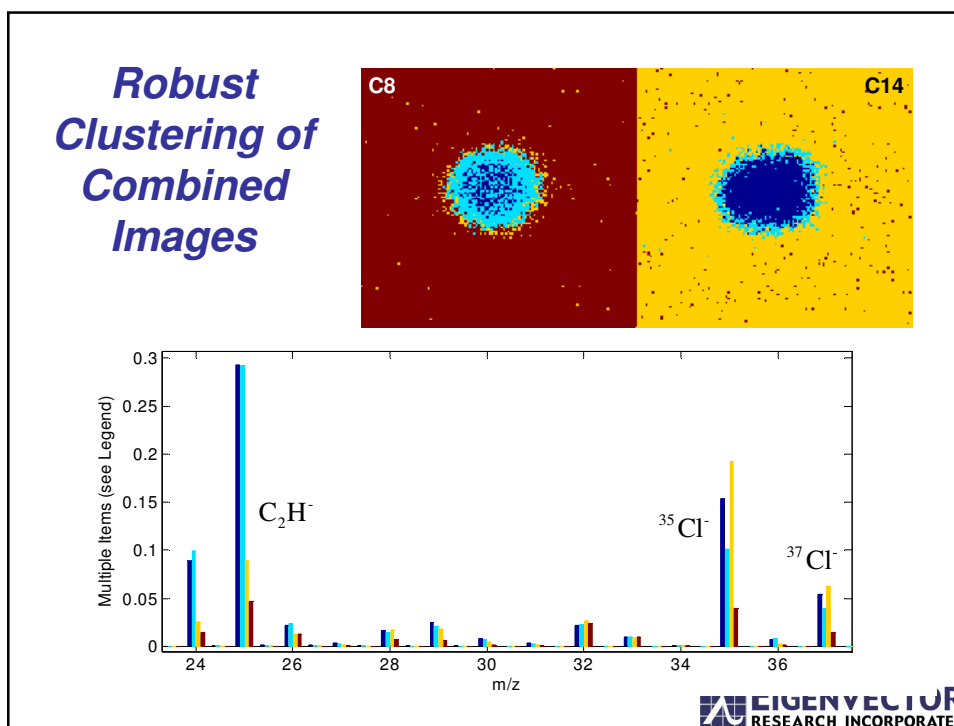
4 Clusters

 **EIGENVECTOR**
RESEARCH INCORPORATED

Robust Clustering of Combined Images



 **EIGENVECTOR**
RESEARCH INCORPORATED



Conclusions

- Pure-Pixel Robust k-Means a fast method for clustering (can also use results as even better initial guess for MCR!)
 - Other robust methods on images
PLS_Toolbox and LIBRA toolbox