Improvements to Expectation-Maximization Approach for Unsupervised Classification of Remote Sensing Data

Thales Sehn Korting

tkorting@dpi.inpe.br

http://www.dpi.inpe.br/~tkorting

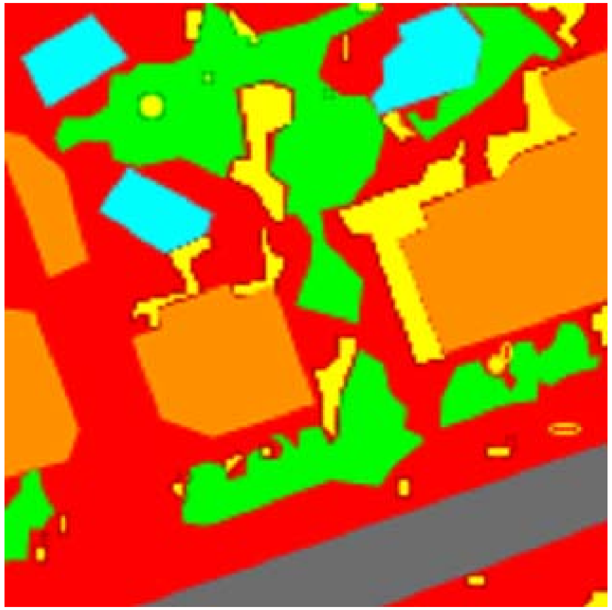
and

Luciano Dutra, Leila Fonseca

Guaraci Erthal, Felipe Silva

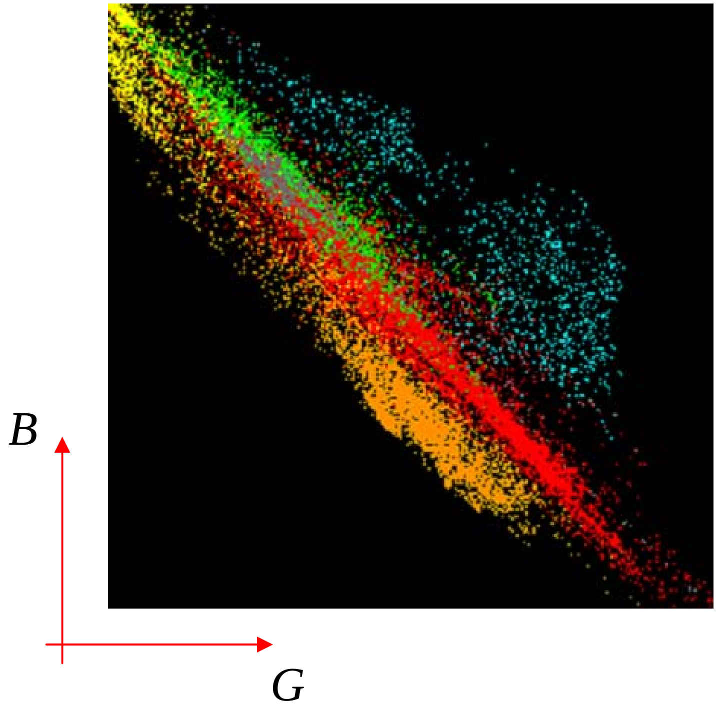
Introduction

- Color compositions of remote sensing imagery changes gradually according (x, y) axis



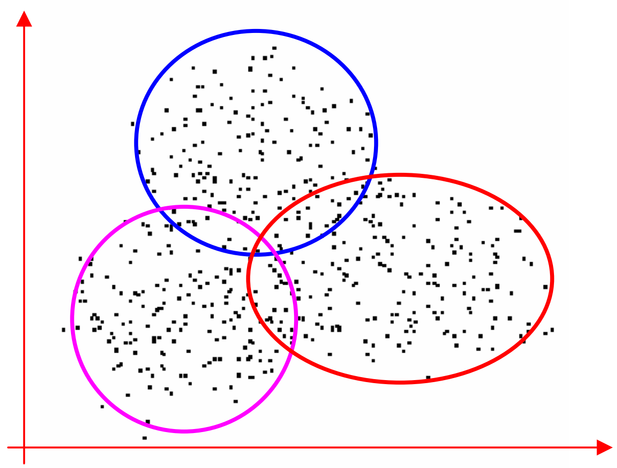
Introduction

- Linear classification is not enough



Introduction

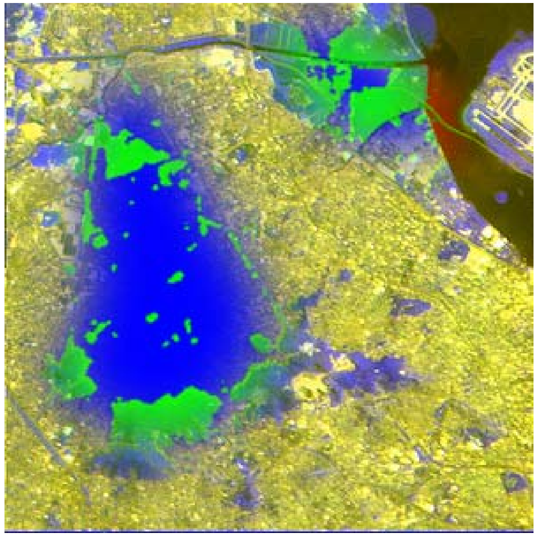
- Data points have membership in one data distribution → firstly unknown
- *Expectation-Maximization* (EM) algorithm

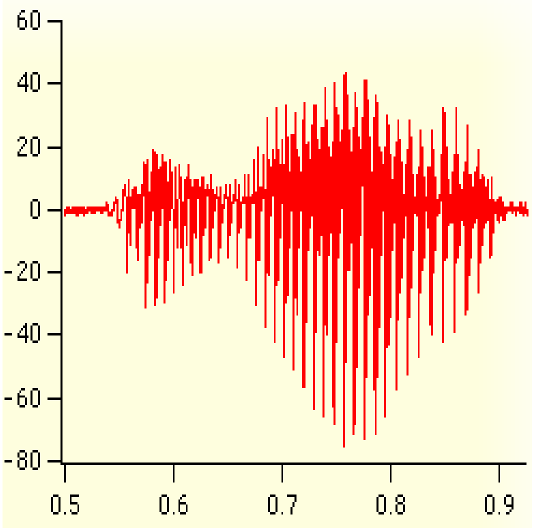


The EM algorithm

- EM is an iterative procedure, that estimates the probabilities from a set of clusters
- Input formed by the image pixels
- The parameters are the *mean* and *variance*

EM applications





The standard EM

- Each class is characterized by mean and variance of pixels
- Based on the Bayesian theory
  - Input data is composed by N vectors

$$\mathbf{x}_k, k = 1, \ldots N$$

  - Aims to estimate M clusters

$$C_j, j = 1, \ldots M$$

  - Calculates probabilities

$$P(C_j | \mathbf{x}_k)$$

Computing EM

- Successive parameter estimation

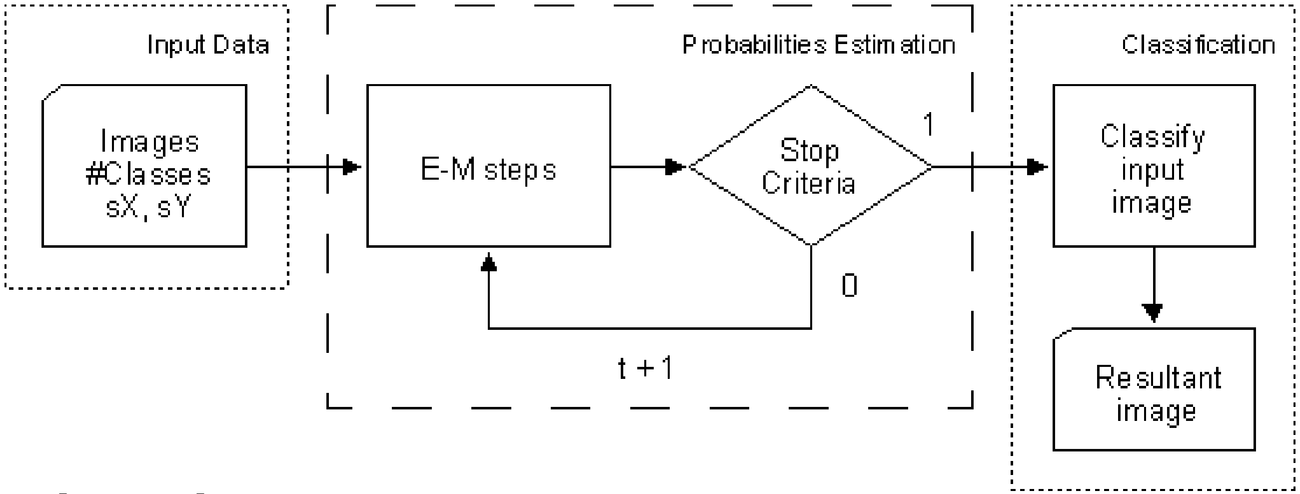
$$\hat{\theta}(t) = \{\mu_j(t), \Sigma_j(t)\}$$

$j = 1, \dots M$

$t = 0, 1, \dots$

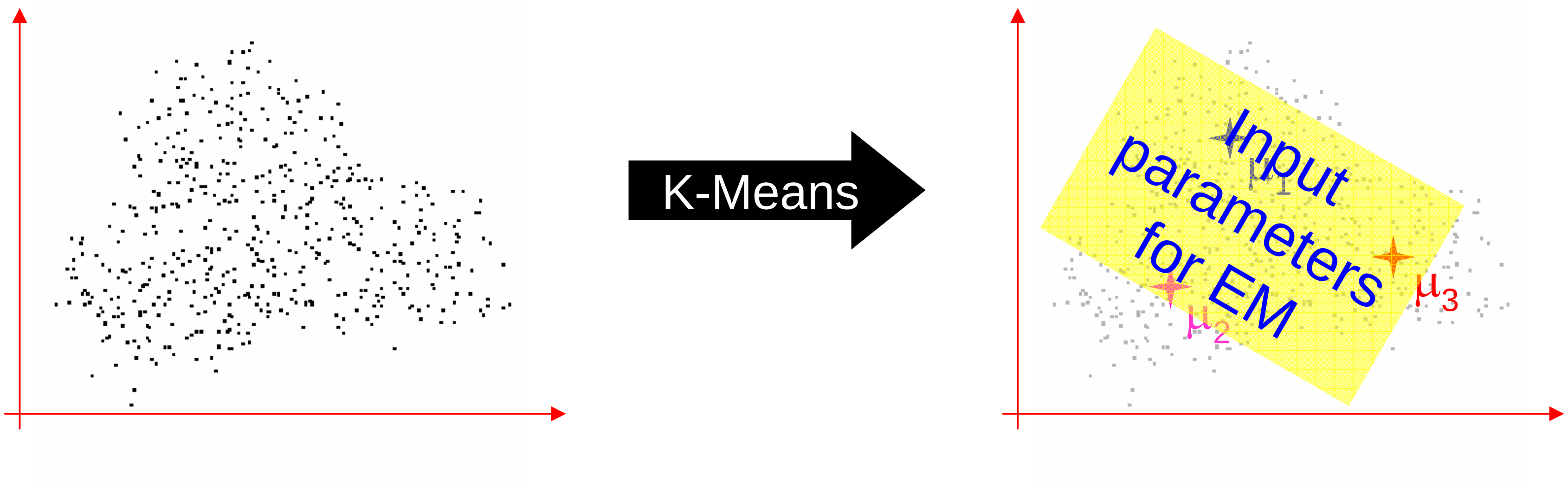
- E-step (Expectation)
  - Calculates *a posteriori* probabilities function
- M-step (Maximization)
  - Updates estimates of $\hat{\theta}(t)$

Improving the approach

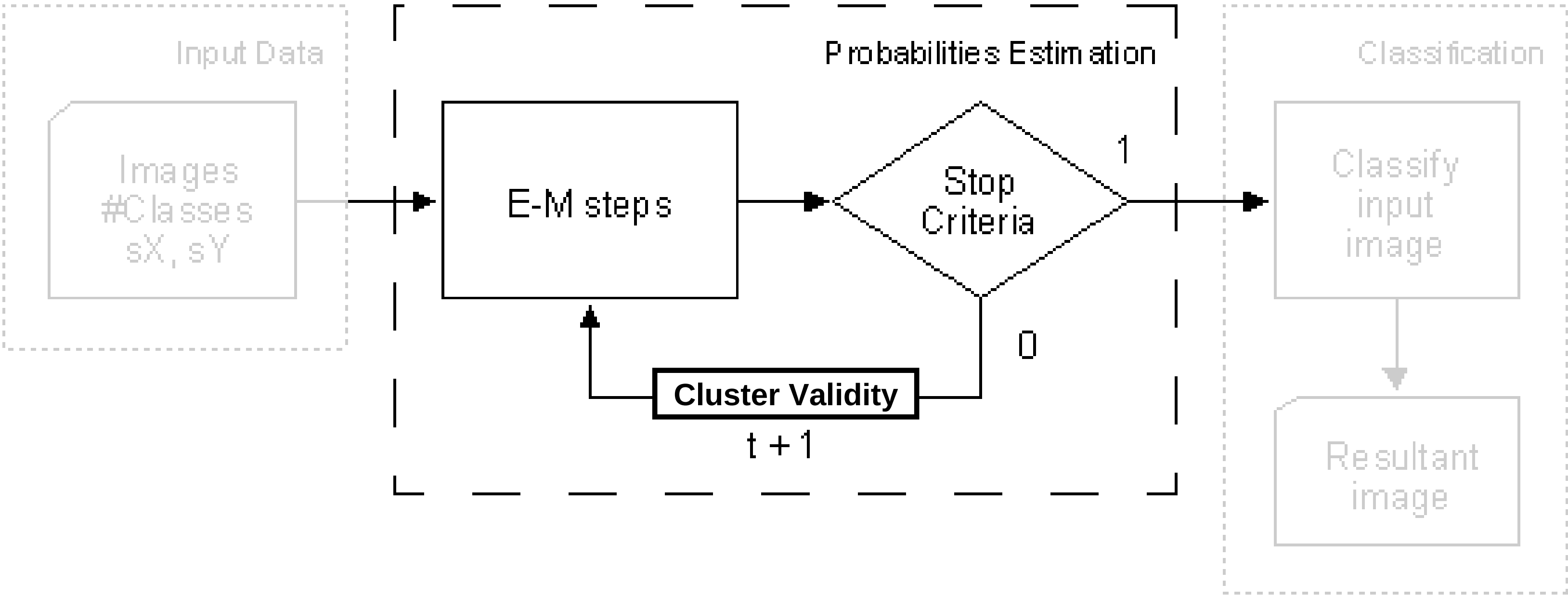


- Drawbacks
  - Random initial parameters
  - Iterative procedure can be time consuming
  - Local *minima* may prejudice results

Improving the approach – Hypothesis



Improving the approach – Cluster validity



Cluster Validity

- Performed at each iteration
- If two clusters are *close* to each other
  - Randomly modifies one of them
- If some cluster has *low* probability
  - Exclude cluster
- Thresholds (*close* and *low*) provided by the user

Results – 0

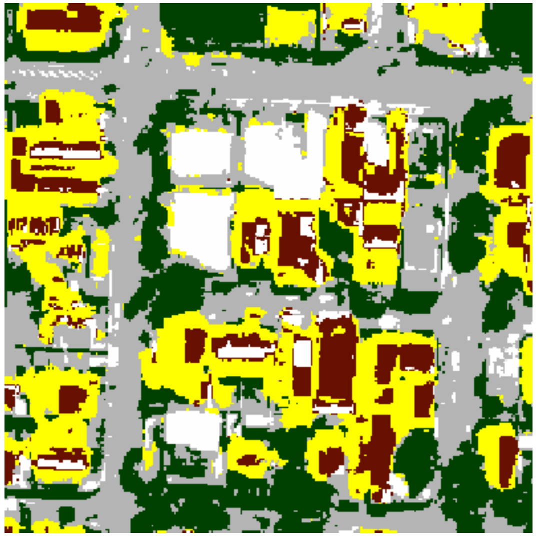
- Manual classification used as reference
- Three methods compared
  - Improved EM
  - K-Means
  - Maximum Likelihood

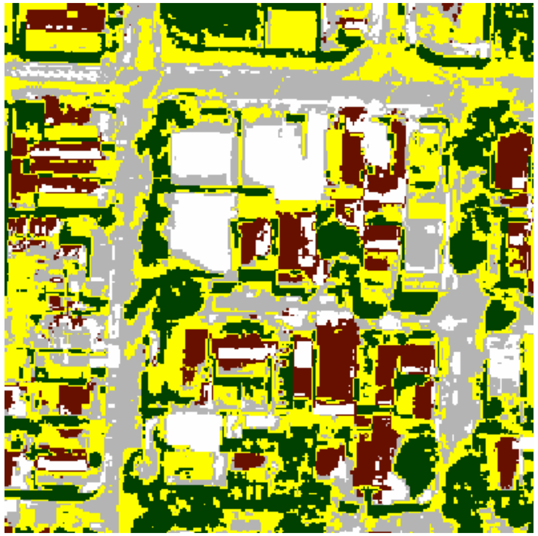
Results – 1

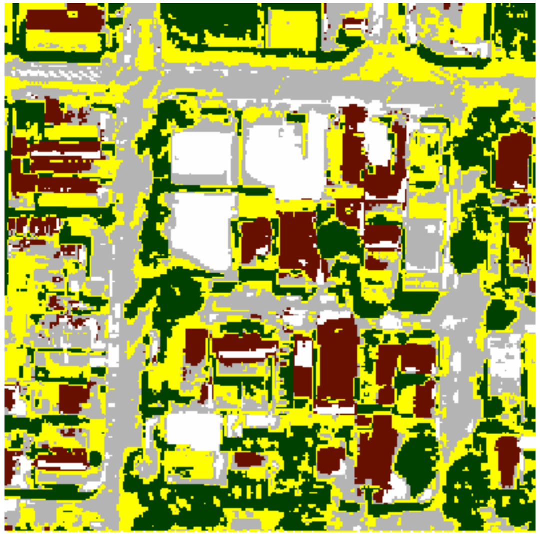
- Urban area of São José dos Campos

Comparison – Other approaches

Improved EM | K-Means | ML







Comparison – Original x Improved

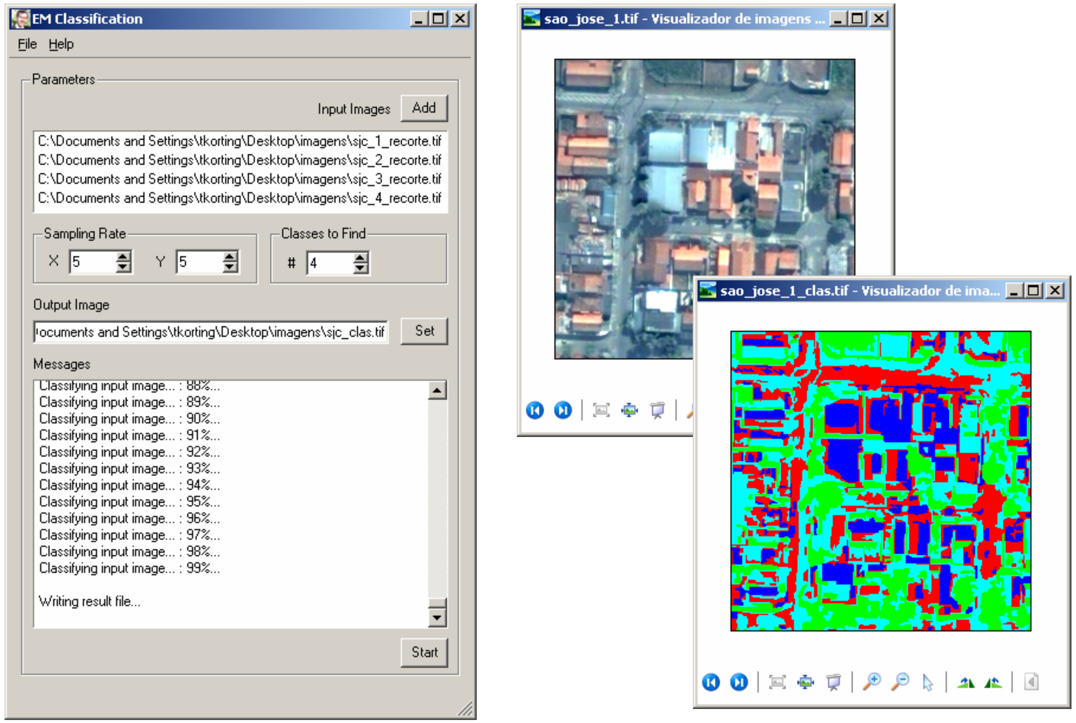
	Image1	Image2	Image3	Image4	Image5
Image size	512×512	512×512	200×200	512×384	264×377
# of classes	4	4	5	6	5
Δt_1 original EM	$467s$	$467s$	$103s$	$402s$	$202s$
Δt_2 improved EM	$140s$	$148s$	$29s$	$105s$	$70s$
$\Delta t_1 / \Delta t_2$	3.335	3.155	3.551	3.828	2.885

Conclusion

- Improved EM approach presented
  - Uses K-Means results as input
  - Performs Cluster Validation
  - Runs ~3x faster than original EM
- EM has smooth convergence
  - Invulnerable to instabilities
  - Wrong initial parameters can decrease performance

Conclusion

- Implementation using TerraLib (3.2.0 RC1)
- GUI → **`http://www.dpi.inpe.br/~tkorting`**



END

Thales Sehn Korting

tkorting@dpi.inpe.br

http://www.dpi.inpe.br/~tkorting

and

Luciano Dutra, Leila Fonseca

Guaraci Erthal, Felipe Silva