# Advanced Processing of Hyperspectral Images

State of Science of Environmental Applications of Imaging Spectroscopy
in honor of Dr. Alexander F.H. Goetz

A. Plaza[1], J. A. Benediktsson[2], J. Boardman[3], J. Brazile[4], L. Bruzzone[5], G. Camps-Valls[6], J. Chanussot[7], M. Fauvel[2,7], P. Gamba[8], A. Gualtieri[9,10], M. Marconcini[5], J. C. Tilton[9], G. Trianni[8]

[1]Univ. Extremadura (Spain), [2]University of Iceland, [3]Analytical Imaging and Geophysics LLC (USA), [4]Univ. Zürich (Switzerland), [5]Univ. Trento (Italy), [6]Univ. Valencia (Spain), [7]INPG Grenoble (France), [8]Univ. Pavia (Italy), [9]NASA/Goddard Space Flight Center (USA), [10]Global Science & Technology (USA)

# Talk Outline

# Levels of Spectral Information in Remote Sensing

**Ultraspectral**

**(1000's of bands)**

**Hyperspectral**

**(100's of bands)**

**Multispectral**

**(10's of bands)**

**Panchromatic**
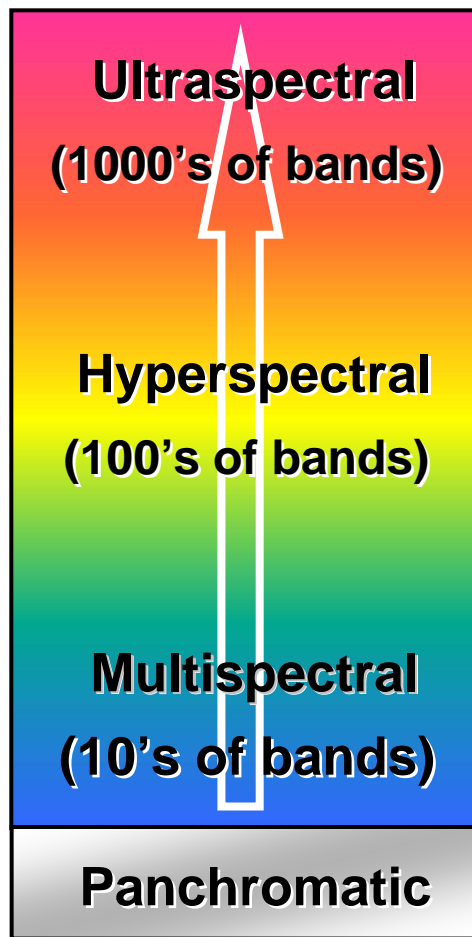
**Spectral mixture analysis:** Determines the abundance of materials (e.g. precision agriculture).

**Characterization:** Determines variability of identified material (e.g. wet/dry sand, soil particle size effects).

**Identification:** Determines the unique identity of the foregoing generic categories (e.g. land-cover or mineral mapping).

**Discrimination:** Determines generic categories of the foregoing classes.

**Classification:** Separates materials into spectrally similar groups (e.g., urban data classification).

**Detection:** Determines the presence of materials, objects, activities, or events.

# Challenges in hyperspectral image processing

- The special characteristics of hyperspectral data pose several *processing problems*:

  1. The high-dimensional nature of hyperspectral data introduces important limitations in supervised classifiers, such as the **limited availability of training samples** or the inherently complex structure of the data.

  2. There is a need to **integrate the spatial and spectral information** to take advantage of the complementarities that both sources of information can provide, in particular, for unsupervised data processing.

  3. There is a need to develop **parallel algorithm implementations**, able to speed up algorithm performance and to satisfy the extremely high computational requirements of time-critical remote sensing applications.

- In this work, we have taken a necessary first step towards the understanding and assimilation of the above aspects in the design of *last-generation hyperspectral image processing algorithms*.

# Support Vector Machines (SVMs)

- High-dimensional spaces are mostly *empty*, making density estimation difficult.

- SVMs consider *geometrical* rather than *statistical* properties of the classes.

  ✓ *Kernel-trick* allows one to work in the *mapped* kernel space.

  ✓ No need to know the *mapping function*: $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle = k(\mathbf{x}_i, \mathbf{x}_j)$
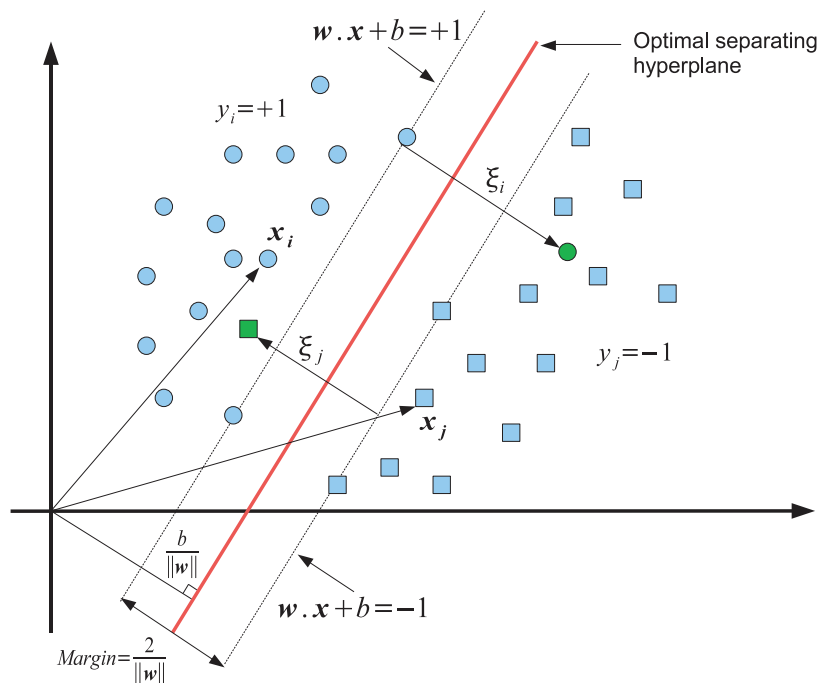
*Polynomial kernel:*

$$k_{poly}(\mathbf{x}, \mathbf{z}) = (\langle \mathbf{x}, \mathbf{z} \rangle + \theta)^d$$

*SAM kernel:*

$$k_{SAM}(\mathbf{x}, \mathbf{z}) = \exp\left(-\gamma \alpha(\mathbf{x}, \mathbf{z})^2\right)$$

$$\alpha(\mathbf{x}, \mathbf{z}) = \arccos\left(\frac{\langle \mathbf{x}, \mathbf{z} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{z}\|}\right)$$

*Gaussian RBF kernel:*

$$k_{gauss}(\mathbf{x}, \mathbf{z}) = \exp\left(-\gamma \|\mathbf{x} - \mathbf{z}\|^2\right)$$

# Composite Kernels for Image Classification

- *Some Properties of Mercer's kernels:*

  ✓ Sum of valid kernels *is* a valid kernel.

  ✓ Scaling a valid kernel by a positive factor *is* a valid kernel

  ✓ Tensor product of valid kernels *is* a valid kernel.

- From *Functional Analysis,* the direct sum of Hilbert spaces allows elegant integration of different information sources in any *kernel* machine.

- To decompose the *spectral* and *contextual/spatial* information and *merge it in the composite kernel.*

- New family of composite kernels developed, accounting for *spatial, spectral* and cross-information simultaneously.
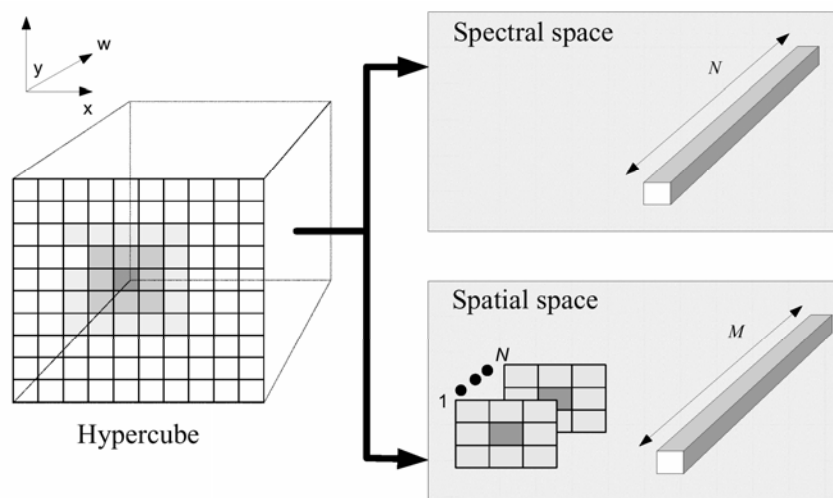
# Transductive Support Vector Machines (TSVMs)

- Based on the joint use of *labeled* and *unlabeled* patterns in the framework of a *transductive iterative learning process.*

**Standard Inductive Approch:** The learning of the classifier is carried out only on the labeled samples of the *training set*
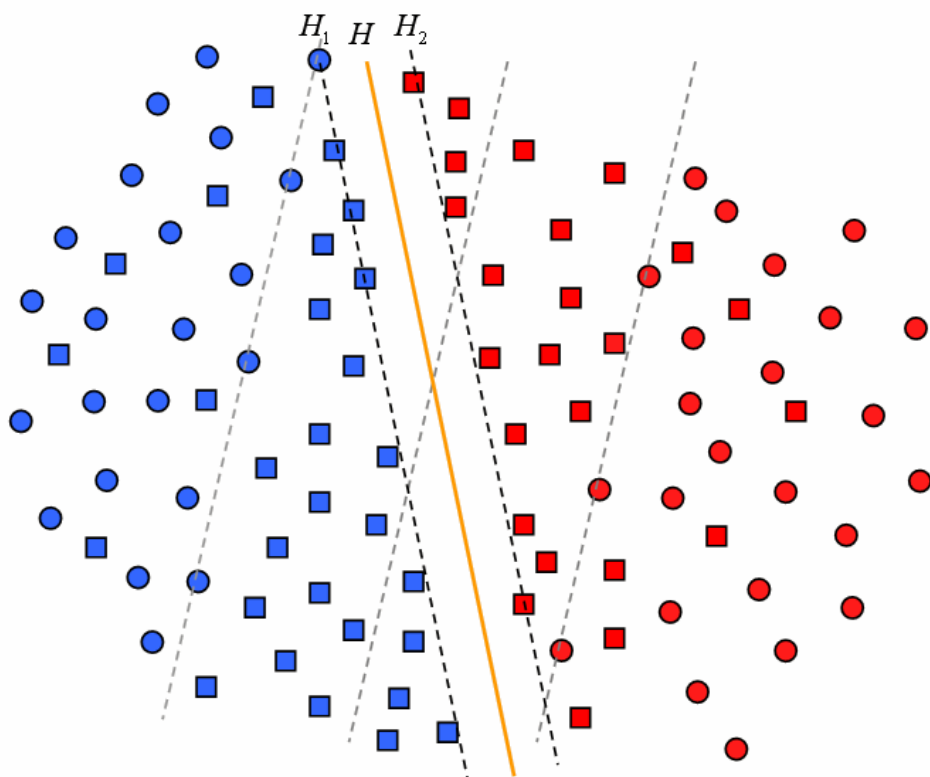
$$\begin{cases} \min_{w,b,\xi} \left\{ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \right\} \\ y_i \cdot (w \cdot x_i + b) + \xi_i - 1 \geq 0, \quad \forall i = 1,\dots,N \\ \xi_i \geq 0, \quad \forall i = 1,\dots,N \end{cases}$$



**TRansductive Approch:** The learning of the classifier is carried out using both *labeled* and *unlabeled* patterns

$$\begin{cases} \min_{w^{(i)},b^{(i)},\xi^{(i)},\xi^{*(i)}} \left\{ \frac{1}{2}\|w^{(i)}\|^2 + C\sum_{i=1}^{n}\xi_i^{(i)} + \sum_{j=1}^{\eta^{(i-1)}}C_j^*\xi_i^{*(i)} \right\} \\ y_l^{(i)} \cdot \left(w^{(i)} \cdot x_l + b^{(i)}\right) \geq 1 - \xi_l^{(i)}, \quad \forall l = 1,\dots,\eta, \quad x_l \in X^{(i)} \\ y_j^{*(i)} \cdot \left(w^{(i)} \cdot x_j^* + b^{(i)}\right) \geq 1 - \xi_l^{*(i)}, \quad \forall j = 1,\dots,\eta^{(i-1)}, \quad x_j^* \in J^{(i} \\ \xi_l^{(i)}, \xi_l^{*(i)} \geq 0 \end{cases}$$

# Transductive Support Vector Machines (TSVMs)

- Important issues:

  ✓ **Selection of transductive samples**. Goals:

  i) To select samples with an *expected accurate labeling;*

  ii) To choose *informative samples*.

  Iterative selection procedure designed to choose a balanced number of *labeled* and *transductive* samples (unlabeled patterns closest to margin bounds).

  ✓ **Threshold criterion**. To select transductive samples in a small solution space, a threshold criterion is used to consider the *density* of the selected area. A pairwise labeling strategy is used to alleviate the problem of unbalanced classes.

  ✓ **Regularization parameter**. Used to control the number of misclassified samples that belong to the original training set and the unlabeled set. The larger the regularization parameter, the higher the influence of the associated samples on the selection of the discriminant hyperplane.

  ✓ **Multi-class extension**. As in standard SVMs, the transductive process is based on a structured architecture made up of binary classifiers. It must be possible to give a classification label to all unlabeled samples (*one versus the rest*).
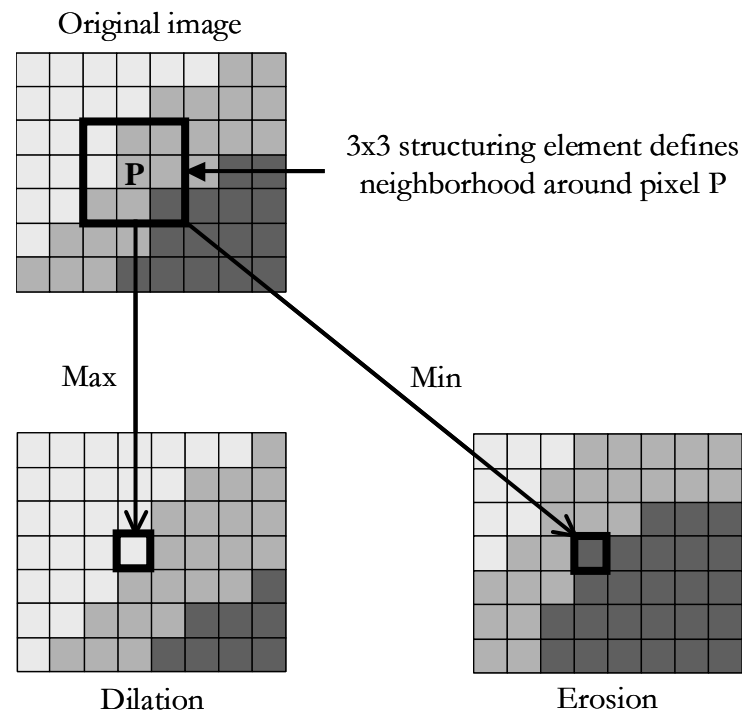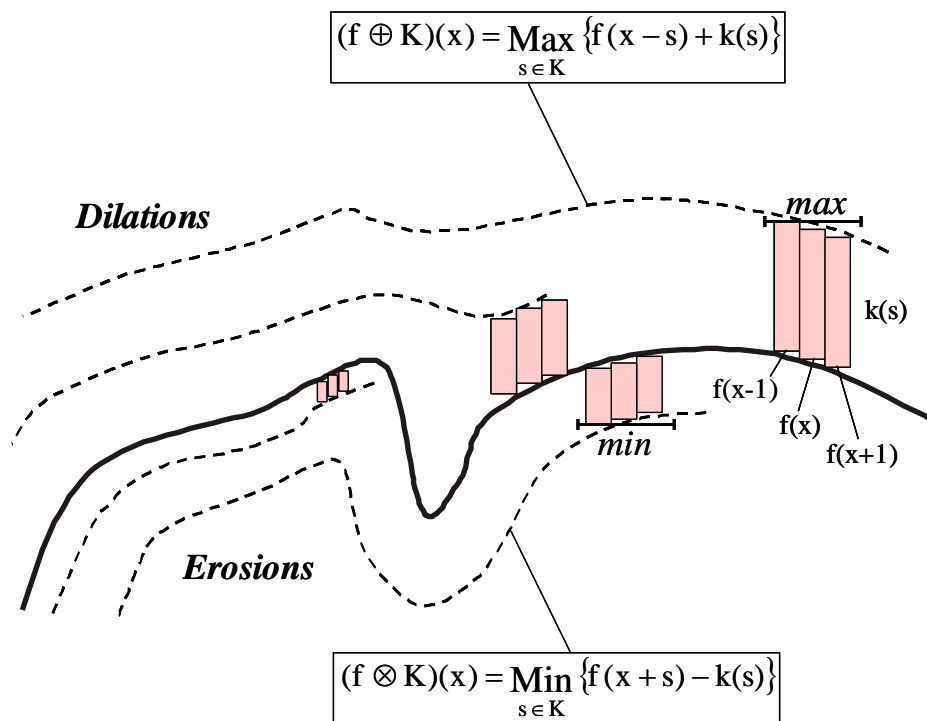
# Why integrated spatial/spectral approaches?

- Much effort has been given to processing hyperspectral image data in *spectral* terms.

- Data analysis is carried out without incorporating information about *spatial* context.
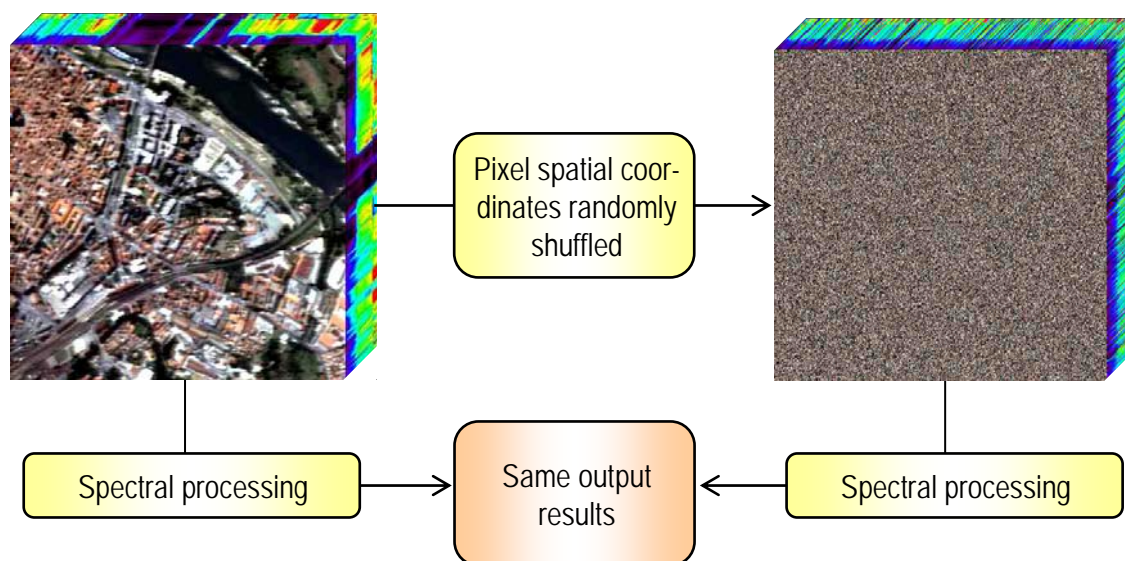


- There is a need to incorporate the *image representation* of the data in the analysis.

- Most available approaches consider spatial and spectral information *separately*.

- Several approaches considered in this work to achieve the desired integration.

# Mathematical Morphology

- *Grayscale* morphology relies on a *partial* ordering relation between image pixels.

$$(f \oplus K)(x) = \underset{s \in K}{\mathrm{Max}} \{f(x-s) + k(s)\}$$

**Dilations**

*max*

k(s)

f(x-1)

f(x)

*min*

f(x+1)

**Erosions**

$$(f \otimes K)(x) = \underset{s \in K}{\mathrm{Min}} \{f(x+s) - k(s)\}$$

Original image

P

3x3 structuring element defines neighborhood around pixel P
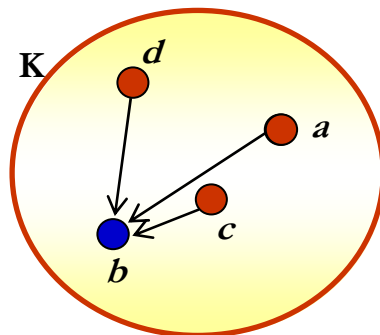
Max

Min

Dilation

Erosion

- Morphological operations for *hyperspectral imagery* require *ordering* of image *pixels*.

- Two strategies explored in this work: vector-based ordering and PCA-based ordering.
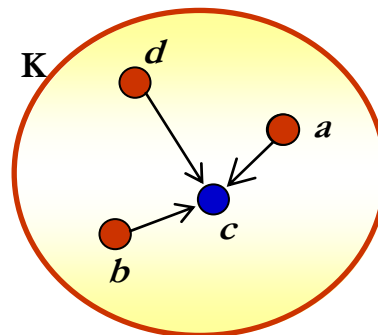
# Vector-based ordering

- Based on a *spectral* distance function (SAD, SID) and a *cumulative* distance measure.
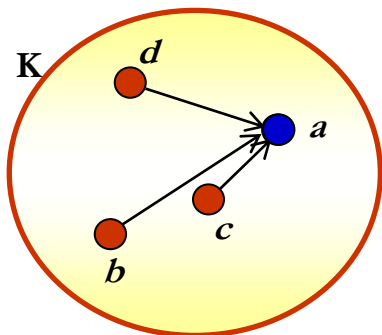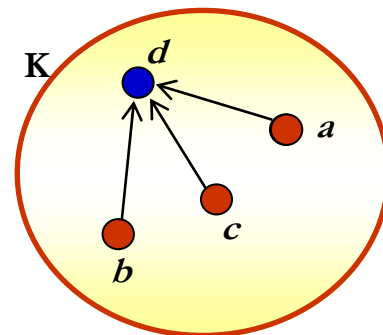


$D_a = $SAD$(b,a)+$
SAD$(c,a)+$
SAD$(d,a)$

$D_b = $SAD$(a,b)+$
SAD$(c,b)+$
SAD$(d,b)$

$D_c = $SAD$(a,c)+$
SAD$(b,c)+$
SAD$(d,c)$

$D_d = $SAD$(a,d)+$
SAD$(b,d)+$
SAD$(c,d)$

- The *greatest* element is the *most spectrally distinct (pure)* in the structuring element.

- The *least* element is the *most spectrally similar (mixed)* in the structuring element.

- *Extended dilation* has the effect of expanding pure spectral areas in the image.

- *Extended erosion* reduces pure spectral areas and expands mixed areas.

- Particularly suited for spatial/spectral endmember extraction.

# Automated Morphological Endmember Extraction.-

- Integration of *spectral* and *spatial* information (computation intensive)

- Selection of the most *spectrally pure* and the *most spectrally* mixed signatures.
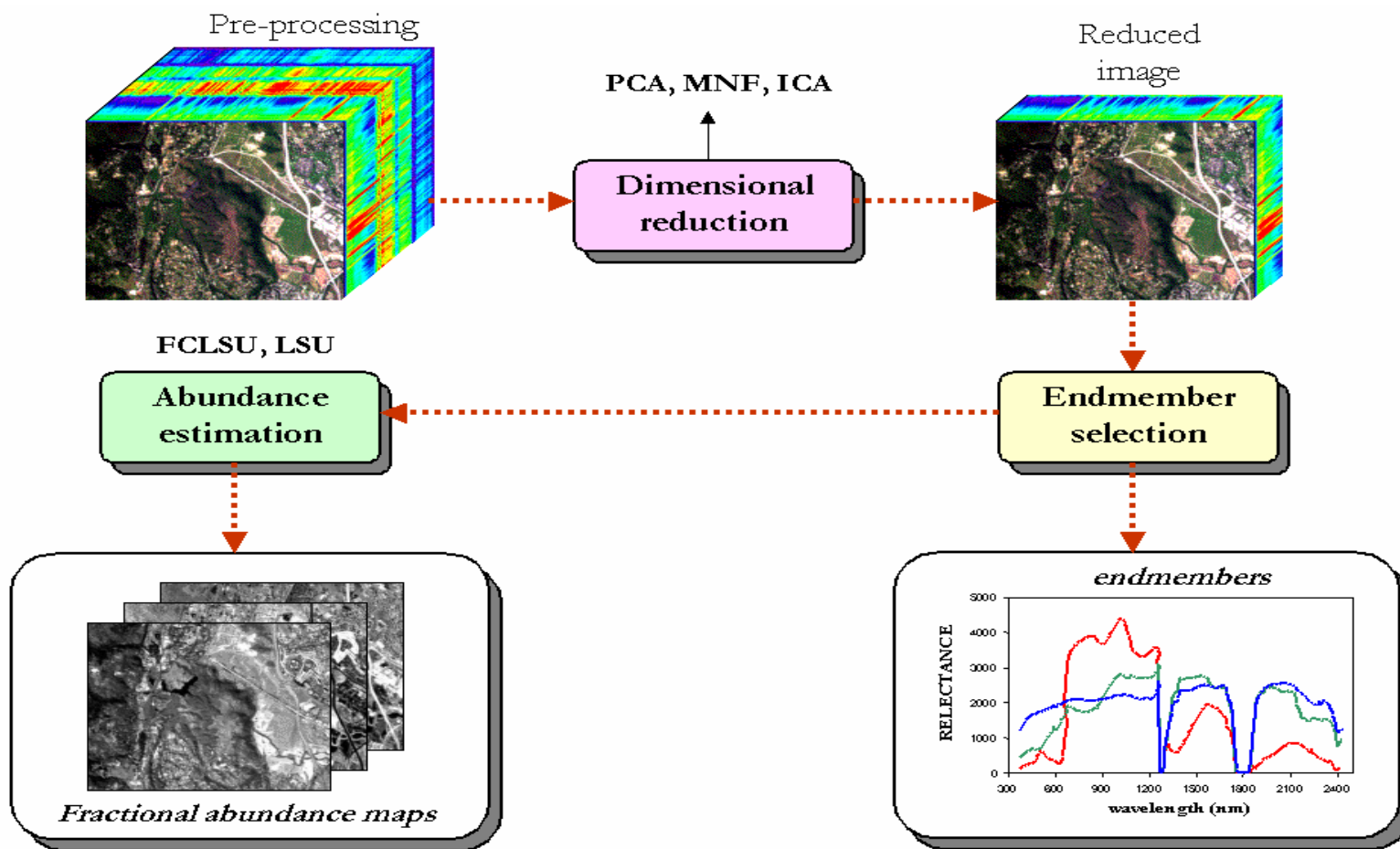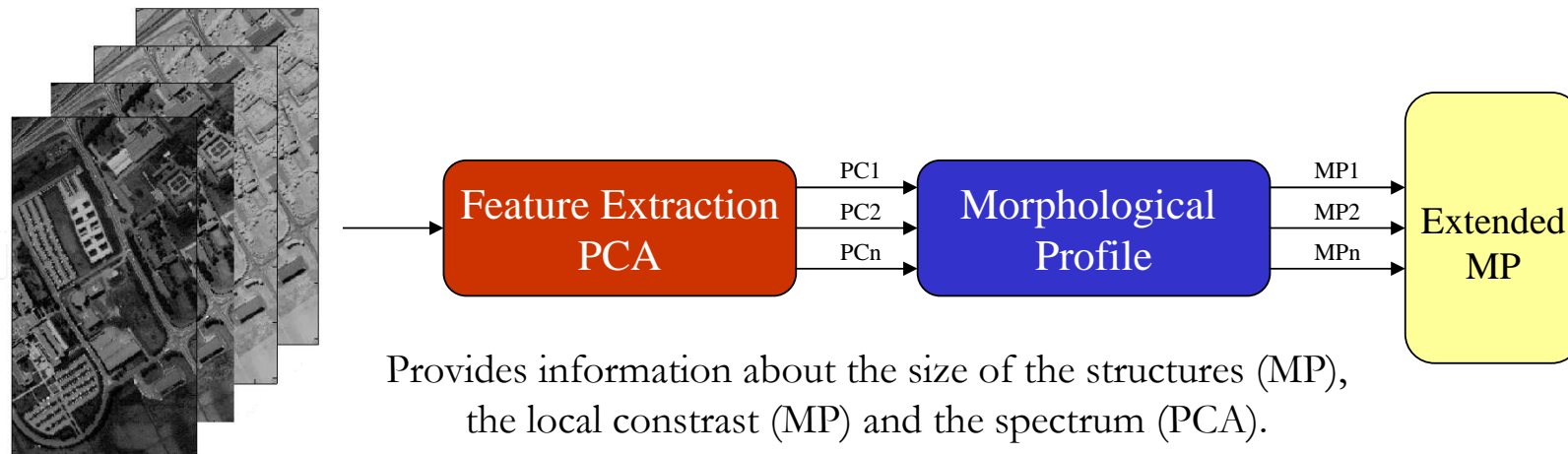
# Classification Using Morphological Profiles.-

Uses opening and closing operations to create a *feature vector* for classification:

$$OP_i(x,y) = \gamma_R^{(i)}(x,y), \ \forall i \in [0,k] \qquad CP_i(x,y) = \phi_R^{(i)}(x,y), \ \forall i \in [0,k]$$

$$MP(x,y) = \{CP_k(x,y), \ldots, f(x,y), \ldots, OP_k(x,y)\}$$

# Extended Morphological Profile.-



Provides information about the size of the structures (MP), the local constrast (MP) and the spectrum (PCA).

$$MP_{ext}(x,y) = \{MP_{PC_1}(x,y), \ldots, MP_{PC_k}(x,y)\}$$

# Spatial/spectral classification using MRFs

- Minimization of a cost function:

$$U(\mathbf{g}(x,y), C(x,y)) = \alpha U_{spectral}(\mathbf{g}(x,y), C(x,y)) + U_{spatial}(\mathbf{g}(x,y), C(x,y))$$

where
$$U_{spatial}(\mathbf{g}(x,y), C(x,y)) = \sum_{(i,j) \in G(x,y)} \beta I(C(x,y), C(i,j))$$

and
$$U_{spectral}(\mathbf{g}(x,y), C(x,y)) = \frac{m}{2} \ln|2\pi \Sigma_k| + \frac{1}{2}(\mathbf{g}(x,y) - \mu_k)^T \Sigma_k^{-1}(\mathbf{g}(x,y) - \mu_k)$$

**SPECTRAL DOMAIN**　　　　　　　　　　**SPATIAL DOMAIN**

| NEURO FUZZY APPROACH | → | MRF ANALYSIS |

- Integration of the pattern recognition capability of a neuro-fuzzy classifier and the spatial/spectral nature of the probabilistic ML-based MRF framework

# Hierarchical Segmentation (HSEG)

- HSEG produces a set of image segmentations (**segmentation hierarchy**):
  - ✓ *Coarser* segmentations produced from merges of regions from *finer* segmentations.
  - ✓ Region boundaries maintained at the full image spatial resolution.

- HSEG is a hybridization of Hierarchical Step-Wise Optimization (region growing) with **spectral clustering**, controlled by *spclust_wght* (an input parameter).

- A recursive approximation of HSEG, called RHSEG, is much more computationally efficient (especially for *spclust_wght* > 0.0):
  - RHSEG **recursively** subdivides the image data and then recombines the results such that the number of regions handled at any point in the program is restrained.
  - The recombination step of RHSEG requires special blending code to avoid **processing window artifacts**.

# Why High-Performance Computing is Crucial?
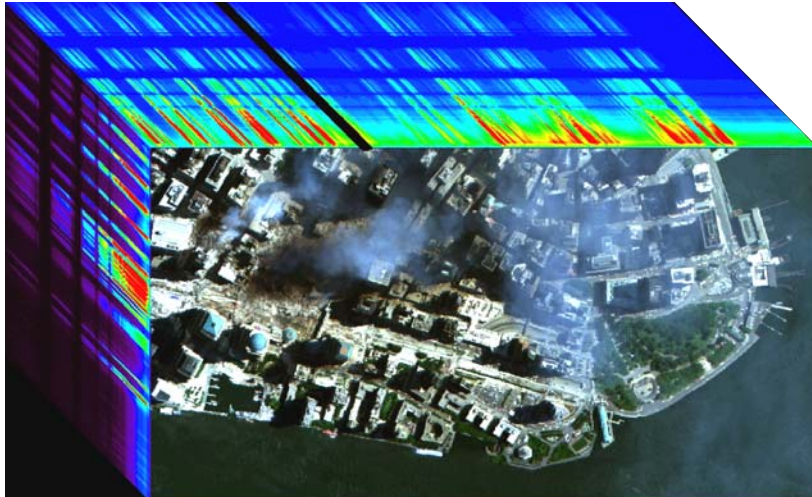


Fire Temperatures
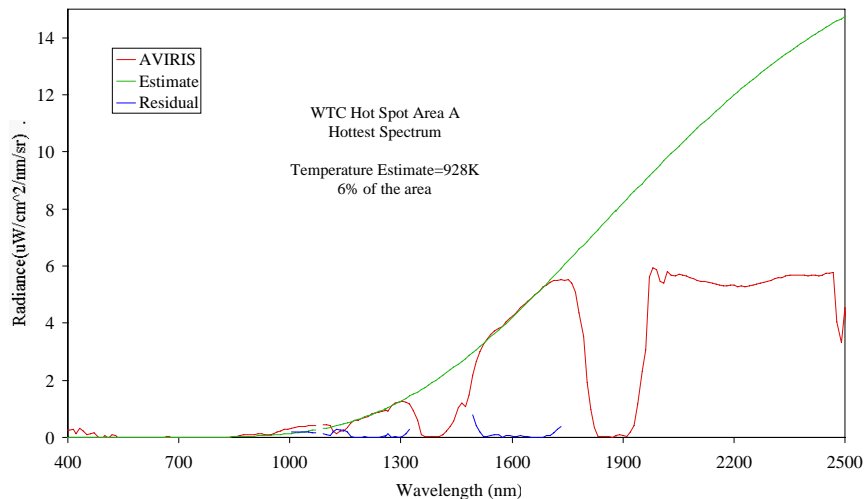


**Biomass Burning**: Sub-pixel temperatures and extent, smoke, combustion products…

**Environmental Hazards**:  Contaminants (direct and indirect), geological substrate…

**Coastal and Inland Waters**: Chemical and biological standoff detection, oil spill monitoring and tracking...

**Ecology**: Chlorophyll, leaf water, lignin, cellulose, pigments, structure, nonphotosynthetic constituents…

**Commercial Applications**: Mineral exploration, agriculture and forest status…

**Military Applications**: Detection of land mines, tracking of targets, decoys...
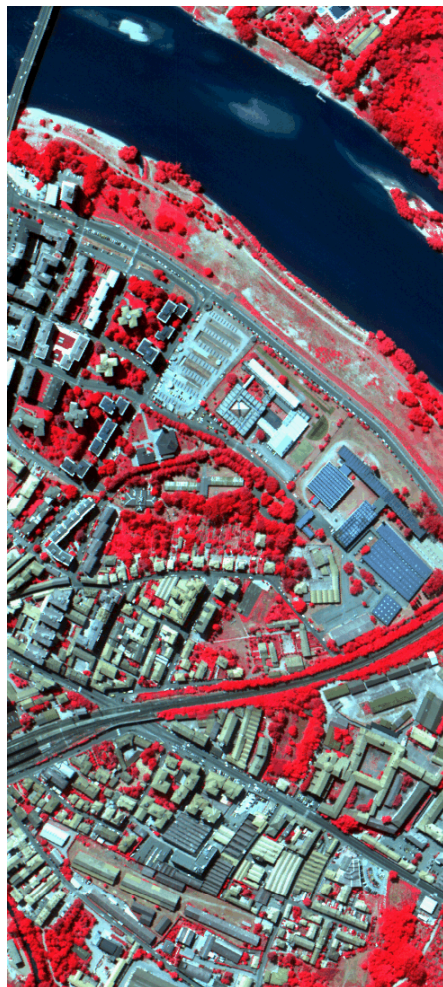
**Others**: Human infrastructure, medical...

# Parallel Framework for Morphological Methods
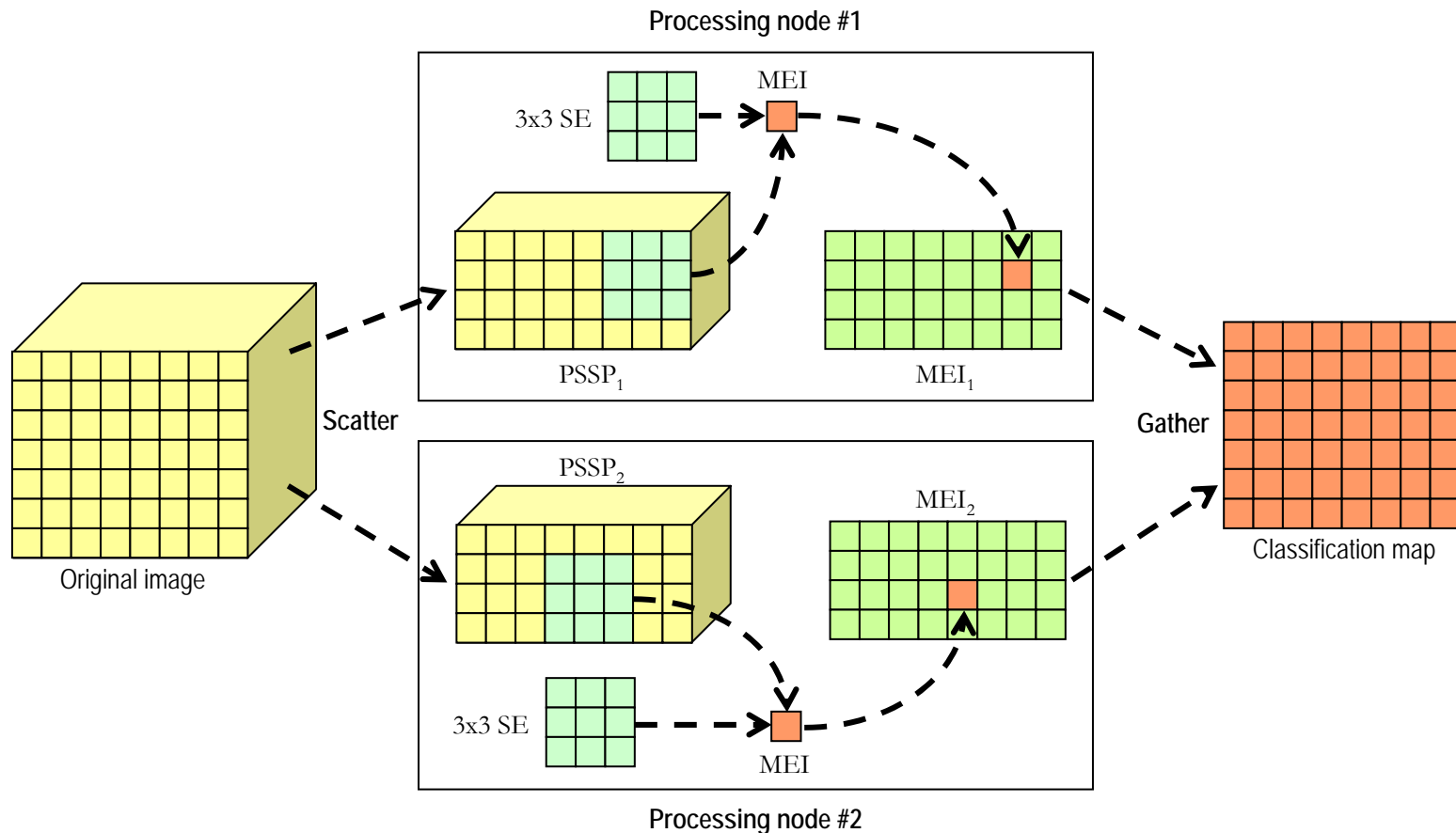


- The *master* processor is in charge of distributing the work among the *workers*.

- Each partition is processed *independently*, and the master gathers the final result.
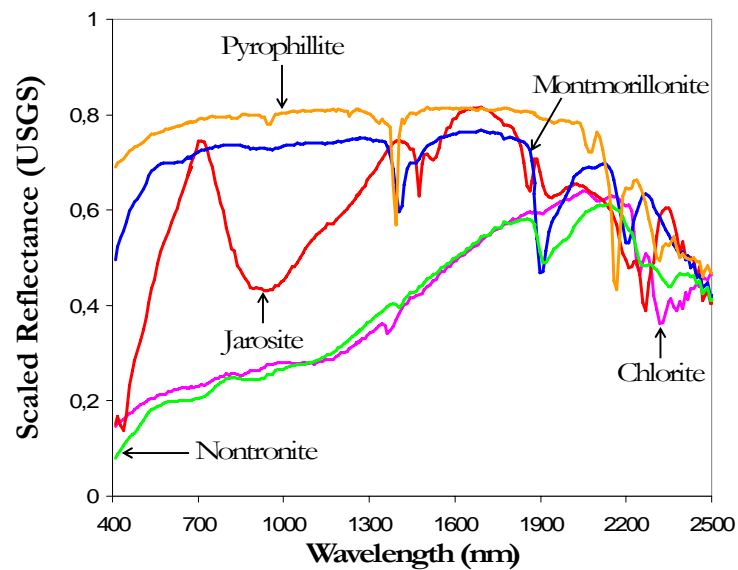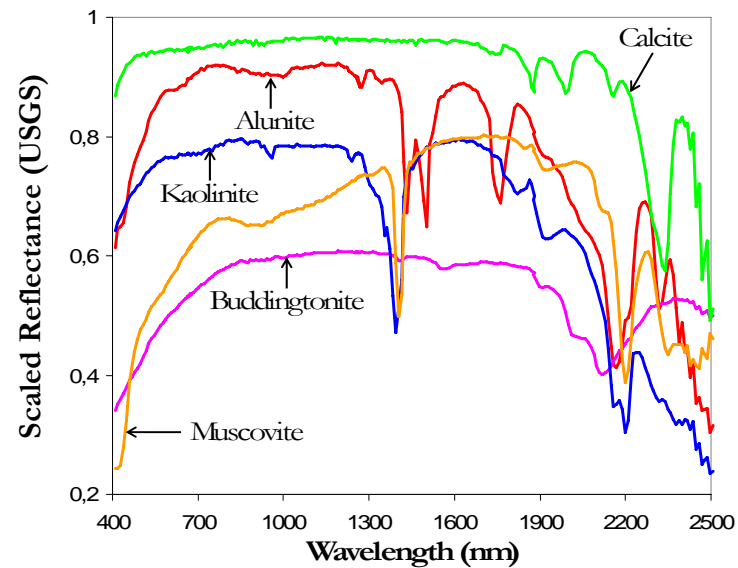
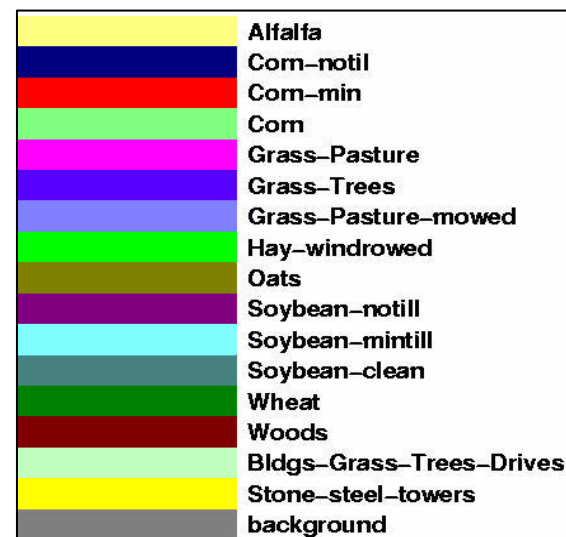# ROSIS Urban Hyperspectral Data Over Pavia, Italy
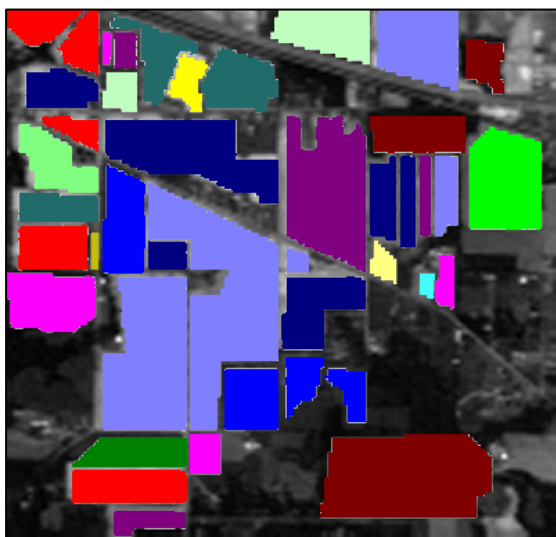


Subset #1

Subset #2

Subset #3

# AVIRIS Data Over Cuprite, Nevada

# Standard SVMs vs Transductive SVMs.-

- Tested using AVIRIS data over Indian Pines, made up of 16 ground-truth classes (7 classes were discarded due to insufficient training samples).

- Gaussian RBF classifier and *one versus the rest* architecture with 9 binary classifiers.



Alfalfa
Corn–notil
Corn–min
Corn
Grass–Pasture
Grass–Trees
Grass–Pasture–mowed
Hay–windrowed
Oats
Soybean–notill
Soybean–mintill
Soybean–clean
Wheat
Woods
Bldgs–Grass–Trees–Drives
Stone–steel–towers
background

| Percentage of Training Samples | Number of Samples | Overall Accuracy | | Kappa | |
|---|---|---|---|---|---|
| | | SVMs | Proposed TSVMs | SVMs | Proposed TSVMs |
| 5% | 237 | 73.41 | 76.20 | 0.68 | 0.71 |
| 10% | 475 | 76.46 | 80.21 | 0.73 | 0.77 |
| 25% | 1189 | 82.17 | 84.83 | 0.79 | 0.82 |

# Composite Kernel-Based Image Classification.-

- Excellent results on a wide range of scenarios (tested on all 16 Indian Pines classes)

| | Overall accuracy | Kappa |
|---|---|---|
| **Spectral classifiers†** | | |
| Euclidean (Tadjudin and Landgrebe, 1998) | 48.23 | — |
| bLOOC+DAFE+ECHO (Tadjudin and Landgrebe, 1998) | 82.91 | — |
| $k_\omega$ (Gualtieri, 1999) | 87.30 | — |
| $k_\omega$ (developed in this paper) | 88.55 | 0.87 |
| **Spatial/spectral classifiers** | | |
| *Mean* | | |
| Spatial | 84.55 | 0.82 |
| Stacked | 94.21 | 0.93 |
| Summation | 92.61 | 0.91 |
| Weighted | **95.97** | **0.94** |
| Cross-terms | 94.80 | 0.94 |
| Summation + Stacked | 95.20 | 0.94 |
| Cross-terms + Stacked | 95.10 | 0.94 |
| *Mean and standard deviation ‡* | | |
| Spatial | 88.00 | 0.86 |
| Stacked | 94.21 | 0.93 |
| Summation | 95.45 | 0.95 |
| Weighted | **96.53** | **0.96** |
| Summation + Stacked | 96.20 | 0.95 |

*TrueMap*

*Spectral SVM*

*Weighted Kernel*

- More advanced contextual/textural extractions can be easily integrated.

- Extensions to *multi-temporal* and *semi-supervised* versions yield good results.

## Subset #1 ROSIS Urban Data



FUZZY ARTMAP

DAFE/MRF

|  | DAFE/MRF | Neuro-fuzzy |
|---|---|---|
| Overall accuracy | 97.27 | 97.29 |
| Water | 99.04 | 99.71 |
| Trees | 91.27 | 93.19 |
| Grass | 97.09 | 94.80 |
| Parking lot | 76.83 | 71.99 |
| Bare soil | 93.33 | 93.36 |
| Asphalt | 99.65 | 81.87 |
| Bitumen | 88.45 | 96.42 |
| Tiles | 98.33 | 99.98 |
| Shadow | 99.86 | 99.93 |

- Similar classification accuracies

- Different performances in border areas

- The DAFE/MRF frawework achieves better geometrical characterization of buildings and roads.

- The neuro-fuzzy procedure performs better in homogeneuos areas.

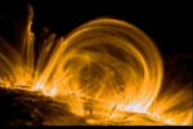# Extended Morphological Profiles vs Spectral Info.-

- PCA was applied to the original spectral information, and the first *three* principal components were retained (99% of the *cumulative variance*).

- Morphological profiles were constructed *for each component*, based on 10 opening/closings by reconstruction and a disk-shaped structuring element.

- Classification results using 102 *spectral bands* vs. 63 *morphological features*:

|  | Original spectral information | Extended morphological profile |
|---|---|---|
| Overall accuracy | 80.99 | 85.22 |
| Average accuracy | 88.28 | 90.76 |
| Kappa | 76.16 | 80.86 |
| Asphalt | 83.71 | 95.36 |
| Meadow | 70.25 | 80.33 |
| Gravel | 70.32 | 87.61 |
| Tree | 97.81 | 98.37 |
| Metal Sheet | 99.41 | 99.48 |
| Bare Soil | 92.25 | 63.72 |
| Bitumen | 81.58 | 98.87 |
| Brick | 92.59 | 95.41 |
| Shadow | 96.62 | 97.68 |

# Thunderhead (NASA)

## http://thunderhead.gsfc.nasa.gov
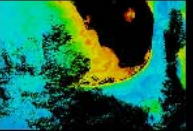
| Aggregate Specification | |
|---|---|
| Number of nodes | 256 |
| Total processors | 512 |
| Total memory (Gb) | 256 |
| Total disk (GB) | 20480 |
| Interconnect 1 | Myrinet 2000 |
| Interconnect 2 | Gigabit Ethernet |
| Total peak performance (Gflops) | 2457.6 |
| **Node Specification** | |
| Motherboard | Tyan Thunder 2720 |
| Number of processors | Dual Intel 4 Xeon 2.4Ghz |
| Memory (Gb) | 1 |
| Local disk (Gb) | 80 |
| Interconnect 1 | Myrinet 2000 |
| Interconnect 2 | Gigabit 10/100/1000 |
| Peak performance (Gflops) | 9.6 |

NASA/Goddard Space Flight Center
High Performance Computing

HYDRA Cluster:
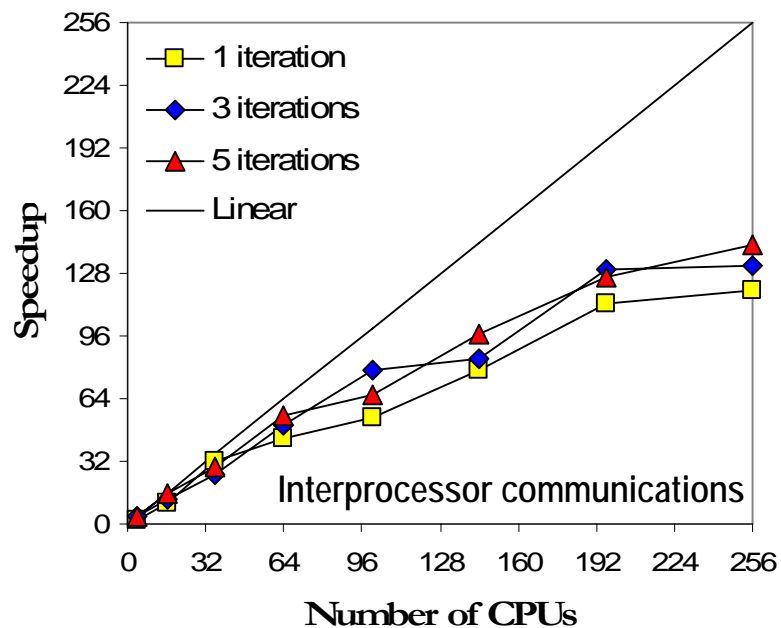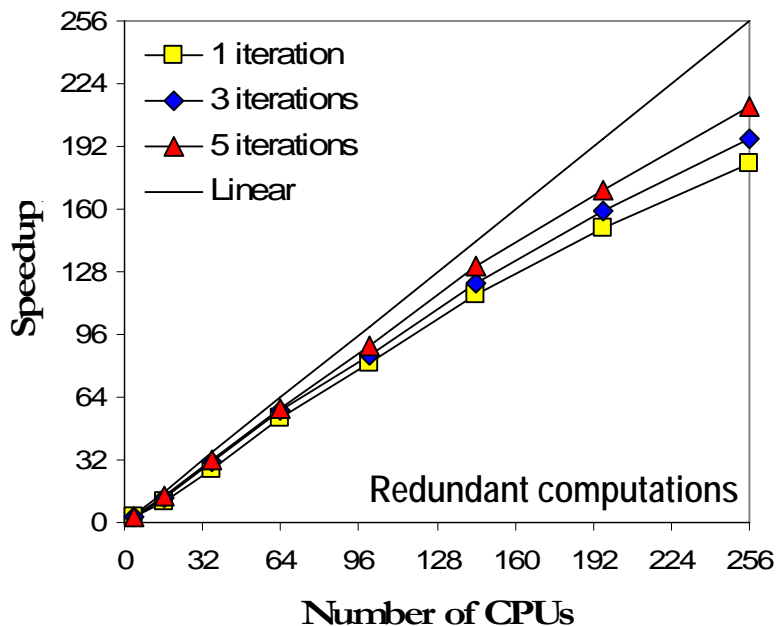Thunderhead, Medusa, Orka, Pivot, BBBLUE

NASA Summer School for High Performance Computing:
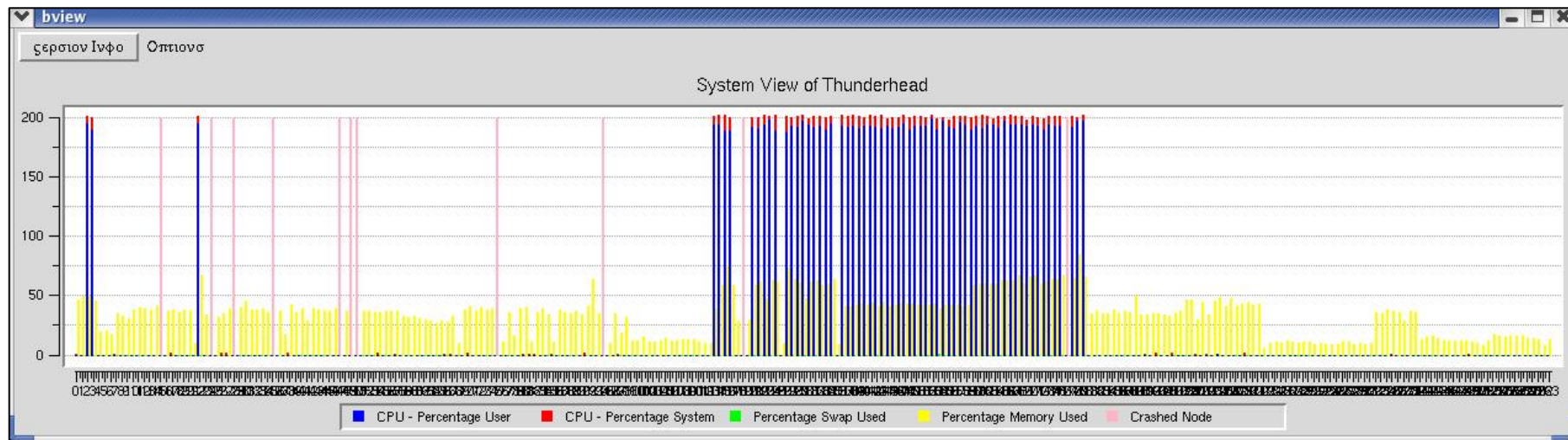
# Performance of Morphological Endmember Extraction.-

- Algorithms were implemented in C++ using calls to Message Passing Interface (MPI).

- Using redundant computations versus communications was crucial:.



**Processing times (seconds) for different numbers of processors on Thunderhead (AVIRIS Cuprite)**

|  | 1 | 4 | 16 | 36 | 64 | 100 | 144 | 196 | 256 |
|---|---|---|---|---|---|---|---|---|---|
| Redundant | 9452+13 | 4075+10 | 917+12 | 381+11 | 205+15 | 128+16 | 89+14 | 65+11 | 50+10 |
| Interprocesssor | 9356+143 | 4062+151 | 889+160 | 371+194 | 205+225 | 124+243 | 83+261 | 62+268 | 49+292 |

A. Plaza et al., "A quantitative and comparative analysis of endmember extraction algorithms," *IEEE-TGARS* (42), 3, 650-663, 2004

# Performance of Parallel Support Vector Machines

MPI Code to Implement **boss_workers**:
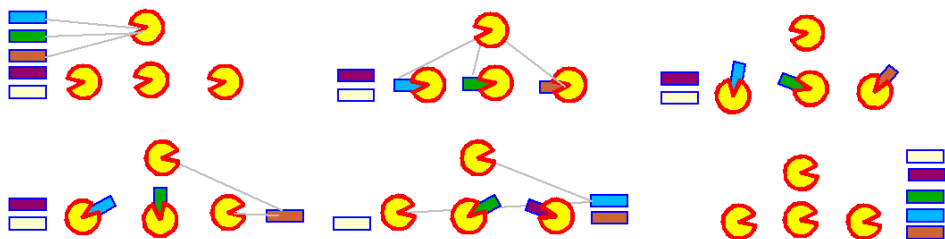


**Processors are _asynchronous_ (not in lockstep):Who finishes first is non-deterministic. Boss processor keeps track, n_proc – 1 workers do the work.  If all sub-tasks take the same time, then speedup is linear for n_tasks < n_proc-1 However..**

**Results for Indian Pines Data 16 classes: 145 x 145 pixels 200 bands/pixels => build 120 pair classifiers: 120 training tasks + 120 testing tasks + 1 voting task (could have done training+testing as a single task)**
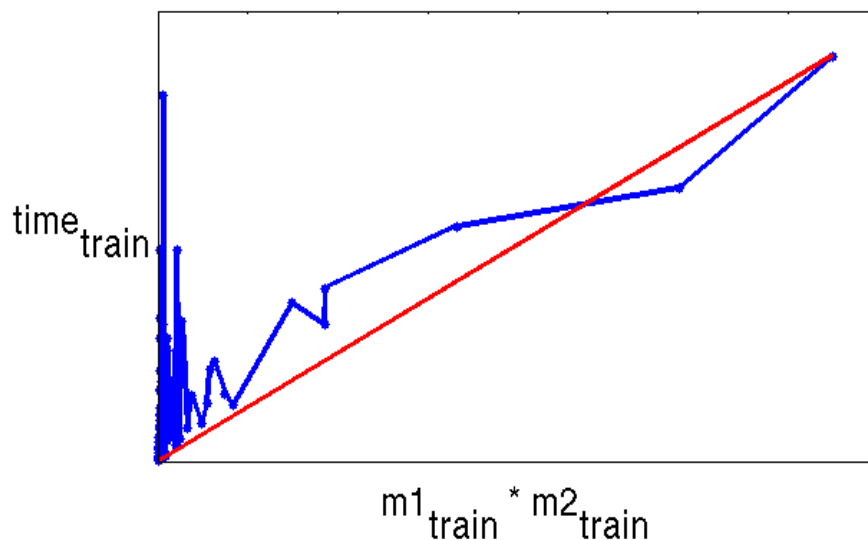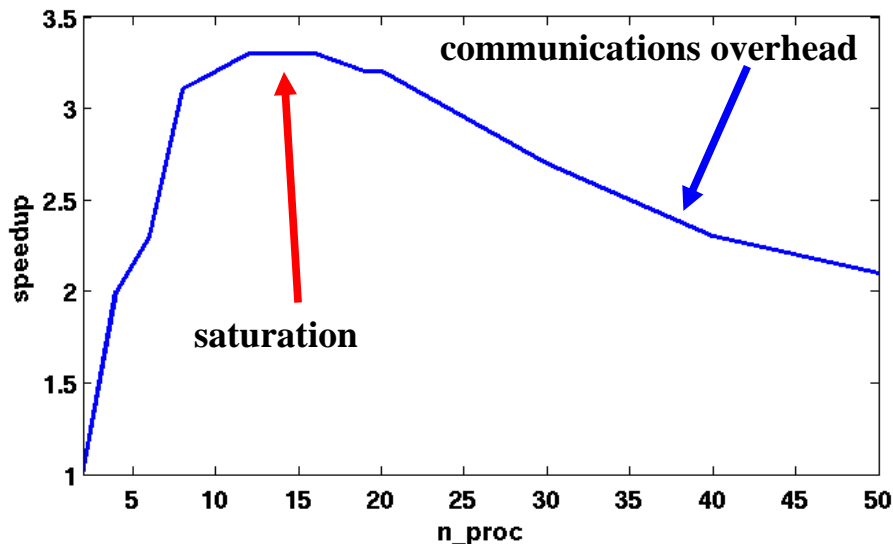
# Performance of Recursive Hierarchical Segmentation (RHSEG).-

- Key concern: selecting an appropriate dissimilarity function.

- Currently available functions: 1-norm, 2-norm, inf-norm, BSMSE, BMMSE.

- Parameter *spclust_wght* gives a lot of flexibility for the integration of spatial/spectral info.

| 1-Norm | 2-Norm | $\infty$-Norm | BSMSE | BMMSE |
|--------|--------|---------------|-------|-------|
| 91.5 | 90.5 | 89.1 | 90.5 | 89.1 |

Impact of dissimilarity function on classification of Subset #3
(Region means modeled as the mean of each ground-truth class)

| $spclust\_wght = 1.0$ | $spclust\_wght = 0.5$ | $spclust\_wght = 0.1$ |
|-----------------------|-----------------------|-----------------------|
| 90.5 (9 regions) | 96.5 (14 regions) | 97.7 (18 regions) |

Impact of *spclust_wght* on classification of Subset #3
(Region means were initialized using the ground-truth data)

**Processing times (seconds) for parallel RHSEG executed on Thunderhead (Subset #3 or ROSIS data)**

|  | 1 | 4 | 16 | 64 | 256 |
|--------|------|-----|-----|------|------|
| Time | 2061 | 568 | 155 | 48 | 25 |
| Speedup | 1.0 | 3.6 | 13.3 | 49.9 | 82.4 |



Recursive Hierarchical Segmentation Software

http://tco.gsfc.nasa.gov/RHSEG/index.html

# Conclusions

- The introduction of the concept of imaging spectroscopy by Alex Goetz established the foundations a field which is still emerging in the design of data processing techniques.

- The special characteristics of hyperspectral images pose *new processing problems*, not found in other types of remote sensing data.

- Kernel methods offer an interesting solution to deal with the *high-dimensional nature* of the data and the *limited availability of training samples* (supervised classification).

- The *integration of spatial and spectral information* allows for the development of enhanced supervised/unsupervised analysis techniques.

- Most of the algorithms discussed in this work are dominated by regular computations (appealing for the design of *parallel implementations*).

- Techniques developed in this work show the increasing sophistication of a field that is rapidly maturing at the intersection of many different disciplines.

# Advanced Processing of Hyperspectral Images

State of Science of Environmental Applications of Imaging Spectroscopy
in honor of Dr. Alexander F.H. Goetz

A. Plaza[1], J. A. Benediktsson[2], J. Boardman[3], J. Brazile[4], L. Bruzzone[5], G. Camps-Valls[6], J. Chanussot[7], M. Fauvel[2,7], P. Gamba[8], A. Gualtieri[9,10], M. Marconcini[5], J. C. Tilton[9], G. Trianni[8]

[1]Univ. Extremadura (Spain), [2]University of Iceland, [3]Analytical Imaging and Geophysics LLC (USA), [4]Univ. Zürich (Switzerland), [5]Univ. Trento (Italy), [6]Univ. Valencia (Spain), [7]INPG Grenoble (France), [8]Univ. Pavia (Italy), [9]NASA/Goddard Space Flight Center (USA), [10]Global Science & Technology (USA)