Hyperspectral image analysis

Review Presentation

Robert P.W. Duin,
Pavel Paclik,
Serguei Verzakov

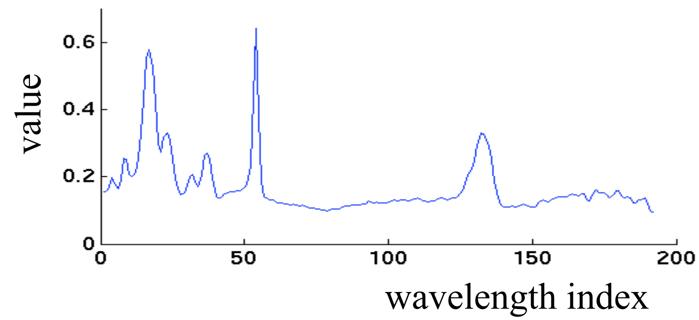
February 2005

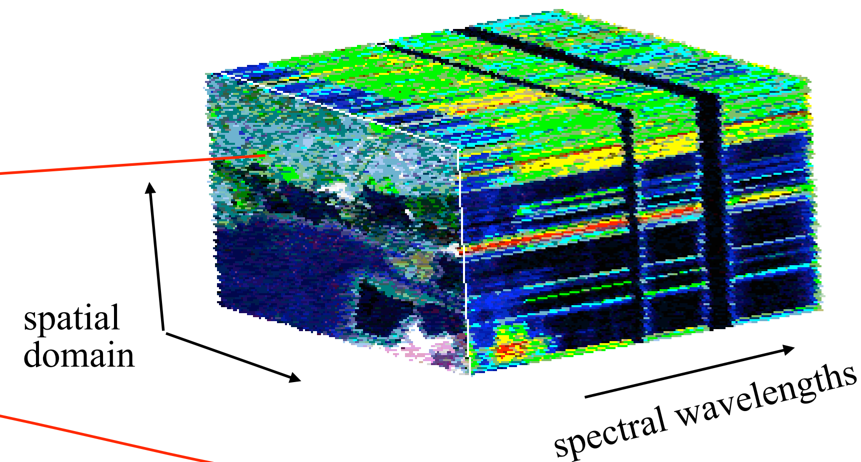
Outline

1. Intro hyperspectral imaging
2. Supervised and unsupervised analysis
3. Spectral Representation
4. Spatial Representation
5. Combined segmentation
6. Examples
   - *project on geological classification of spectra*
   - *GLDB feature extraction, multi-class GLDB*
   - *SVM band-shaving*
   - *sprectra unmixing (factor analysis)*
   - *tangent kernels and invariances*
7. Hypertools toolbox

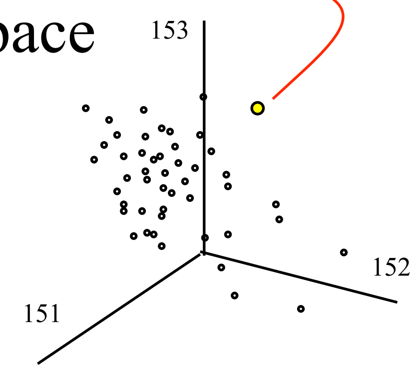
Hyperspectral images?

- spectral connectivity in each spectra
- spatially connected image data





- representing a spectrum by a point in 180D space
  we neglect our apriori knowledge
- properties such as shape or existence of
  a peak must be reconstructed using many
  training examples



How to use apriori information about data connectivity
in a building of data representation?

Hyperspectral images - in remote sensing

...... 10 ...... 35 ...... 60 ...... 85 ...... 110 ......

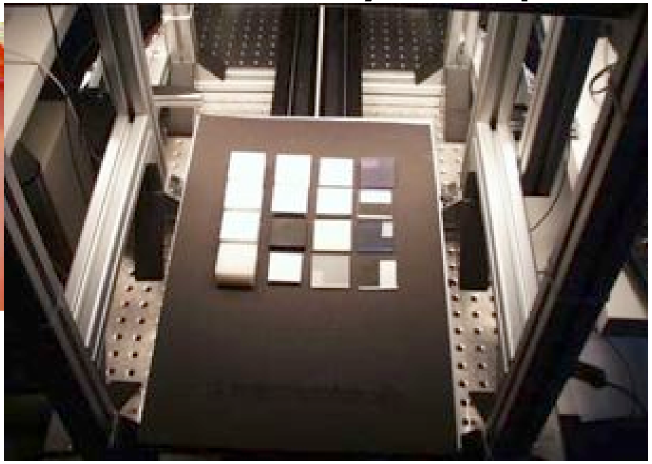
wave numbers

Not anymore just in remote sensing ...

HyperspectralImage Examples

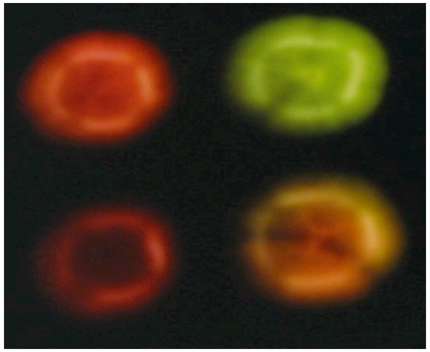
plastic sorting applications...

laboratory setup

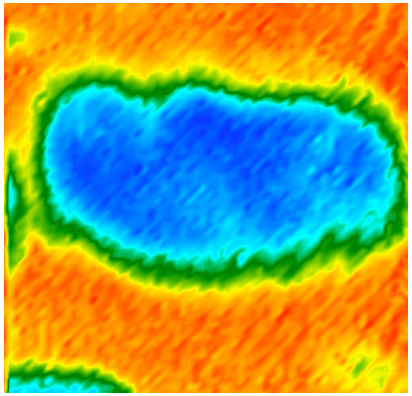


flower variety characterization

estimating tomato ripeness



microstructure characterization in detergents



What are our questions and goals?

We think connectivity is important apriori knowledge that should be exploited in building pattern recognition systems.

What are the benefits of incorporating data connectivity in a representation?

Could we use less training examples to reach the same performance?

How to use the expert's knowledge about spectral shape in defining the representation?

How robust is the representation based on the notion of spectral shape?

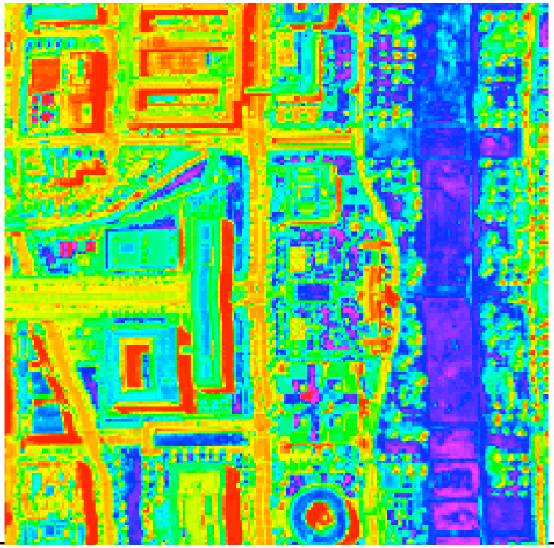
How to use spatial connectivity in clustering or classification of spectral images?

Supervised - unsupervised analysis

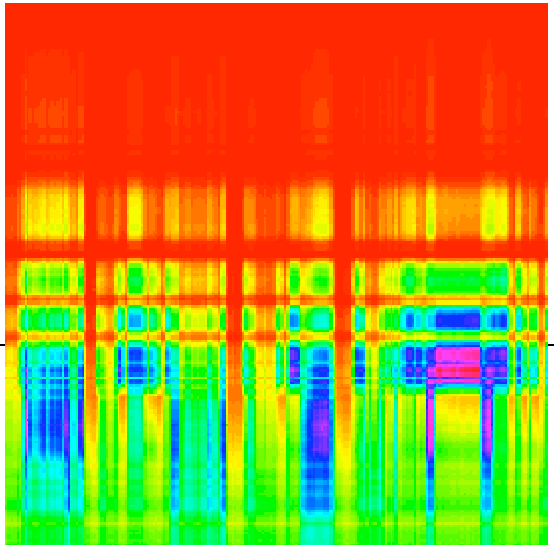
- *Supervised:* Analysis on labeled parts of an image
- *Semi-supervised:* Combine with unlabeled parts
- *Unsupervised:* Analysis on unlabeled pixels only (cluster analysis)

DC_Mall Hyper-spectal Image

Example of single band

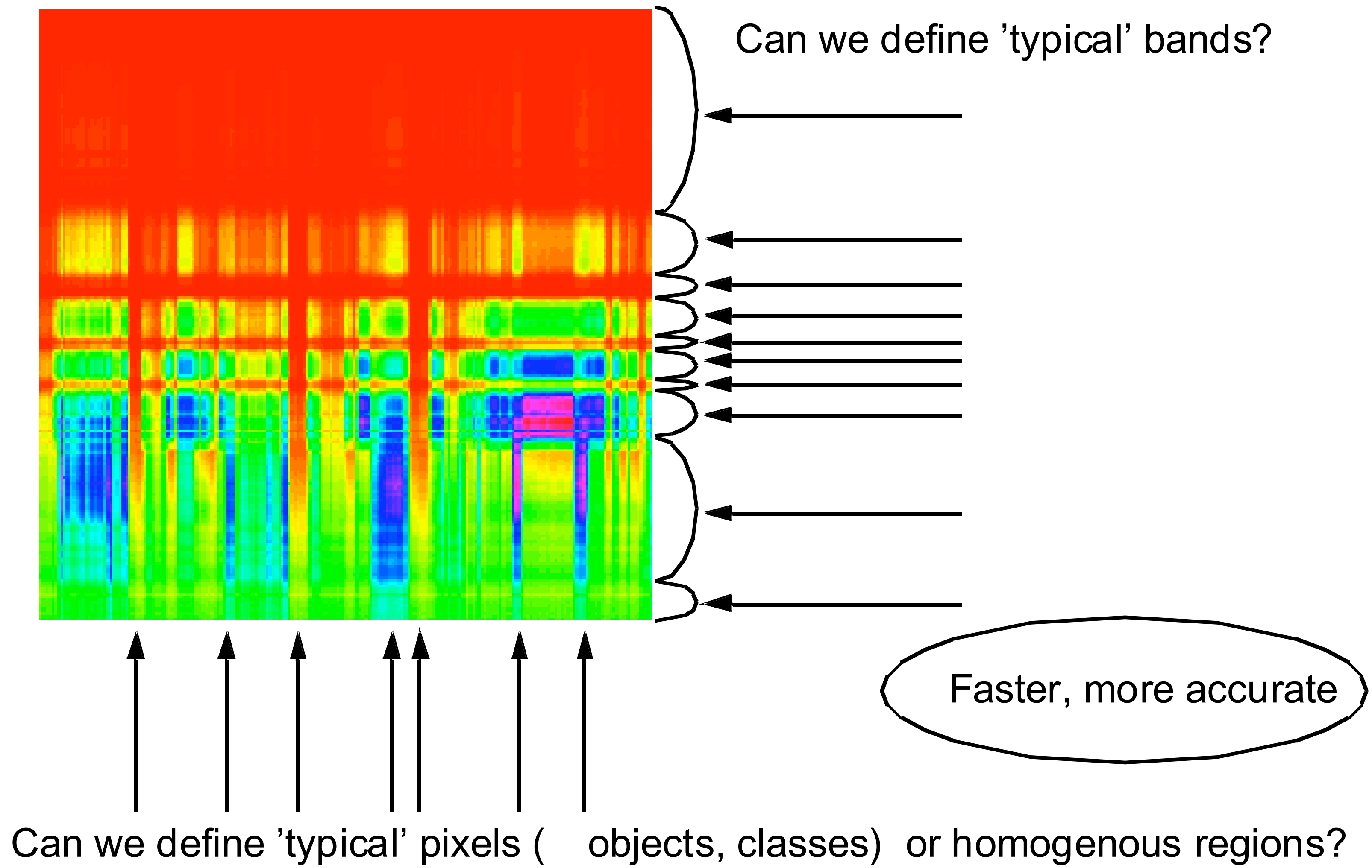


Same in false color



Spectra of bottom line

Spectral and Spatial Analysis



Spectral Analysis

- Reduction:
  - find bands based on spectral connectivity and average
  - useful, when designing a multi-spectral system
- Extraction:
  - find bands and weight wavelengths inside each band
    - Single band: GLDB (linear), Kernel (non-linear)
    - Multi-band: SVM band-shaving
    - Trainable prototype-based similarities

Spatial Analysis

- Prototype selection:

  - Dissimilarity based (random, clustering, feature selection)

  - Pure pixels

- Segmentation:

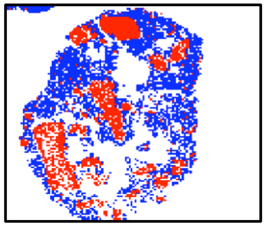
  - Spectral features

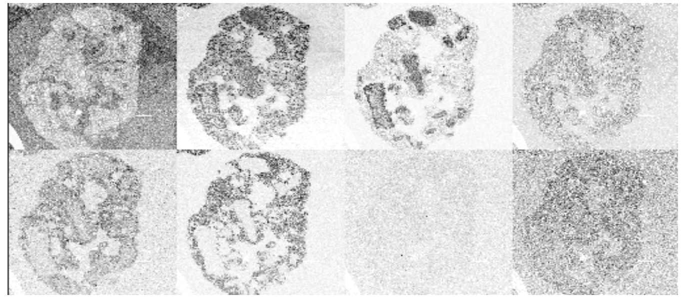
  - Spatial features (textures), possibly per band

  - Combined

Washing Powder Example

Initial noisy labeling

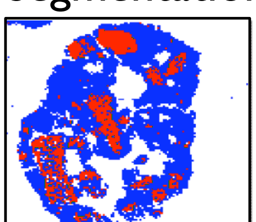


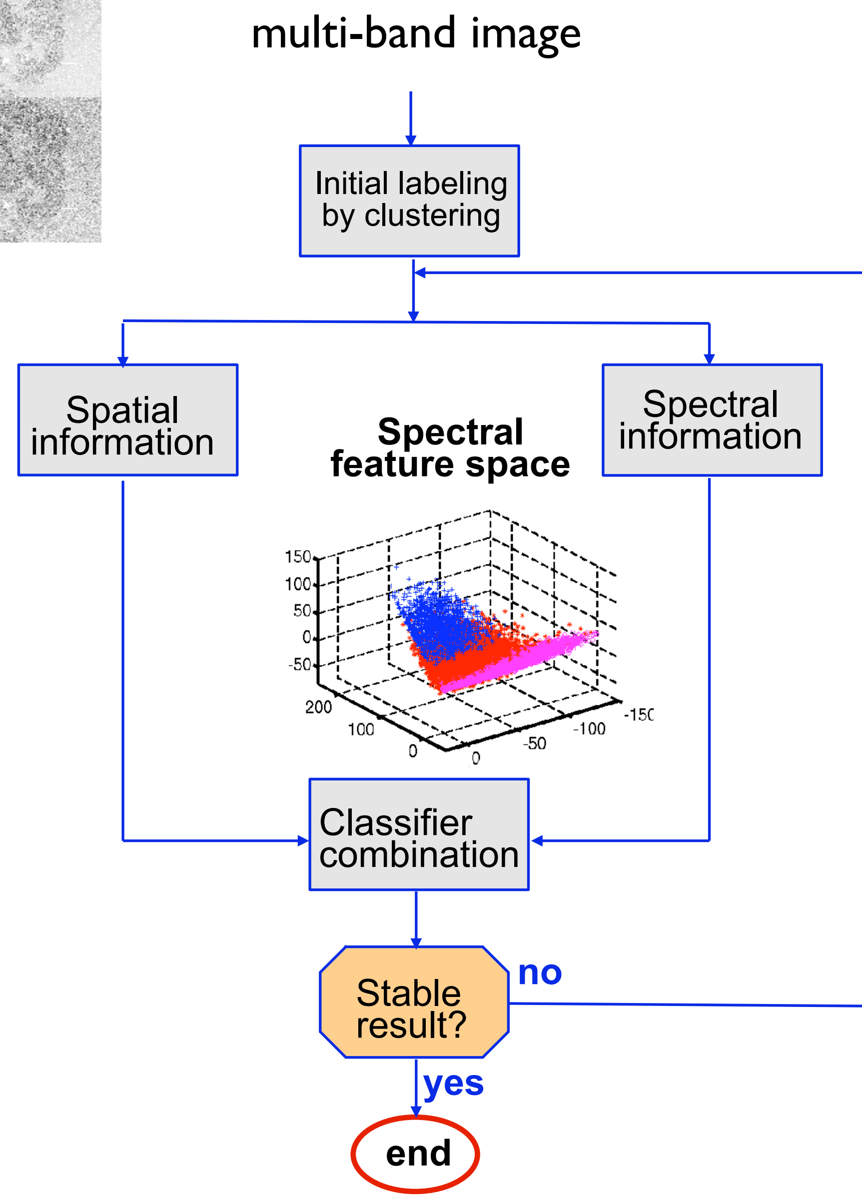


Binary class images

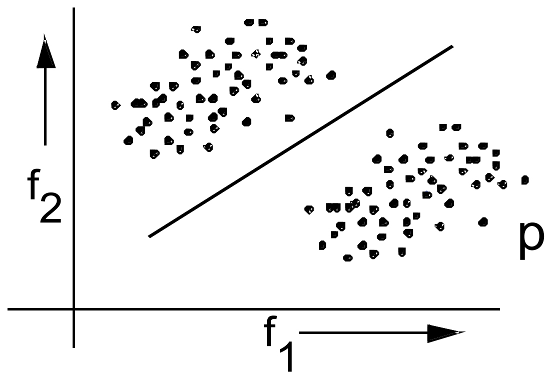
Local spatial information by convolution

Segmentation result





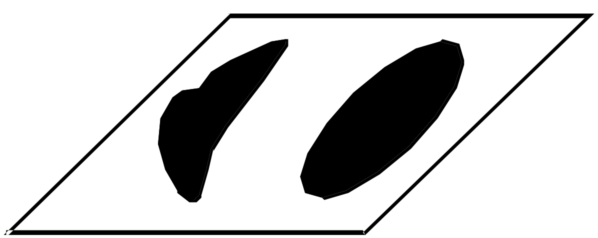
Combined Classifier Approach



Feature Space

pixels

Normal Density based Bayes Classifier



Image

Combine

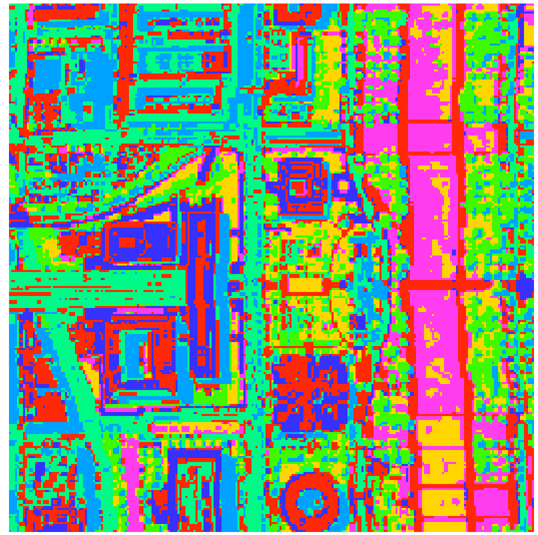
Use EM Clustering

Can spatial domain be treated as a feature space?

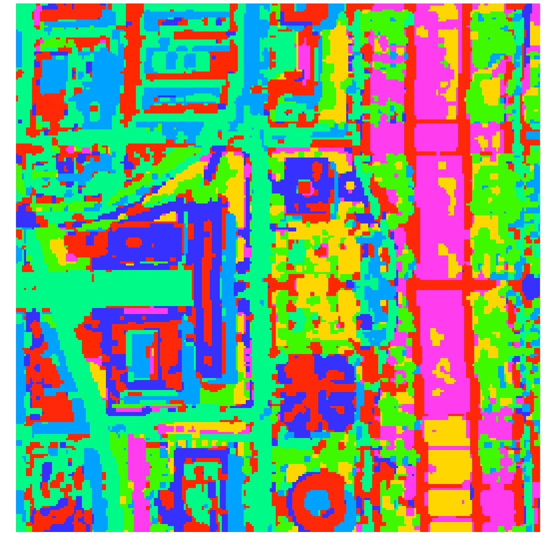
Parzen Classifier

DC_Mall Example

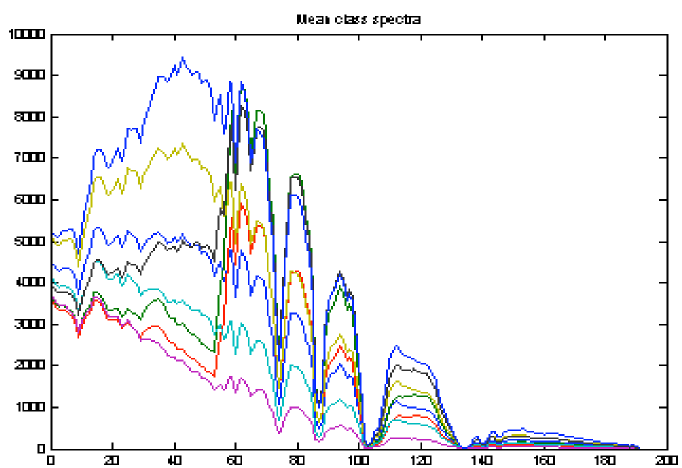
Example of single band



Spectral Clustering

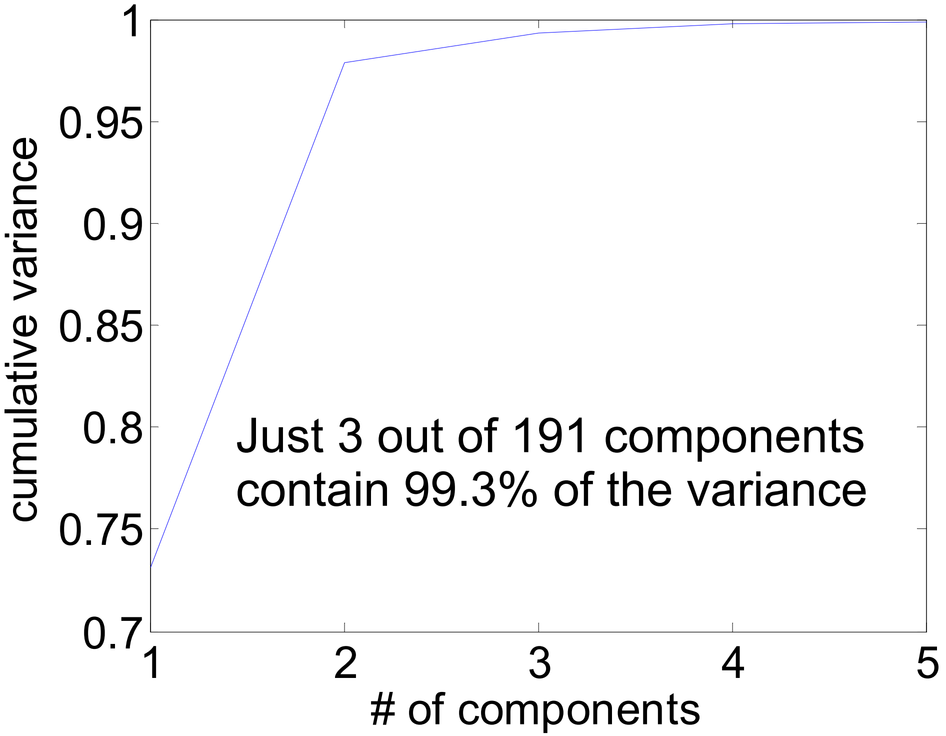


Spectral/Spatial Clustering

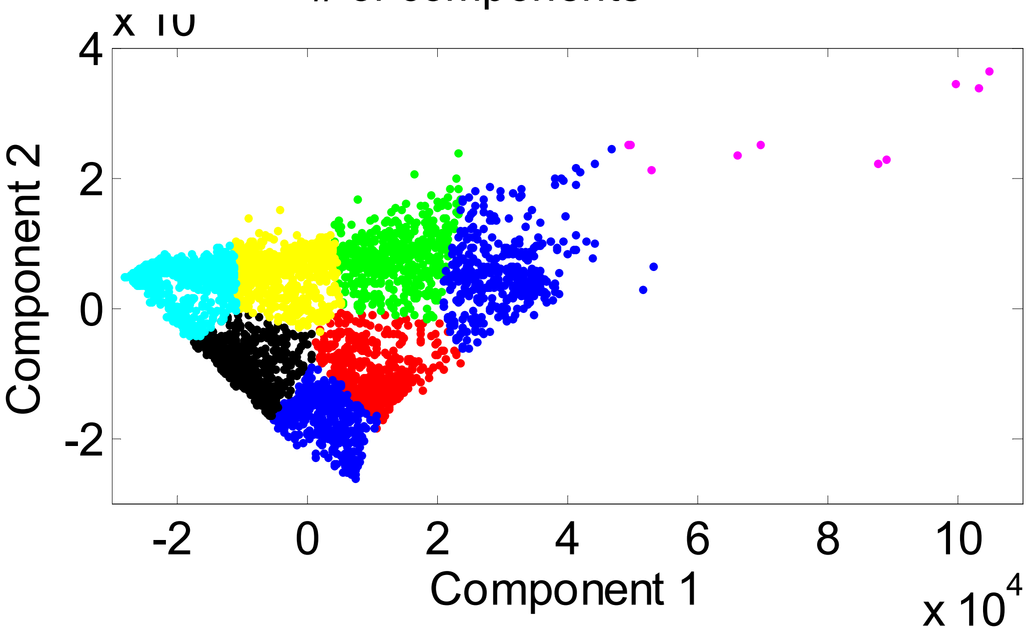


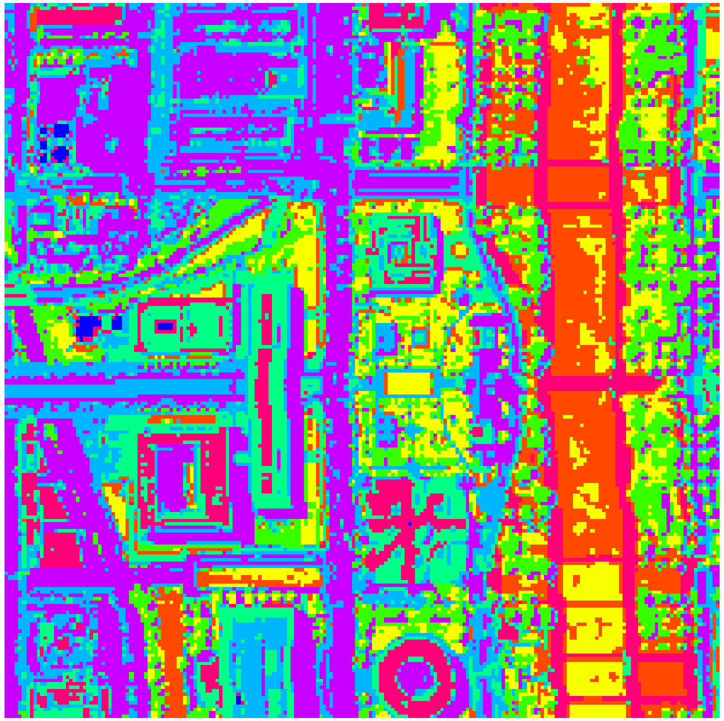
Spectra of class means

PC Analysis



The first 3 eigenimages





The labeled image

Object Classification

- Labeling of an entire image region
- Spectral information
- Shape
- Composition

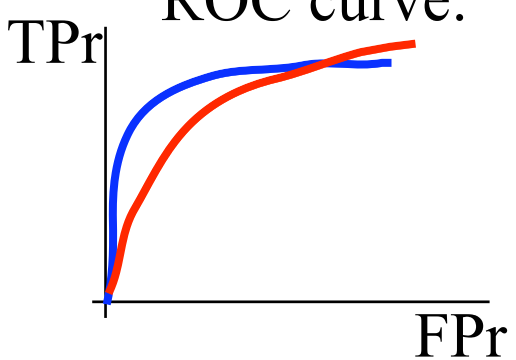
Project on analysis of geological data

- geological exploration problem (classification of spectra)
- large dataset (thousands of examples, tens of classes)
- heavily skewed distribution (very rare target minerals vs abundant non-targets)
- misclassifications are costly
- fast processing needed (single batch contains cca million of grains)
- applied project: how to build such a classification system?
- many scientific challenges
- five work packages:
  - WP1: setup of evaluation methodology, testing classical methods
  - WP2: building spectra-specific data representations
  - WP3: one-class classification
  - WP4: classifier combining
  - WP5: active learning
- our output: toolbox for prototyping and technical reports

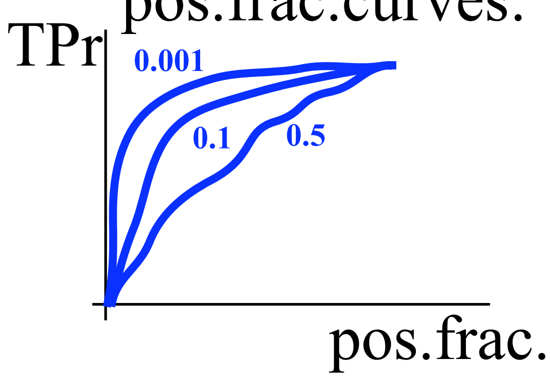
Geological project - research challenges

- usual assumption: *training dataset is representative*
- in our project:
  - training on large balanced dataset, testing on extremely skewed (unknown priors)
  - how to evaluate such classification systems?
- our research:
  - it is not enough to look at average error or ROC curves alone
  - additional operating characteristics are needed (positive fraction)

ROC curve:

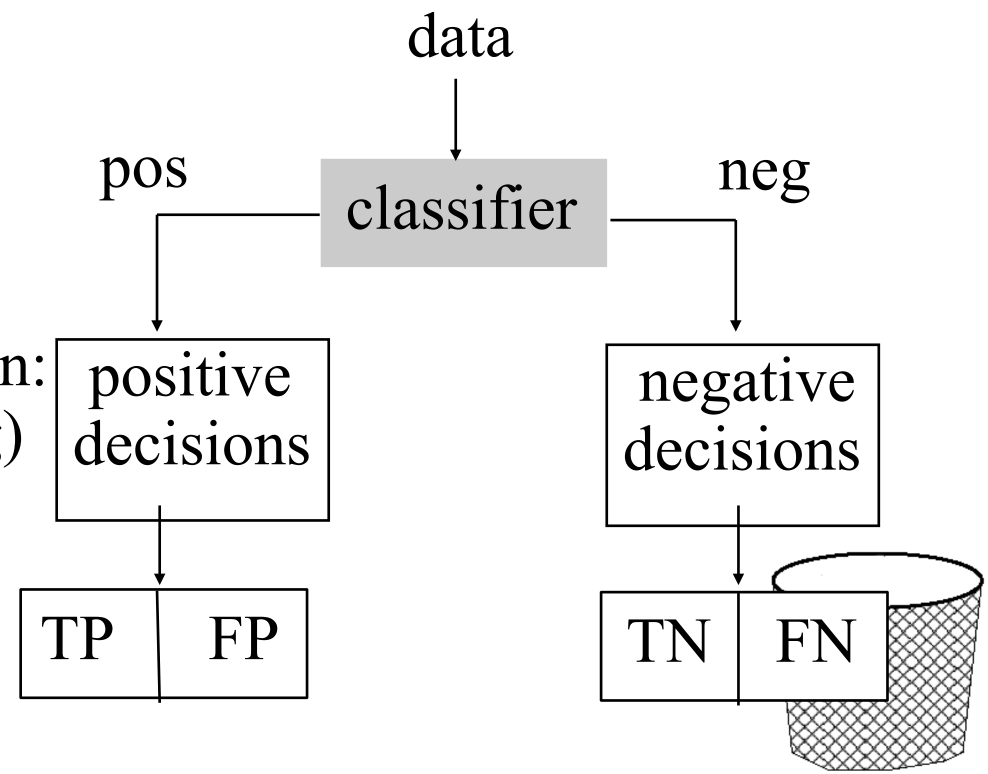


pos.frac.curves:



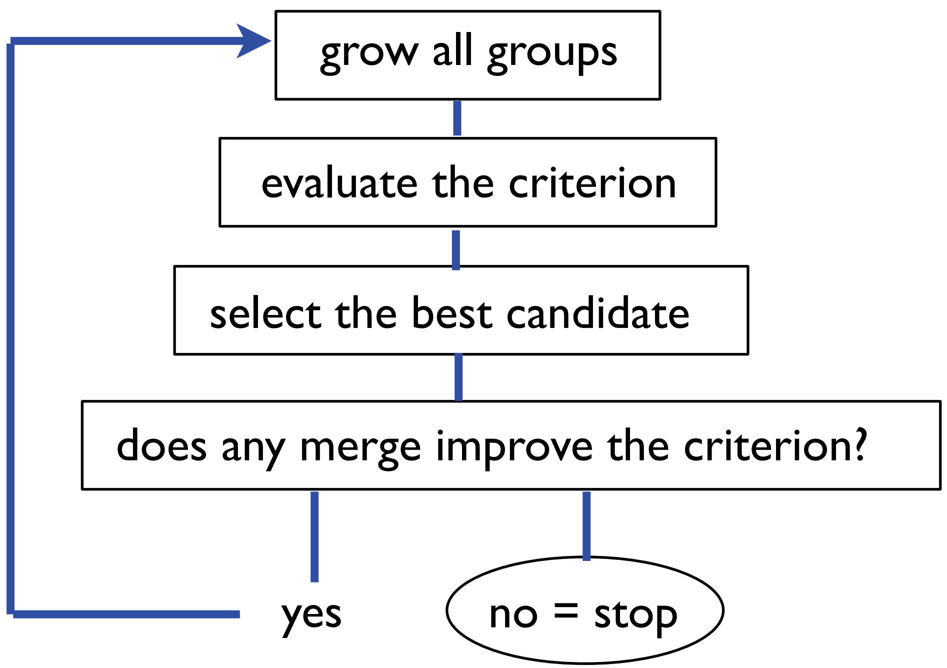
Purity:
TP/pos

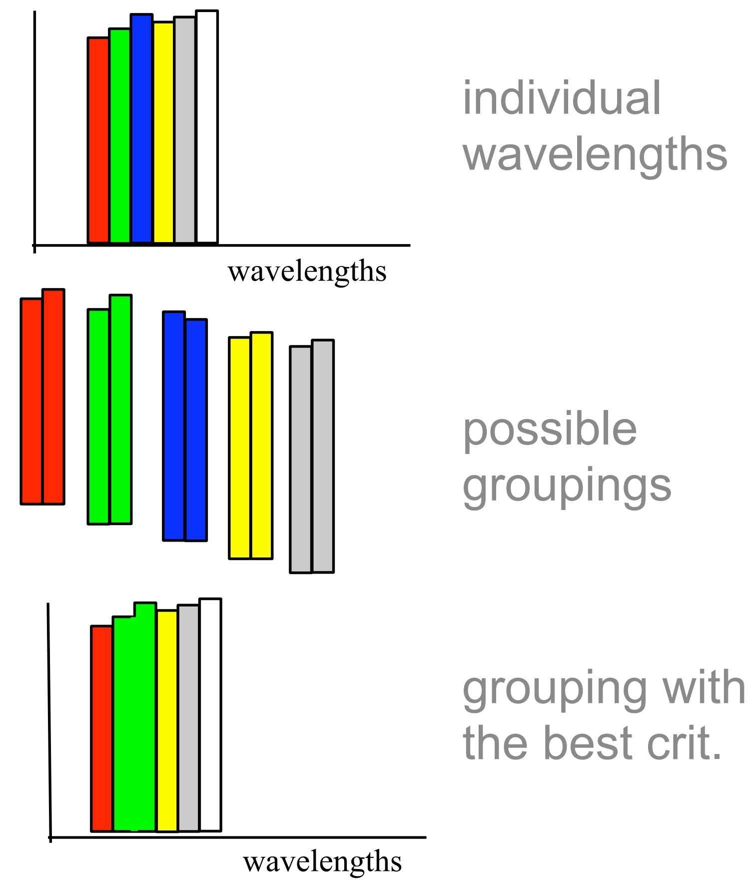
Positive fraction:
pos/(pos+neg)



GLDB feature extraction

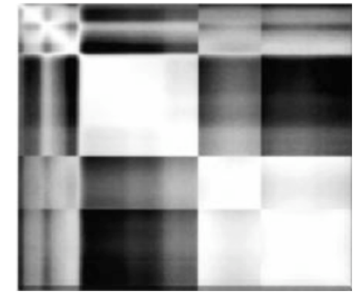
- Generalized Local Discriminant Bases by Kumar, Ghosh, and Crawford
- extracts low-dimensional feature space from spectral data
- identifies non-overlapping groups of wavelengths
- top-down or *bottom-up* approach
- bottom-up starts with all singleton wavelengths, top-down starts from full spectra





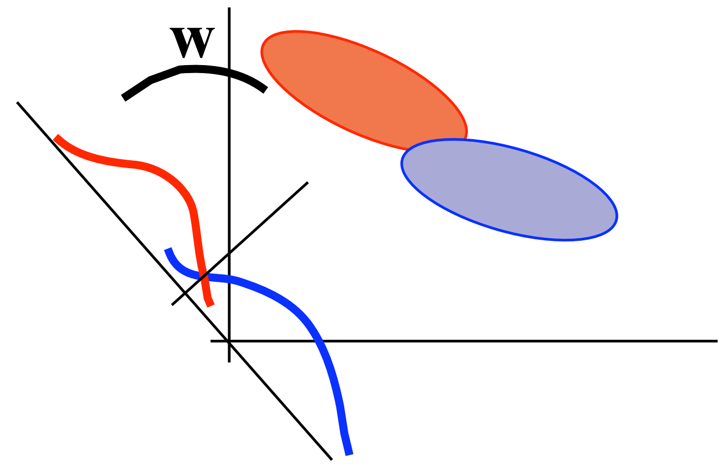
GLDB feature extraction - criterion

- *correlation* between wavelengths



$$C(\text{group}) = \min_{i,j \in \text{group}} \text{corr}(i,j)$$

- *separability* between classes using Fisher criterion



$$D(\text{group}) = \frac{\mathbf{w}'_{\text{group}} \mathbf{B}_{\text{group}} \mathbf{w}_{\text{group}}}{\mathbf{w}'_{\text{group}} \mathbf{W}_{\text{group}} \mathbf{w}_{\text{group}}}$$

- combining both by

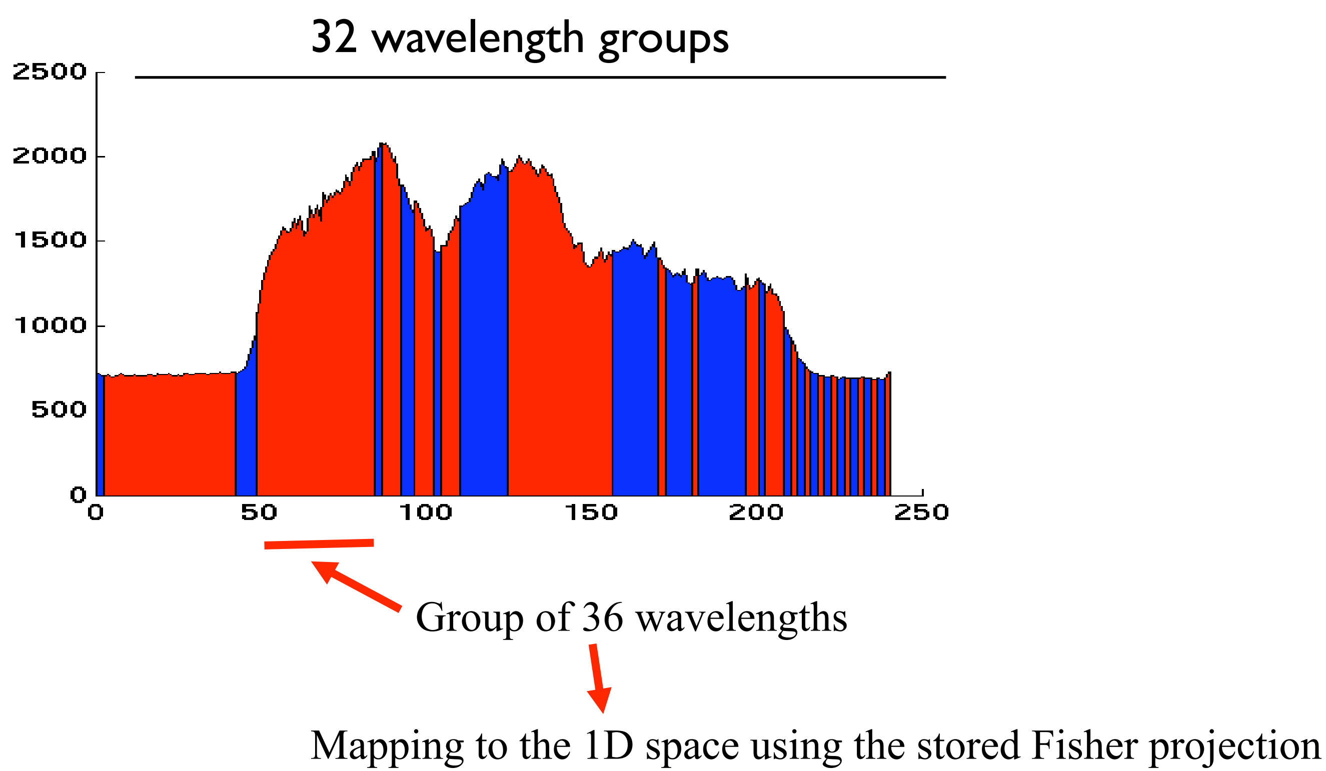
$$\mathcal{I}(\text{group}) = C(\text{group}) D(\text{group})$$

By maximizing this criterion, we choose groups of highly correlated (usually adjacent) wavelengths separating the classes

GLDB feature extraction - applying to new data

- for each wavelength group, Fisher projection vector *w* is stored
- new data are mapped into a lower dimensional space by applying the trained Fisher projection to each group of wavelengths

Example:



- each spectrum is mapped from the original 240D to the 32D space
- because some groups are uninformative, feature selection or extraction should follow

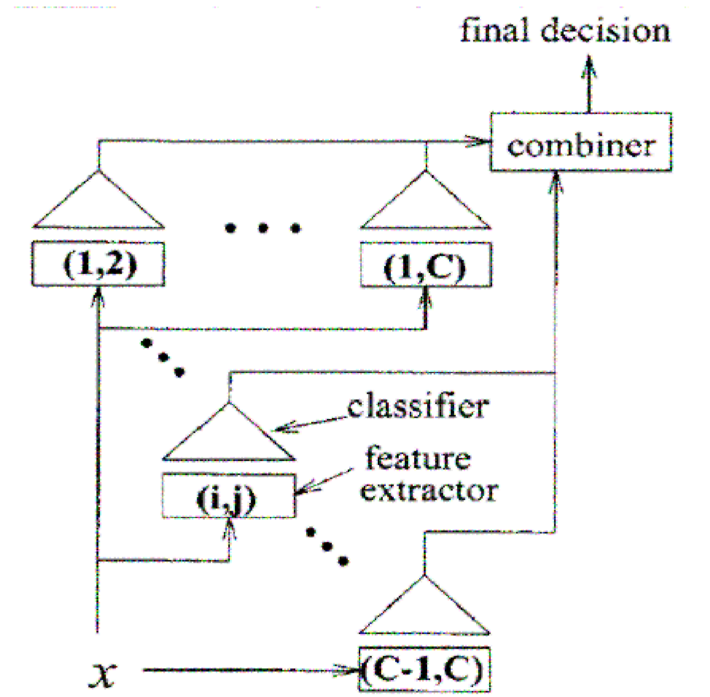
GLDB extraction - original multi-class extension

Kumar et.al. use 2-class criteria and a separate feature extractor for each par of classes

For a C-class problem, you need to train C(C-1)/2 extractors and classifiers.

Unrealistic for larger problems:

17 class problem: 136 mappings and classifiers
42 class problem: 861 mappings and classifiers



- they claim it is advantage to have a different representation for each pair of classes

  We tried to find if it really helps and if not, what is a better multi-class extension of GLDB

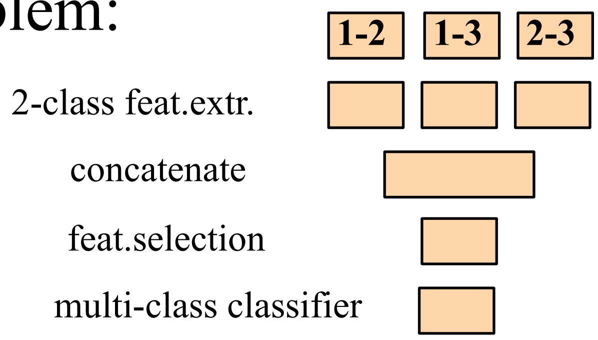
- studying both classification accuracy and complexity in execution

GLDB - proposed multi-class extensions

3-class problem:

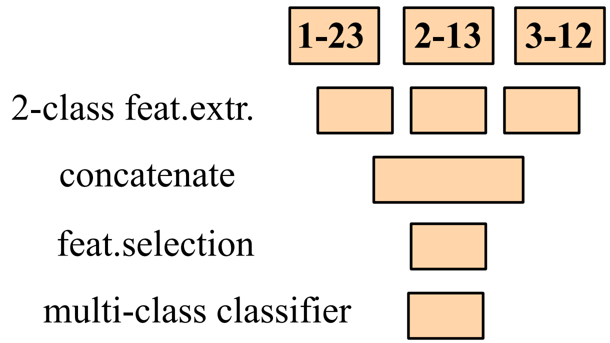
Feature selection

- train all pair-wise extractors
- collect all features
- run 2nd stage feature selection



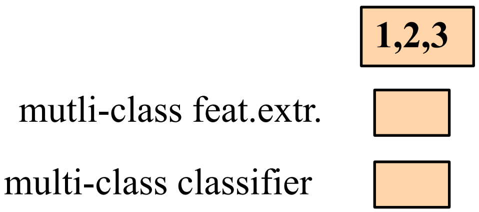
One-against-all

- for C-class problem, train C extractors
- each class vs other classes



Multi-class GLDB criterion

- optimize a multi-class GLDB criterion directly
- more difficult (generalized eigenvalue problem)



- in all cases, a single multi-class classifier is trained

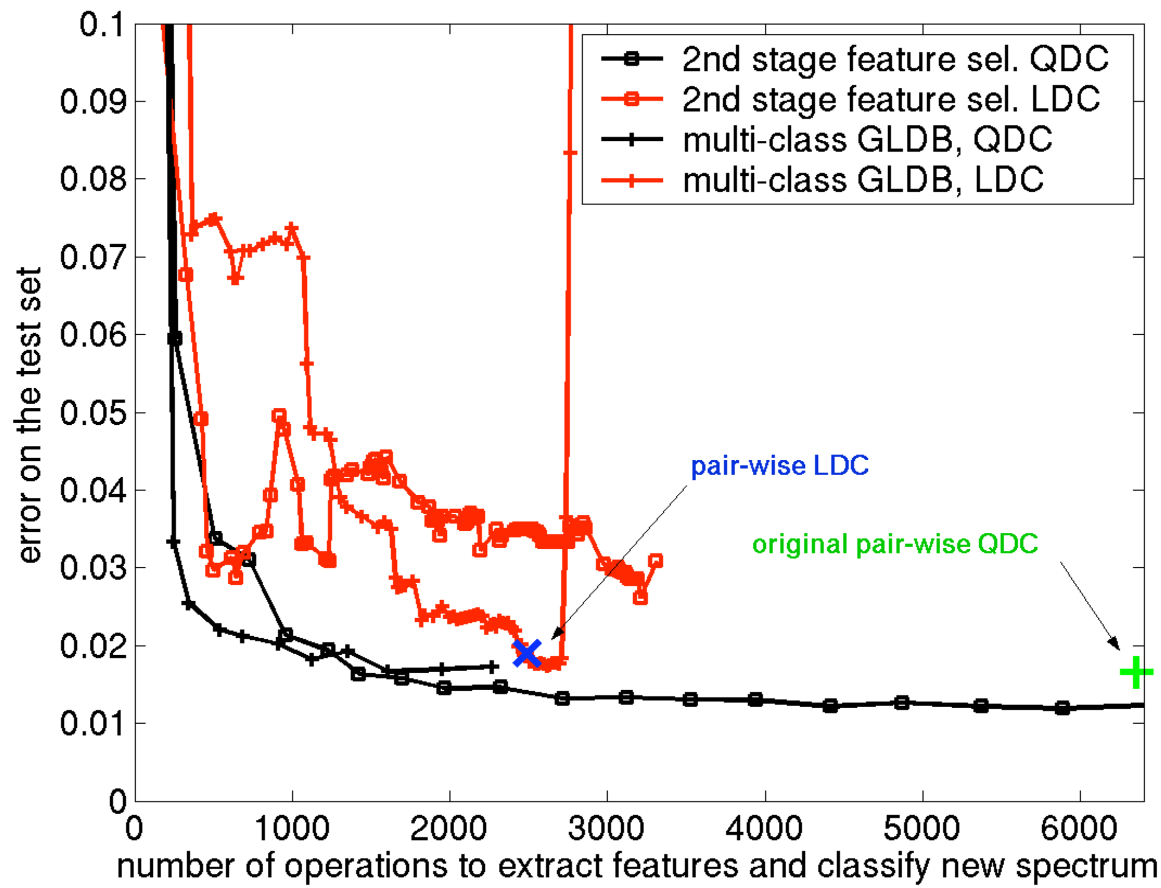
GLDB extraction - performance and speed

- comparing the original pair-wise GLDB to our multi-class extensions
- looking at both classification error and speed of classification

- DC-Mall remote-sensing dataset
  7 classes, 191 wavelengths
  9969 spectra (1400 training, 8569 testing)

- plotting the test error for different subsets in feature selection

- speed of the original pair-wise classifier may be significantly improved



- interesting finding: proposed methods significantly speed-up extraction but the multi-class classifier becomes more computationally expensive

Shaving (recursive feature elimination)

Principle:

Rank each feature $\mathbf{x}_j \rightarrow \mathbf{r}_j$

Remove a group of features with the lowest rank

Repeat until the desired performance and/or the number of features is reached

Guyon et al proposed to use weights of linear SVM classifier as a ranking criterion.

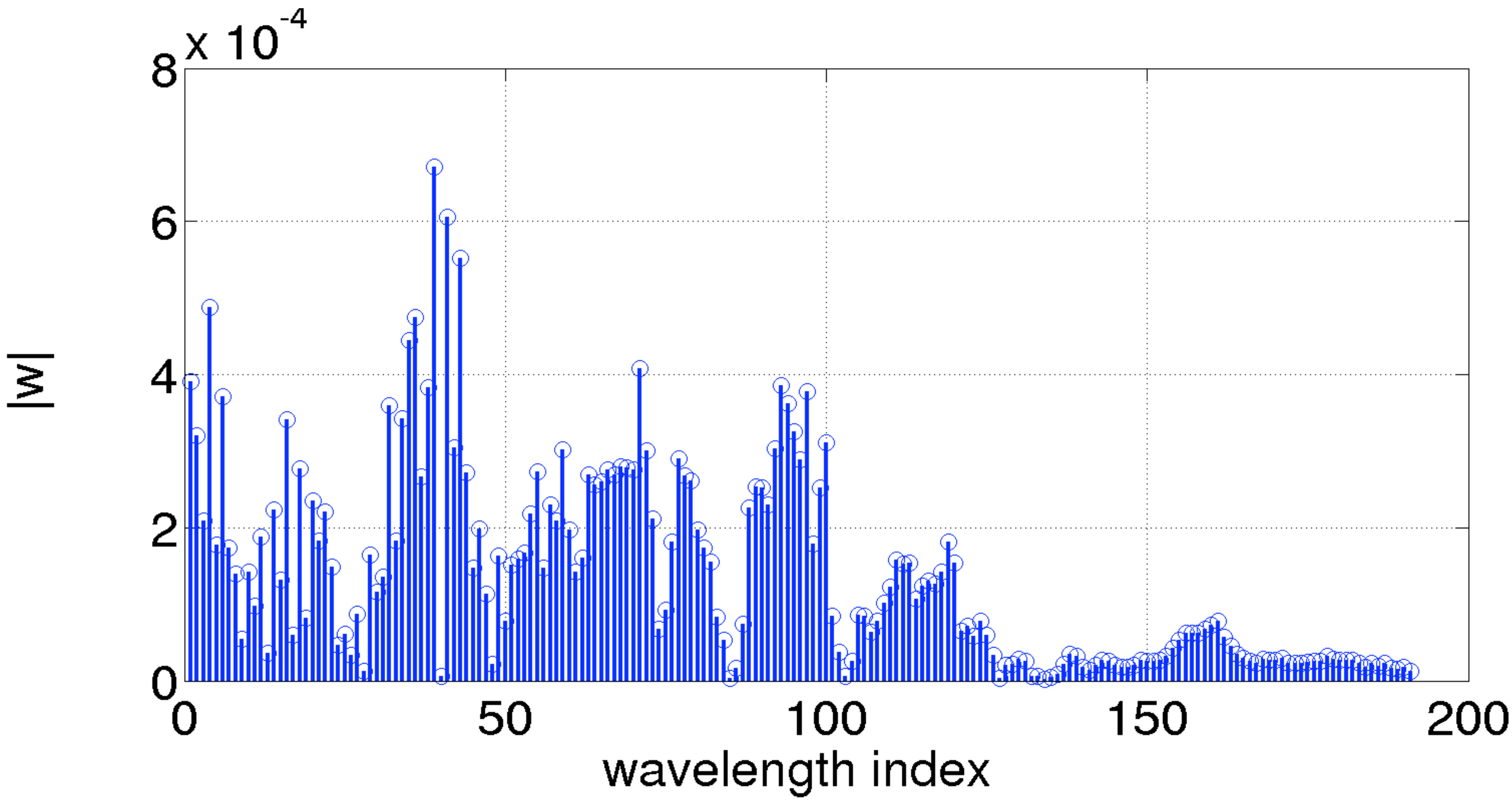
$$f(\mathbf{x}) = \sum_{j=1}^{D} w_j x_j + b \quad \mathbf{r}_j = |w_j|$$

Formally, any linear classifier could be used but SVM has a reputation to be a robust to the curse of the dimensionality.

We proposed **SVM band-shaving** algorithm which exploits connectivity in hyperspectral data (high correlation between neighboring wavelengths).

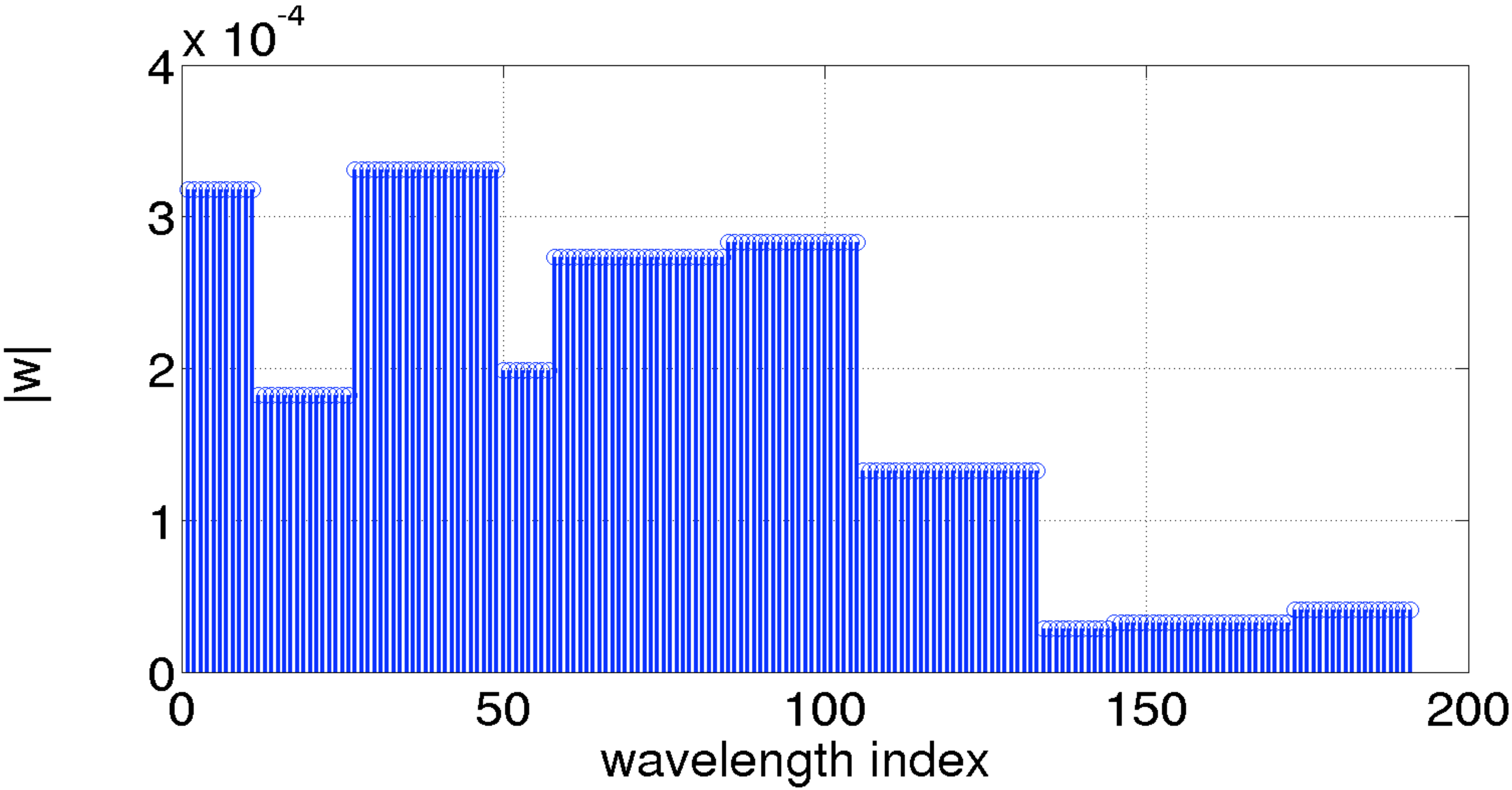
SVM band-shaving

As in the original algorithm, we start from the training SVM classifier applied to the whole feature set.



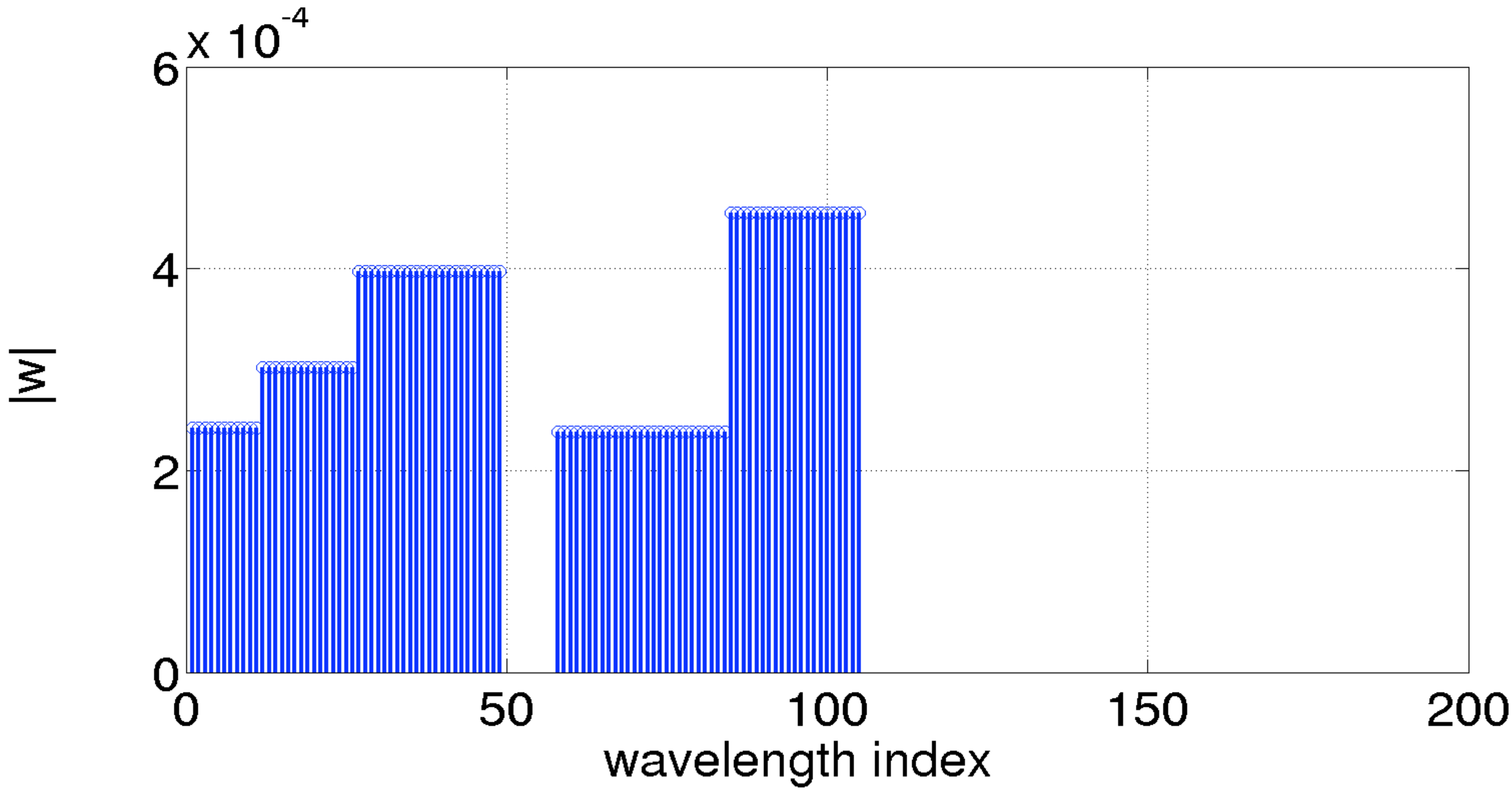
SVM band-shaving

But then we use w profile to extract informative bands (peaks separated by local maxima).



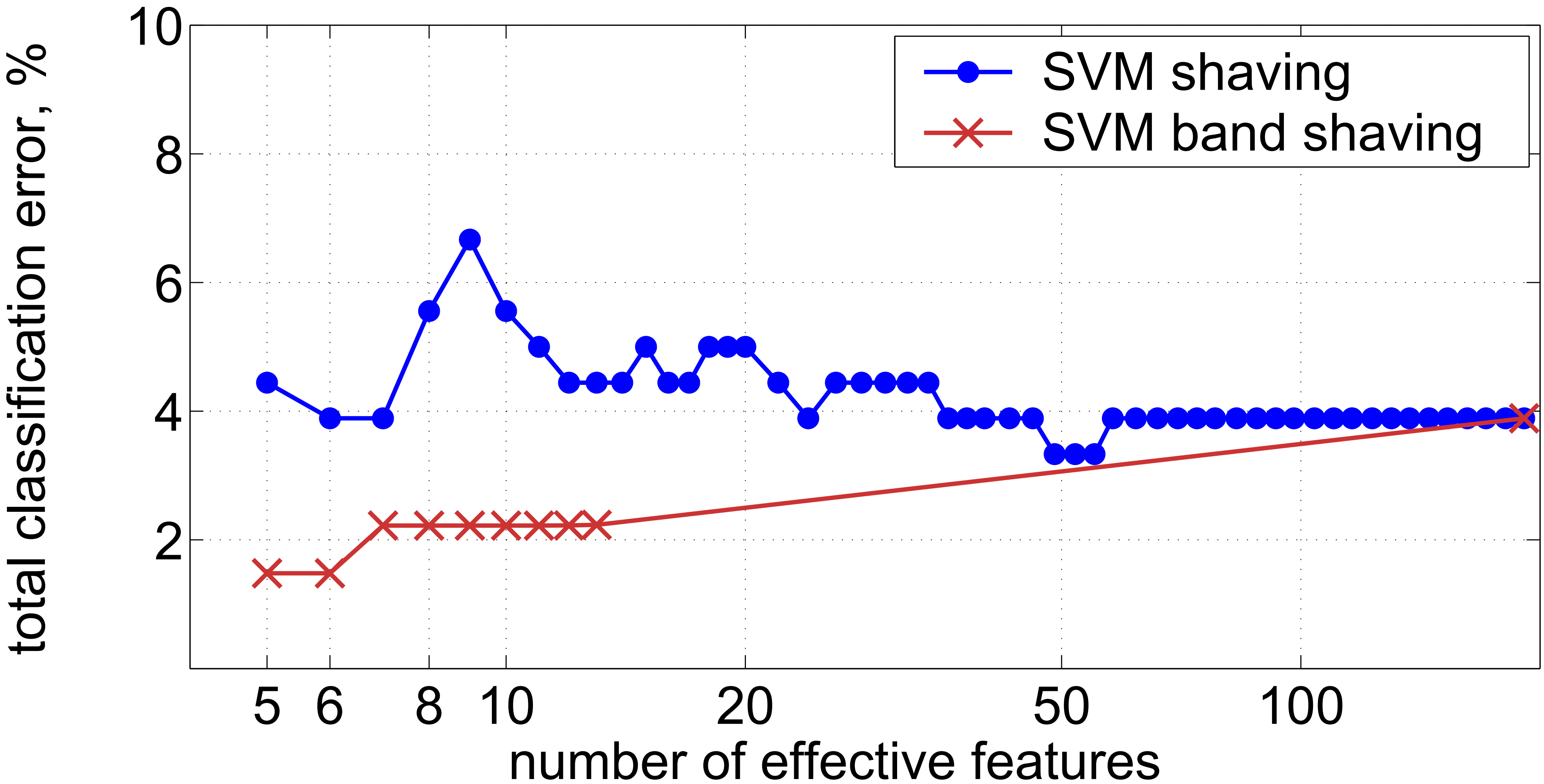
SVM band-shaving

After this we perform standard SVM shaving on the set of new features.



SVM band-shaving - results

SVM band shaving may outperform the standard technique



Spectral unmixing

In hyperspectral images each pixel can contain response from a few (maybe unkown) substances:

$$I(\mathbf{x},\mathbf{y};\nu) = \sum_{i=1}^{m} C^i(\mathbf{x},\mathbf{y})S^i(\nu)$$

pixel unmixing is necessary.

Unmixing techniques

PCA
Finds a subspace

VARIMAX
Rotates axes to the position in which they are most similar to the original spectra

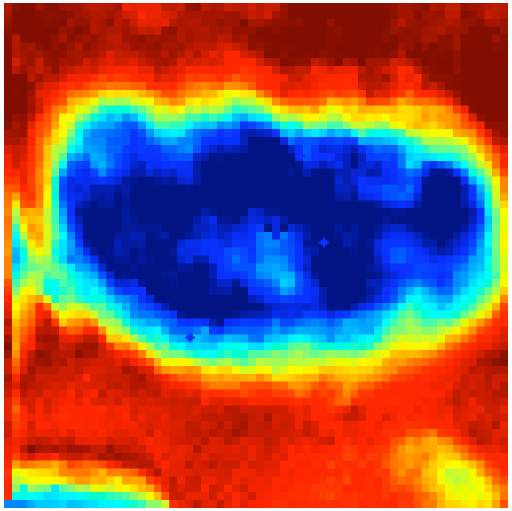
OPA/SIMPLISMA
Looks for the puriest spectra/wavelengths

Alternating Least Squares
Performs least squares with constraints: spectra and concentrations are positive

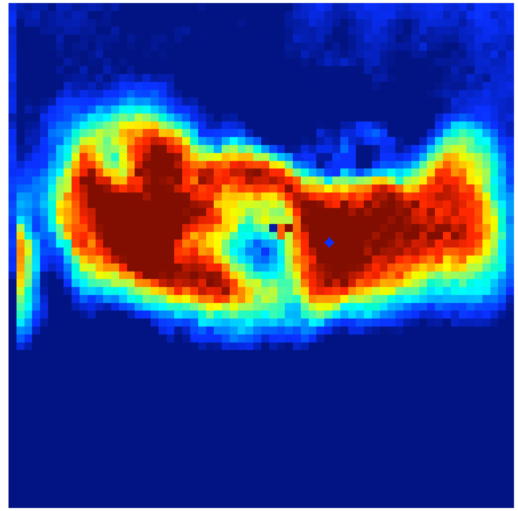
Unmixed spectra and concentrations

Example: OPA-ALS on detergent images

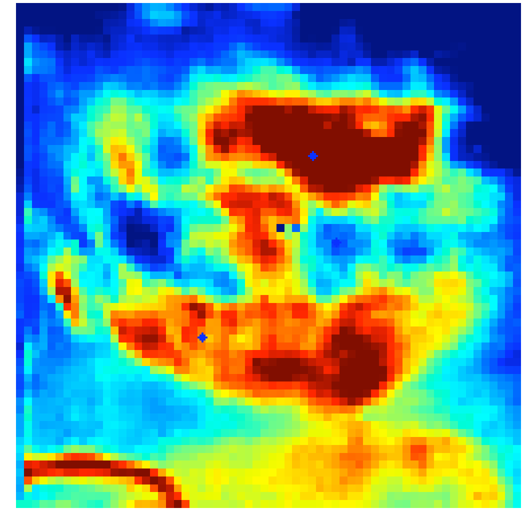
Factor 1



Factor 2

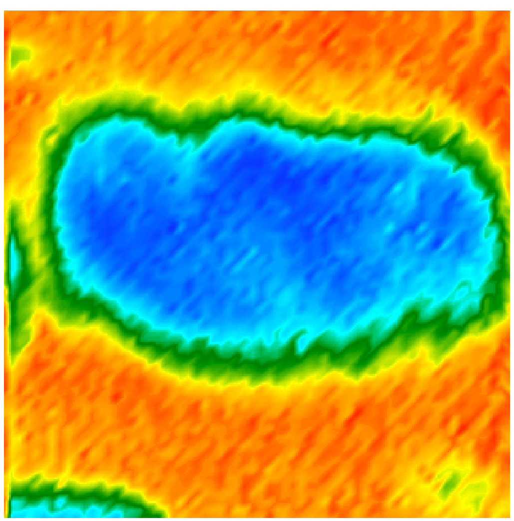


Factor 3

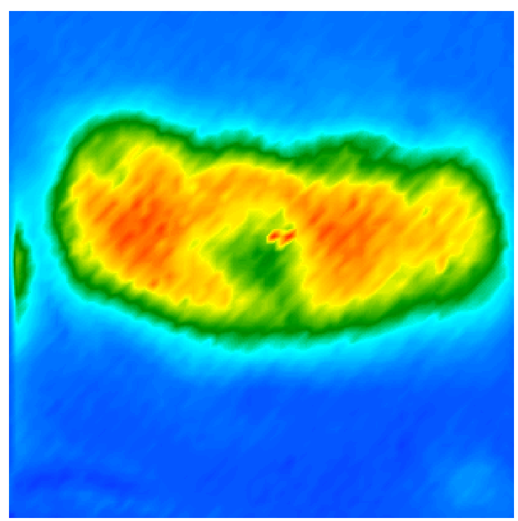


Ground-truth based on the expert knowledge of material chemistry:

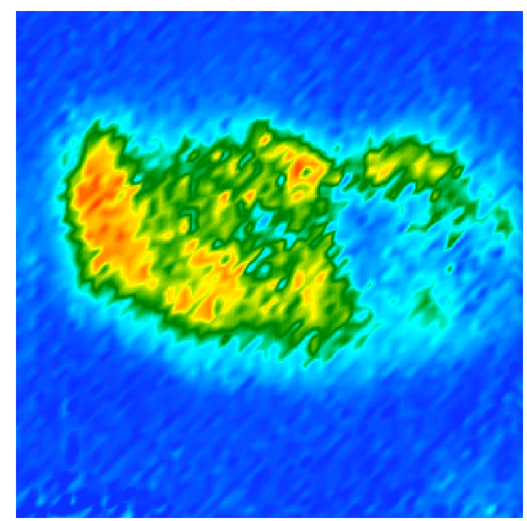
Embedding material



Class 1



Class 2



Tangent Kernels and Invariances

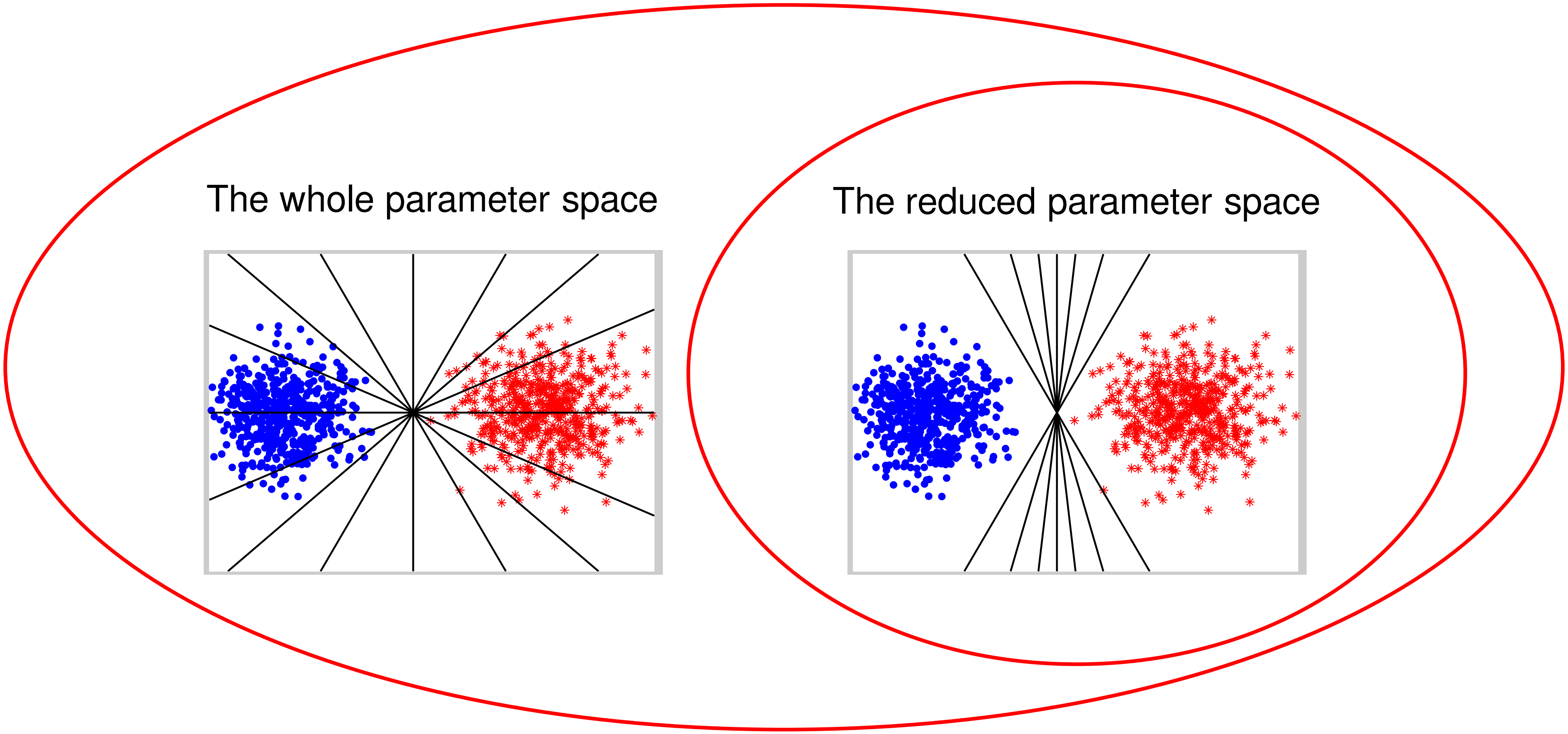
Ways to get a good performance.

One must collect a lot of training data

or

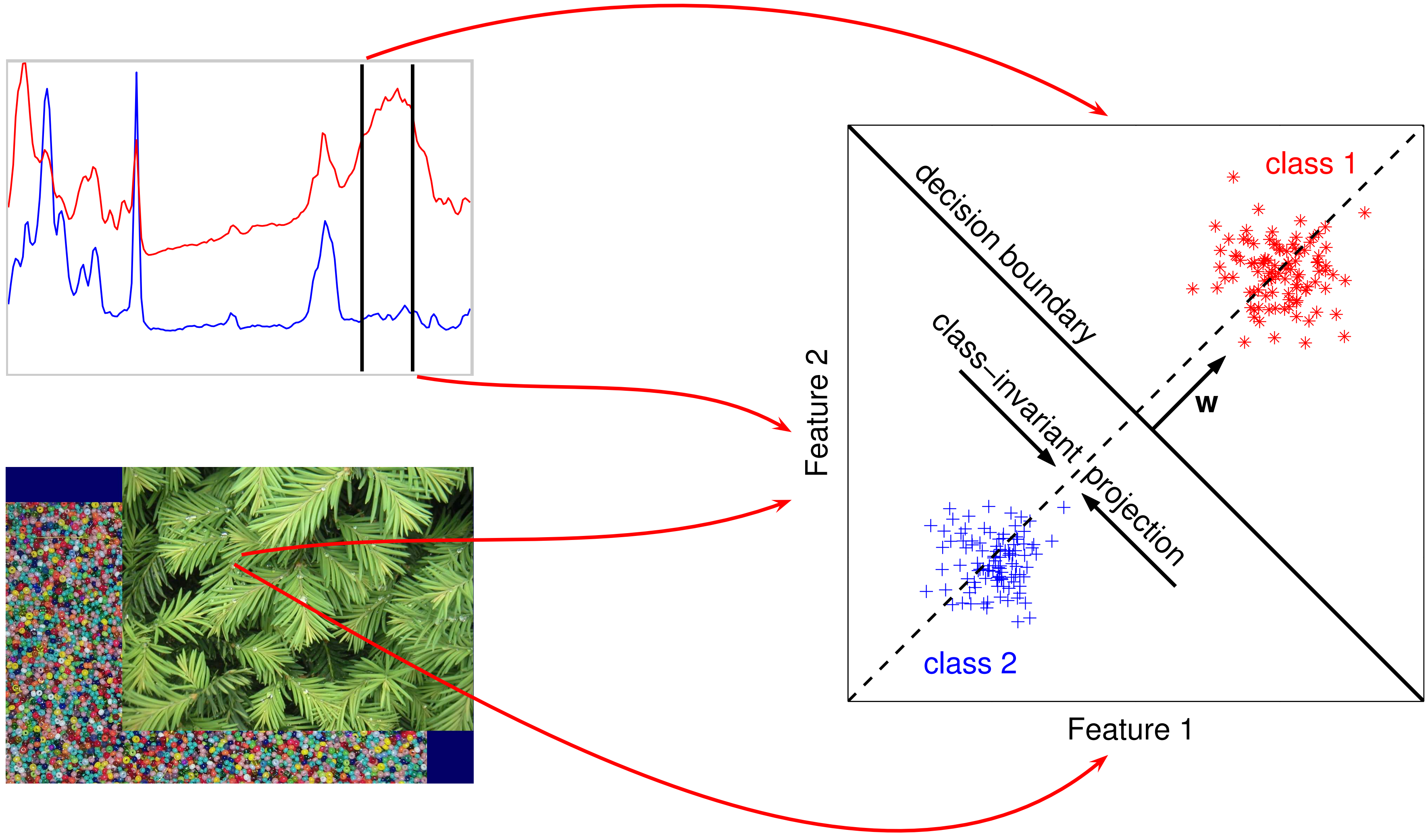
use a prior knowledge!

How does a prior knowledge work?



Tangent Kernels and Invariances

Connectivity in spectral and spatial domain.



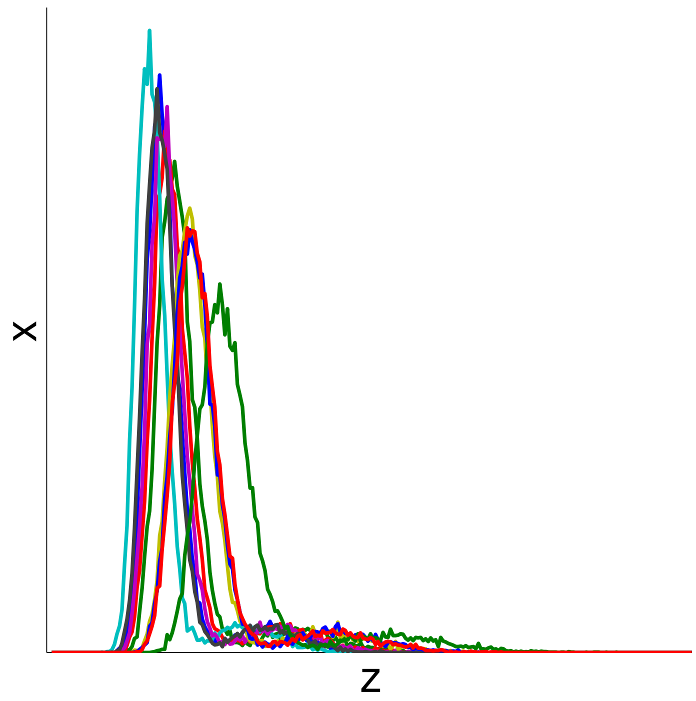
Tangent Kernels and Invariances

Information about calibration imprecisions.

Affine calibration imprecisions:

$\tilde{z} = az + b$

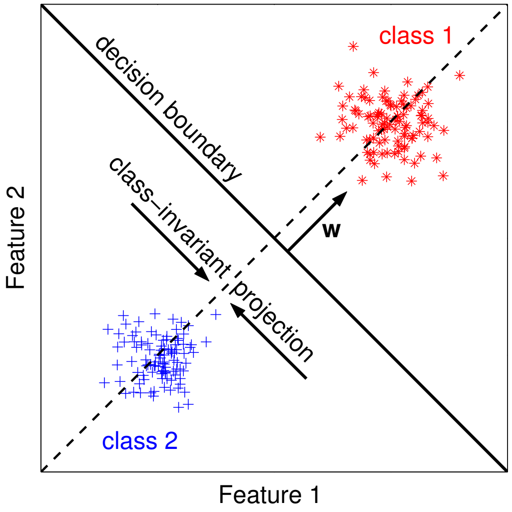
- Gray value image histograms.
- Photometric experiments.
- Normalized spectra.



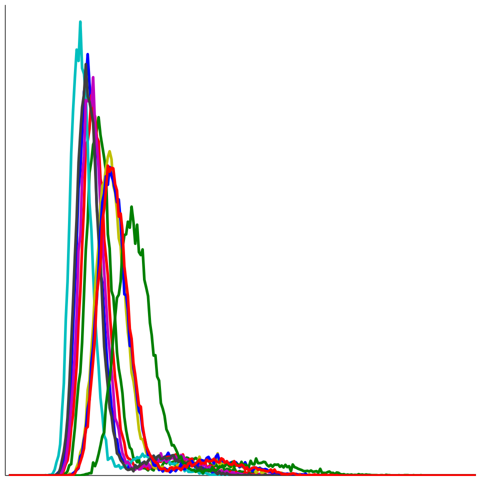
Tangent Kernels and Invariances

Often prior knowledge implies an invariance.

Connectivity: projection to the spectral bands or image regions does not change object membership.

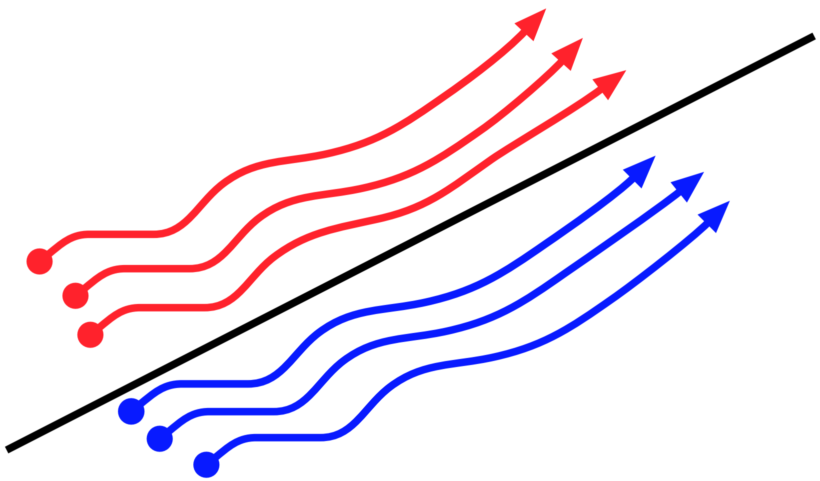


Calibration: scaling and shifting do not change object membership.



Tangent Kernels and Invariances

Transformation invariant classification.



$$\mathbf{x}(t) = \mathcal{L}_t \mathbf{x}$$

$$\mathbf{v}(t) = \frac{\partial \mathbf{x}(t)}{\partial t}$$

Decision boundary should be parallel to the speeds of objects

Tangent Kernels and Invariances

Definition.

Schölkopf, et al proposed tangent kernel approach: substitution of the similarity (inner product) by the robust one.

$$\mathbf{x}^{\mathrm{T}}\mathbf{y} \to \mathbf{x}^{\mathrm{T}}((1-\gamma)\mathbf{I} + \gamma\mathbf{C})^{-1}\mathbf{y}$$

$$\mathbf{C} = \sum_i \mathbf{v}_i \mathbf{v}_i^{\mathbf{T}}$$

$$\mathbf{v}_i = \left.\frac{\partial \mathcal{L}_t \mathbf{x}_i}{\partial t}\right|_{t=0}$$

Tangent Kernels and Invariances

Connectivity.

$$\mathcal{L}_t\mathbf{x} = \mathbf{x} - (1 - e^{-t})(\mathbf{x} - \mathbf{U}\mathbf{U}^{\mathrm{T}}\mathbf{x})$$

$$\mathbf{v} = -(\mathbf{x} - \mathbf{U}\mathbf{U}^{\mathrm{T}}\mathbf{x})$$

U is the band extraction matrix.

Tangent Kernels and Invariances

Invariance to the calibration.

If $x(z)$ is a normalized spectrum or an image channel histogram.

Translation

$$\mathcal{L}_t x(z) = x(z - t)$$

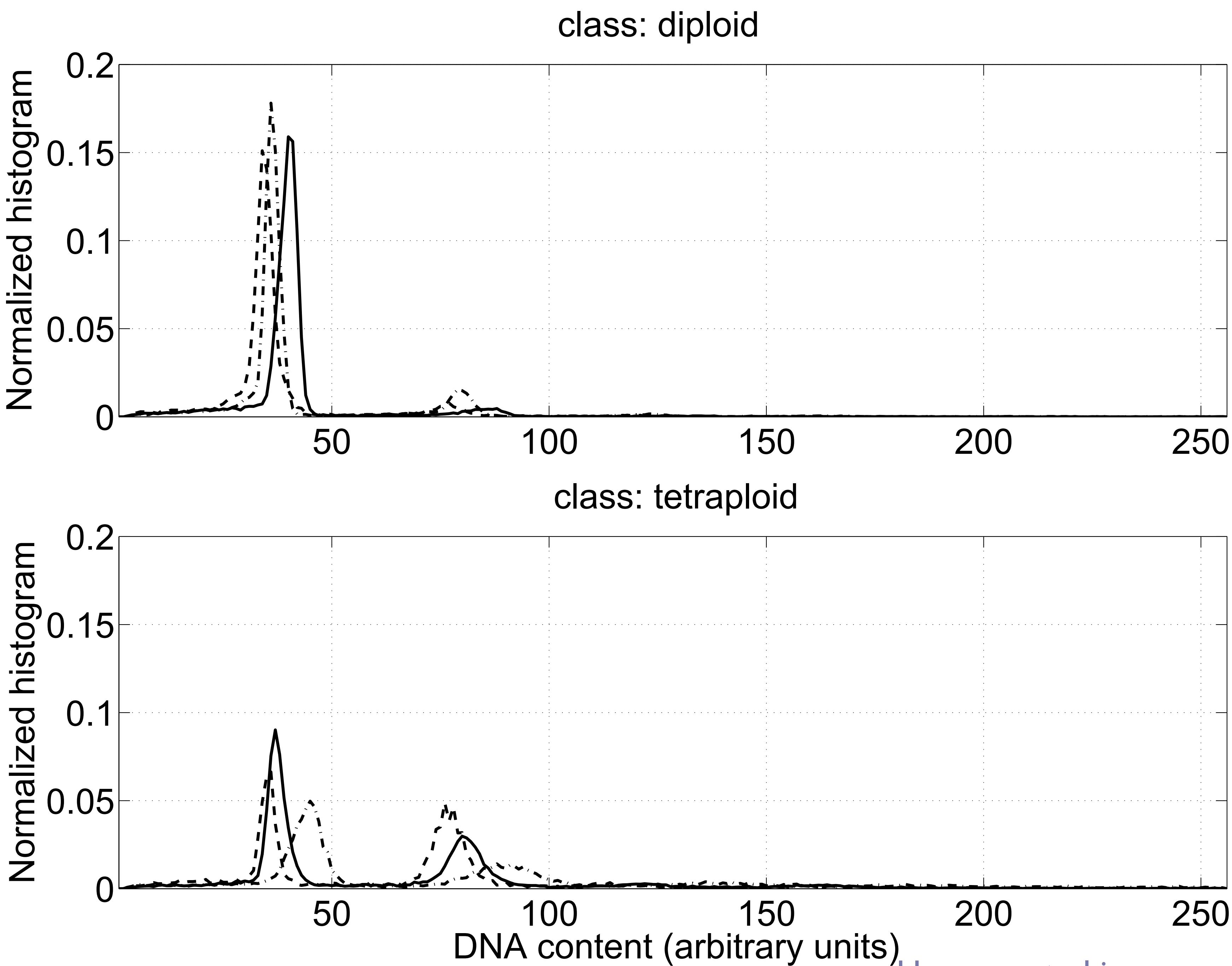
$$v(z) = -x'(z)$$

Scaling

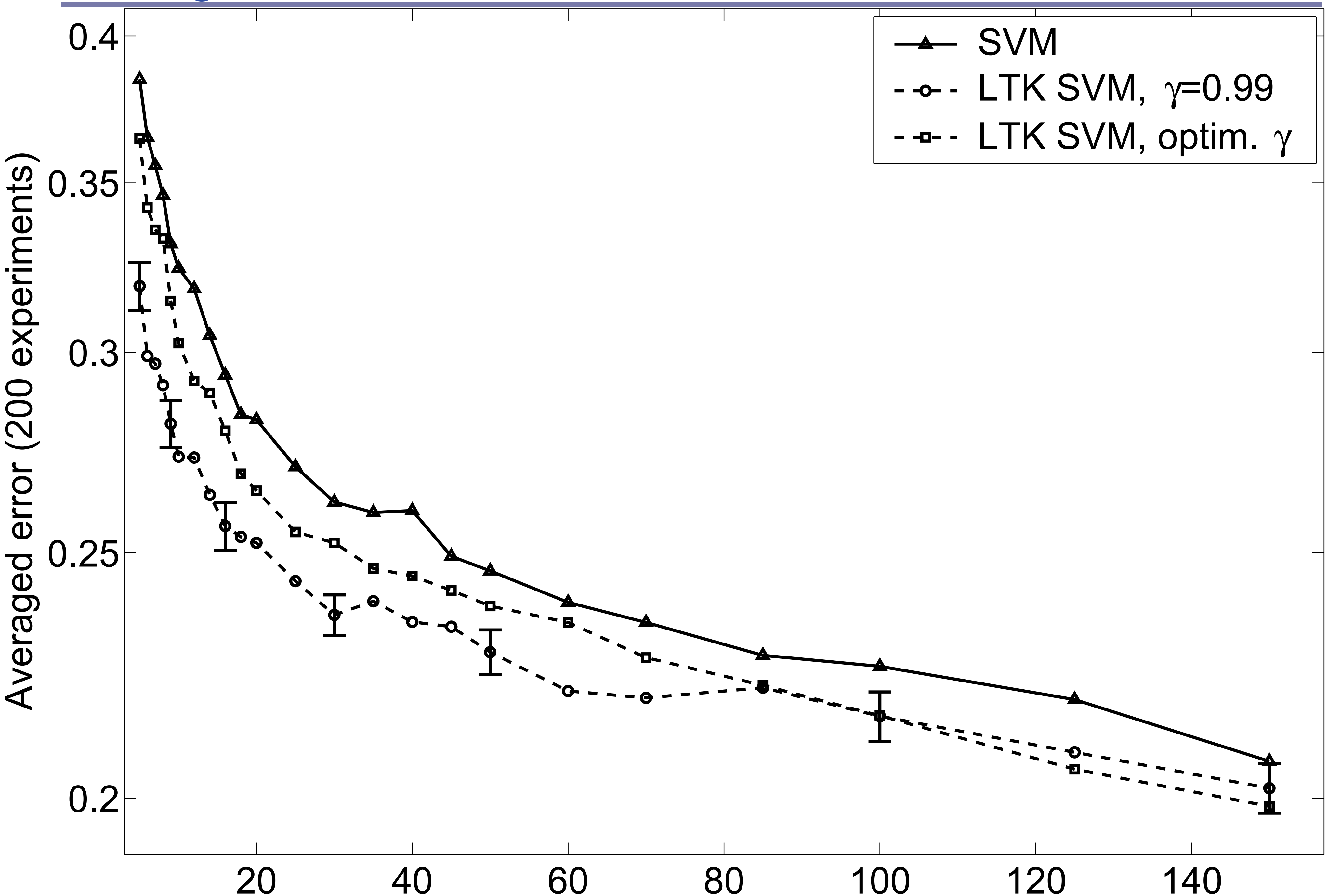
$$\mathcal{L}_t x(z) = e^{-t} x(e^{-t} z)$$

$$v(z) = -x(z) - z x'(z)$$

Tangent Kernels and Invariances



Tangent Kernels and Invariances



Hypertools toolbox

Matlab toolbox for analysis of hyperspectral images and spectral data

Based on *PRTools* (pattern recognition) and *DIPImage* (image proc.)

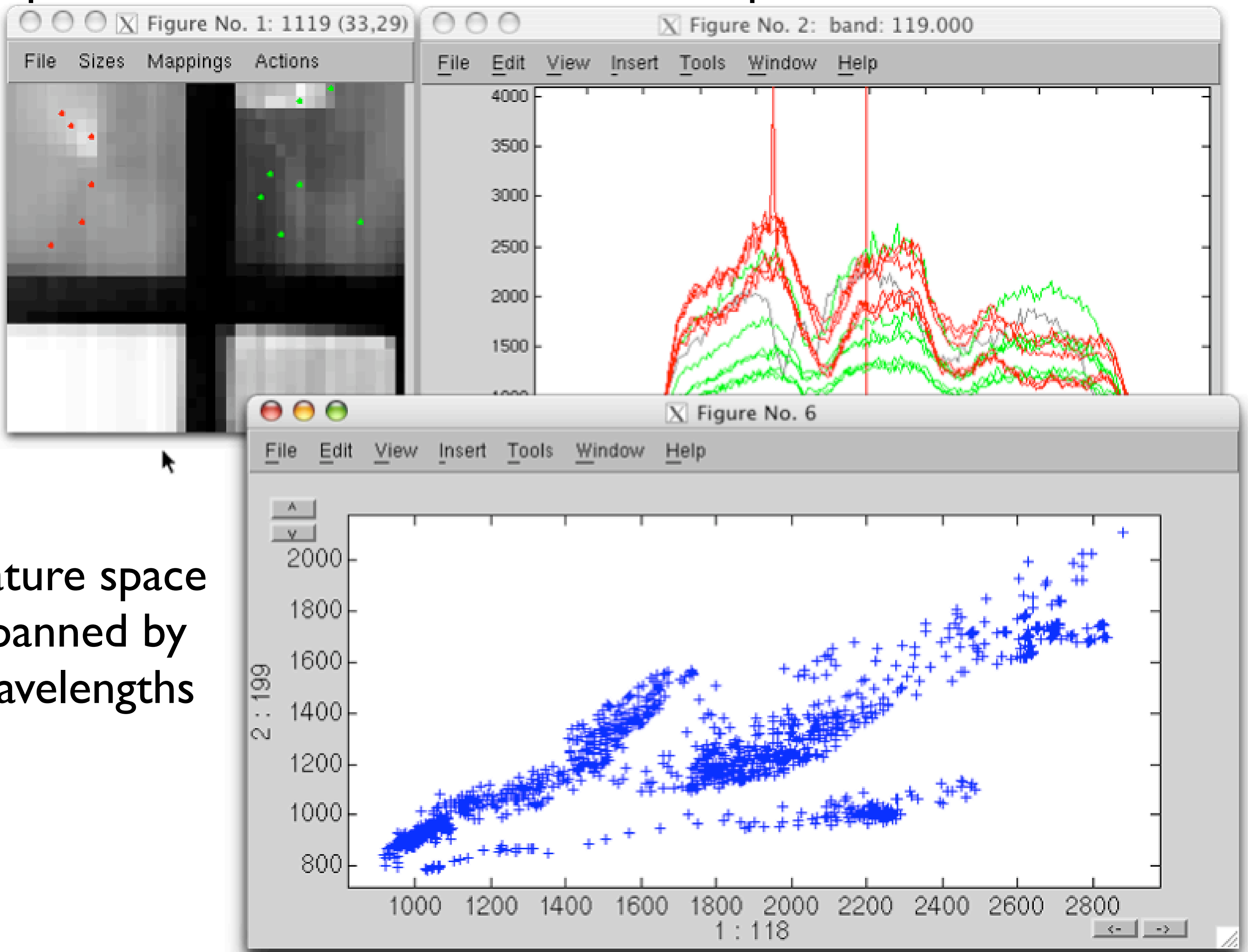
Free for academic use.

- baseline subtraction, smoothing, normalizations
- SIMCA, PCA
- GLDB extraction
- unmixing algorithms (VARIMAX,OPA,SIMPLISMA,ALS)
- spectra-specific dissimilarity measures
- visualization

Hypertools - data visualization

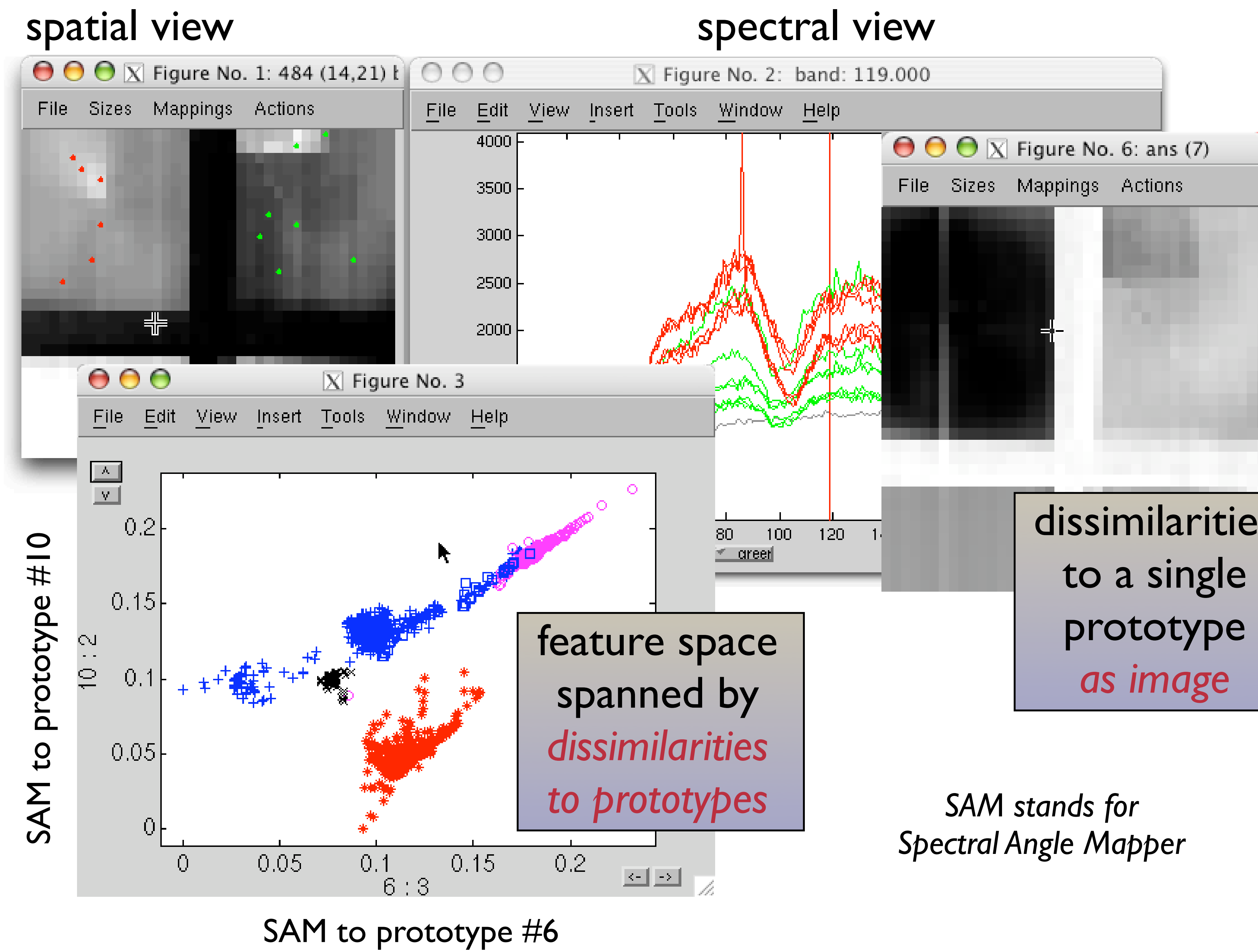
spatial view

spectral view



feature space spanned by wavelengths

Hypertools - dissimilarity representations



feature space spanned by *dissimilarities to prototypes*

dissimilarities to a single prototype *as image*

SAM stands for Spectral Angle Mapper