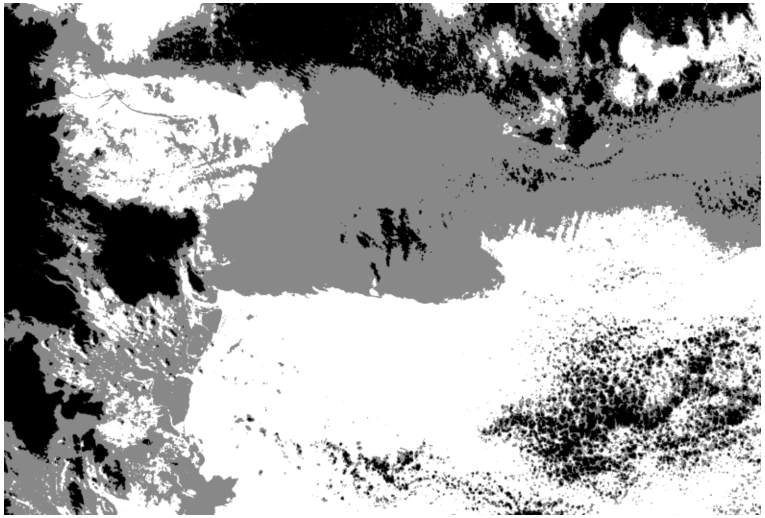
On HPC for Hyperspectral Image Processing

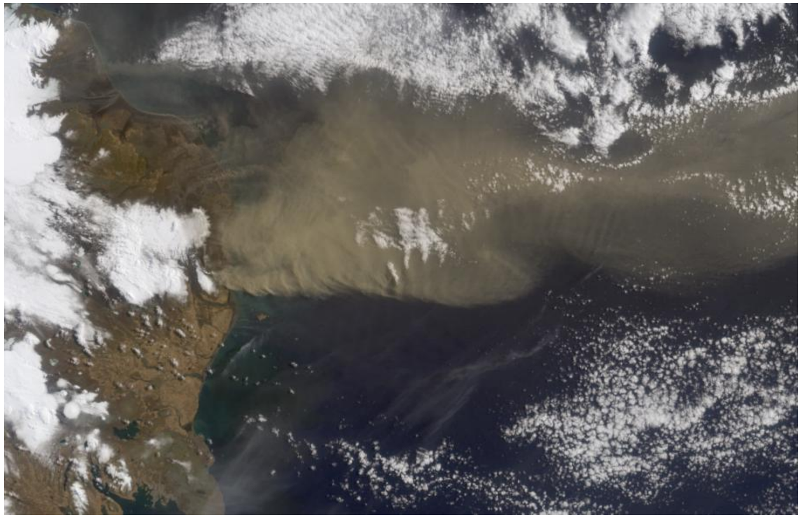
Silviu Panica, Daniela Zaharie, Dana Petcu
West University of Timisoara, Romania

Motivation

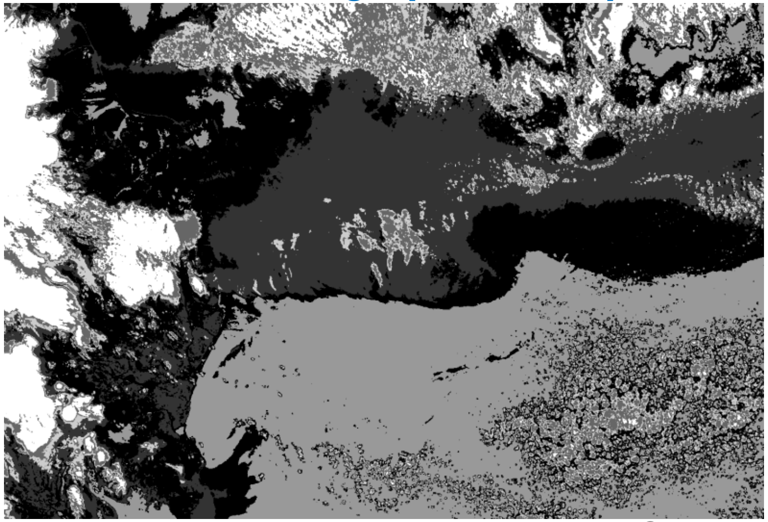
- Unsupervised classification (clustering) = identify regions in the image characterized by similar feature values



Classified image (3 classes)



Original image (Heights of the Eyjafjallajökull Eruption Plume - April 19, 2010)



Classified image (6 classes)

Motivation

Challenges in unsupervised classification of satellite images:

- Pixels may contain spectral information corresponding to different ground components
  - Possible solutions:
    - Assign each pixel to several classes based using membership values (Fuzzy Clustering)
    - Extract so-called endmembers which would correspond to pure pixels and express the image pixels as combinations of pure pixels
- The image can contain noise because of the limited sensors sensitivity
  - Possible solution: use both spectral and spatial information (spatial variants of Fuzzy Clustering)
- Increase of spatial and spectral resolution of sensors led to large images (high number of pixels and/or high number of spectral bands)
  - Possible solution: spatial (or spectral) domain partitioning and parallel (or distributed) implementation of clustering algorithms

Agenda

Parallel implementations of:

- Spatial Fuzzy C means (SFCM – [Chuang et al. 2006])
- Automated Morphological Endmember Extraction (AMEE - [Plaza et al., 2006])

Testing the parallel implementations on large images using:

- A BlueGene/P supercomputer

Comparative analysis:

- Efficiency
- Execution time

Fuzzy C-Means

- Iterative algorithm for fuzzy unsupervised classification (clustering) of images
- **Input data:**
  - Image = $\{x_1, ..., x_n\}$, n = number of pixels
    $x_i = (x_{i1},..., x_{id})$, d = number of spectral bands
    (set of vectors corresponding to all pixels and containing the values corresponding to the spectral bands)
  - Number of classes to be identified (c)
- **Output data:**
  - Membership matrix (of size c x n) $=(u_{ij})$, j=1..c, i=1..n
    u_{ij} is a value in [0,1] specifying the degree of membership of pixel i to class j
  - Classes centroids = $\{v_1,...,v_c\}$
  - Classified image = $\{y_1,...,y_n\}$, y_i = value related to the label of the class to which x_i belongs

SFCM: Spatial Fuzzy C-Means

Aim: reduce the number of spurious blobs caused by noisy pixels

Main idea: adjust the membership values using averages over a neighborhood [Chuang et al, 2006]

SFCM Algorithm

- Initialization of the membership values
- DO
  - Compute the centroids
  - Estimate the membership values (u_{ij})
  - Adjust the membership values (u'_{ij})

  WHILE (there are significant changes in the membership values)
- Construct the classification

Parallel SFCM

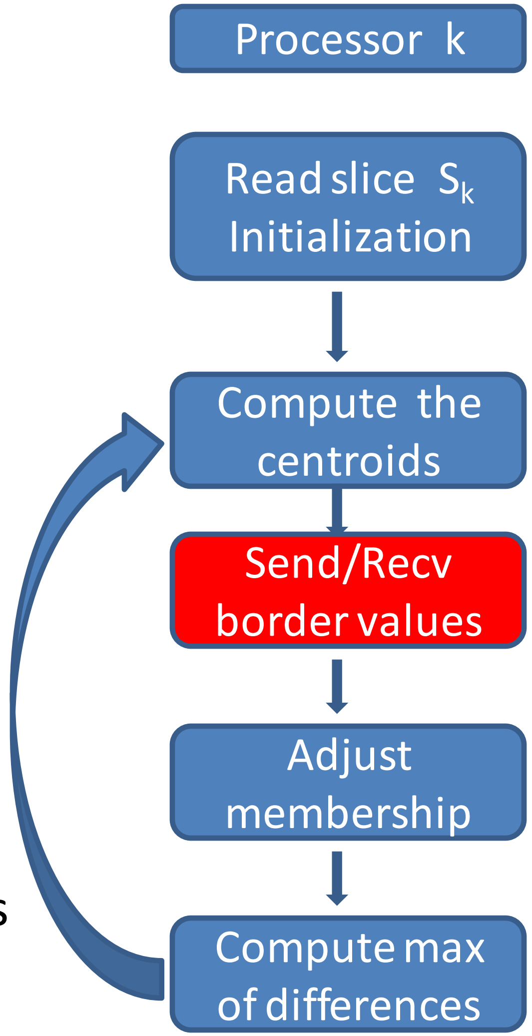
General structure: same idea as in parallel FCM

Processor k computes:

- The corresponding membership values
- The partial sums involved in the centroids computation
- The local maximal difference between the membership values at two consecutive iterations

Particularity:

- The computation of spatial information for pixels on the border of the image slice needs the communication of some membership values between processors



Automated Morphological Endmember Extraction

Algorithm 1 AMEE Parallel

Scatter N partial data structures $\{PSSP\}_{n=1}^{N}$ of F
$i = 1$
$MEI(x, y) = 0, \forall(x, y) \in PSSP_n$
while $i < I_{max}$ **do**
    Move kernel through each pixel
    Compute minimum and maximum
    Update MEI with SAM between minimum and maximum
    $i = i + 1$
    if $i == I_{max}$ **then**
        break
    else
        Replace $PSSP_n$ with its dilation
    end if
end while
Select P endmembers with highest MEI scores
Master gathers all endmembers and forms a unique set of P endmembers by computing all possible pairs

Experiments: Environment

BlueGene/ P

- Nodes: 32 nodes x 32 compute cards x 1CPU

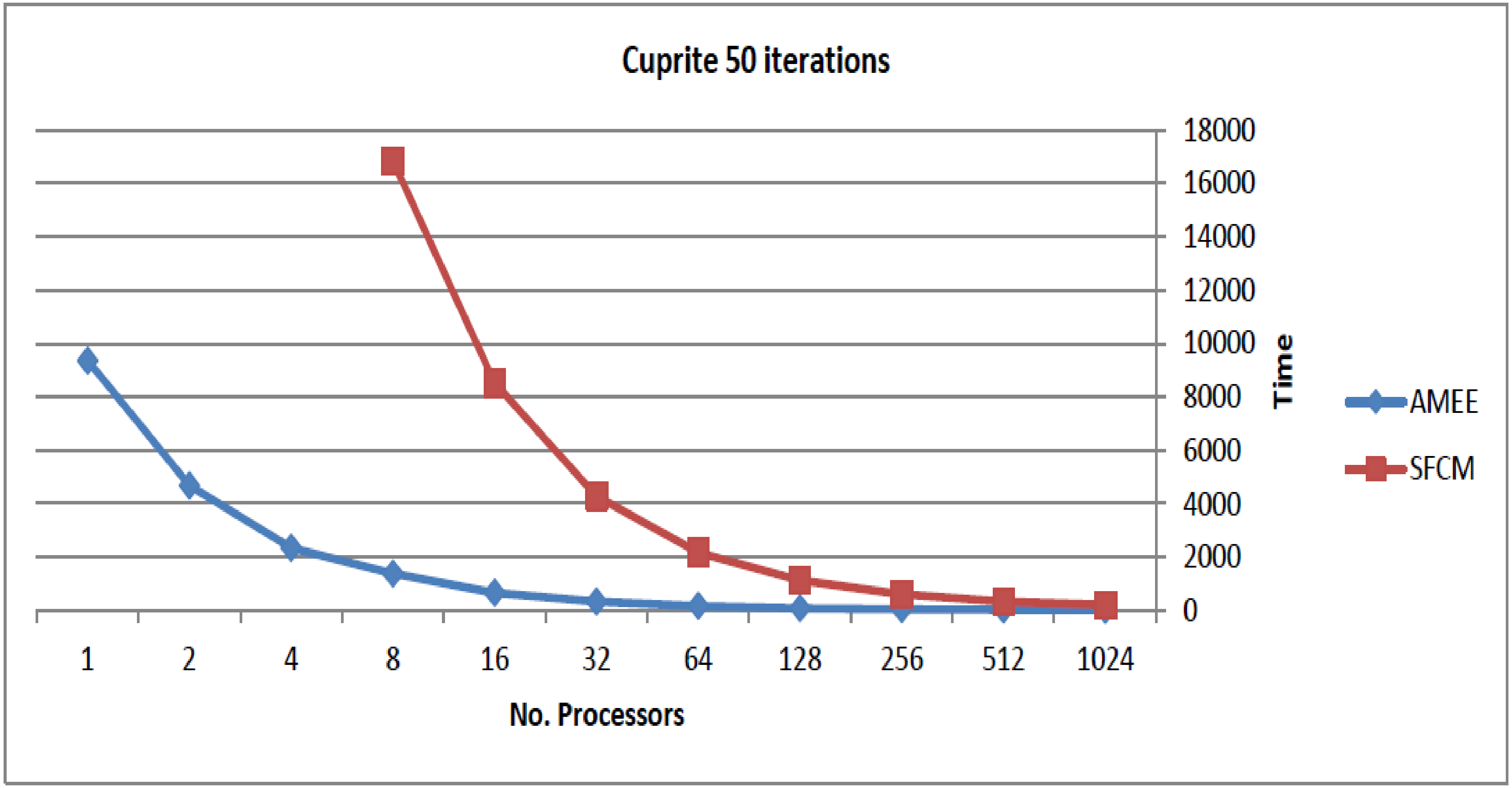
- CPU: 850Mhz PowerPC 450d, 4 cores per CPU (32 bits mode);

- RAM: 1GB / core;

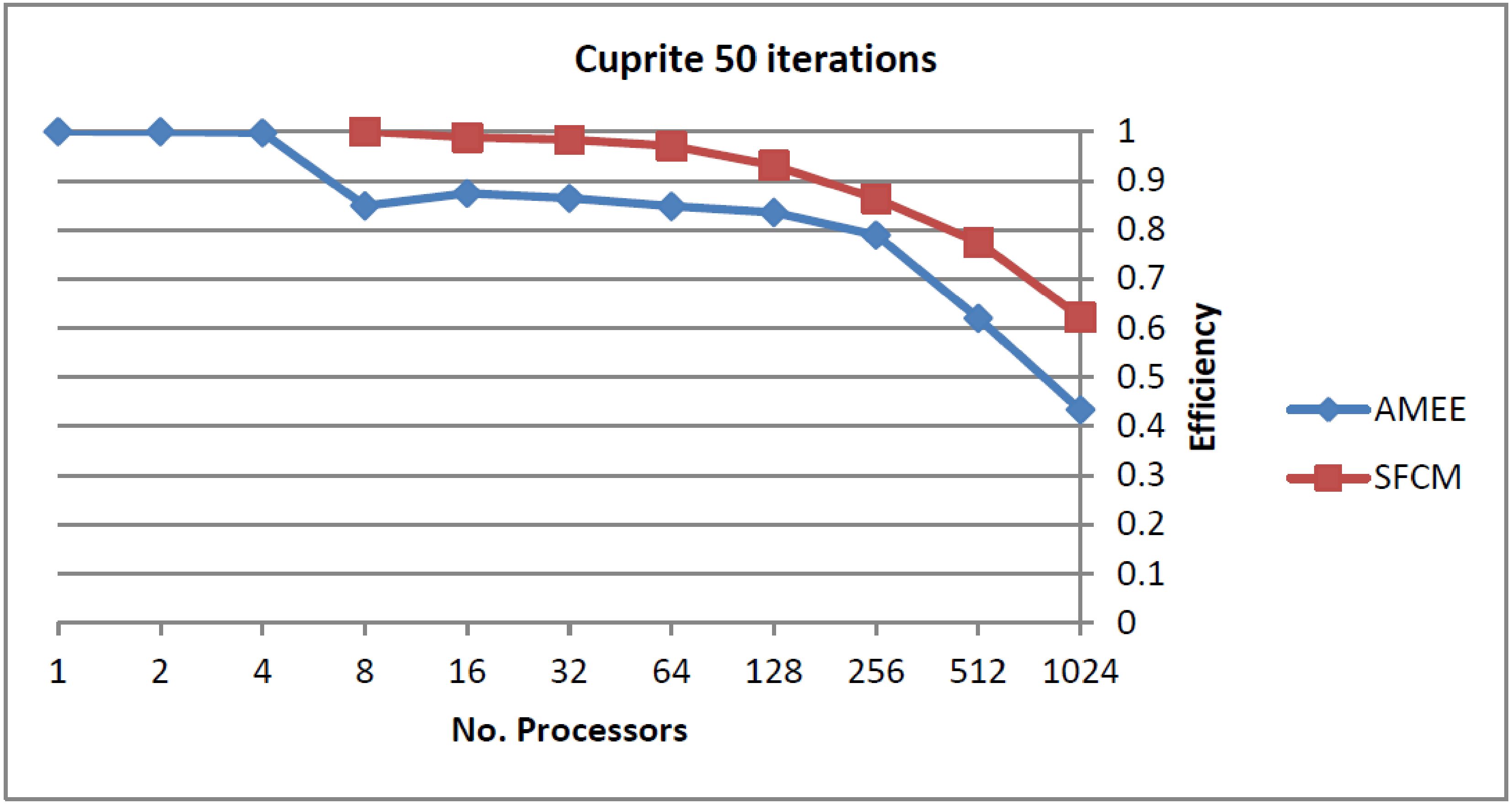
- High-speed interconnect: 3D Torus 40Gbps bandwith (3µs response time on MPI communication)

- Collective interconnect: 53Gbps bandwidth (5µs response time for MPI communication)

Results: execution time



Results: efficiency



Results: outputs

Input

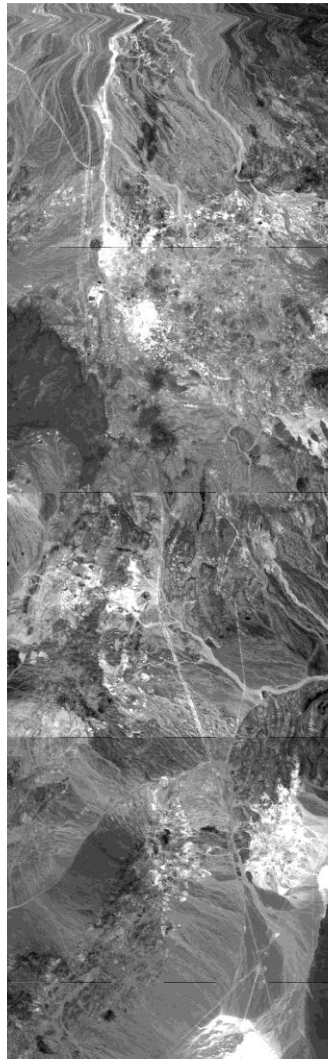
Image with 224 spectral bands

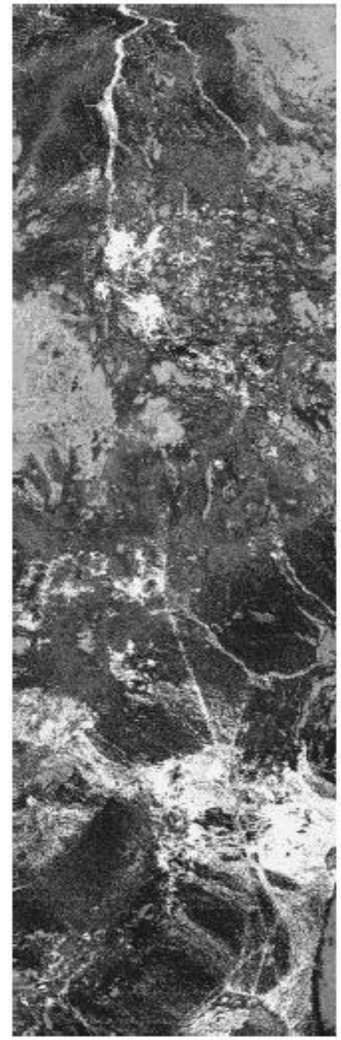
Output

(left to right):

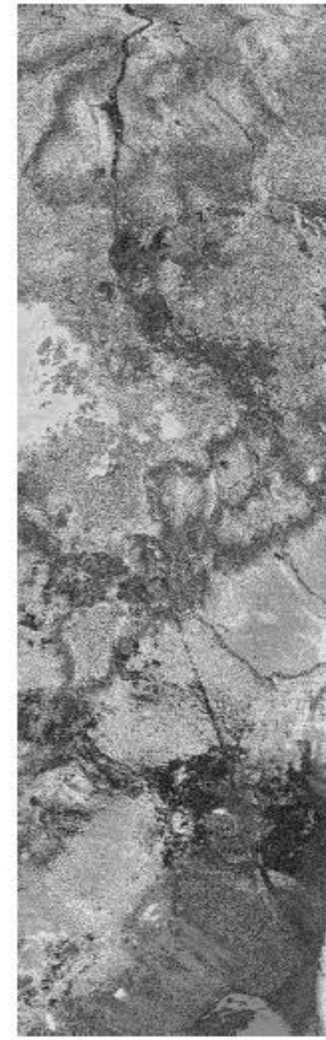
- Original
- AMEE
- SFCM

(VK= clustering quality index; smaller is better)





(a) AMEE, $V_K = 51.9$



(b) SFCM, $V_K = 13.04$

Concluding Remarks

The parallel processing of hyperspectral images raises some computational issues related to:

- Image splitting and scattering slices to processors
  - SFCM is sensitive to the splitting style while AMEE is not so sensitive
- Partitioning the computation:
  - The usage of collective operations instead of send/recv proved to be beneficial both for SFCM and AMEE
  - Synchronization (by MPI_Barrier) proved to be very useful for AMEE implementation

Further work

- Extend the comparative analysis between various endmember extraction and corresponding clustering algorithms
- Finalize the implementation of SFCM on a GPU cluster
- Conduct a systematic comparative analysis between BG/P and the GPU cluster for hyperspectral images tasks

THANK YOU!