Enseeiht

Projet Long

Mthodes de clustering parallles

Test Plan

Project Manager:
Tristan Soriano

Supervisor:
Laurent Beugnet

Client IRIT:
Ronan Guivarch
Sandrine Mouysset

Contents

1 Presentation

1.1 Name

New clustering methods integration into a parallel code.

1.2 Sponsort

IRIT : Ronan Guivarch, Sandrine Mouysset.

1.3 Presentation

The purpose of the project is to add new clustering methods into an existing piece of software. The clustering method are used on dense and sparse matrix which involve heavy computing calculus and time processing. In order to solve such issue, the existing code run on a master slave architecture. The original image is splitted among different slaves that compute initially a single clustering method: spectral clustering.

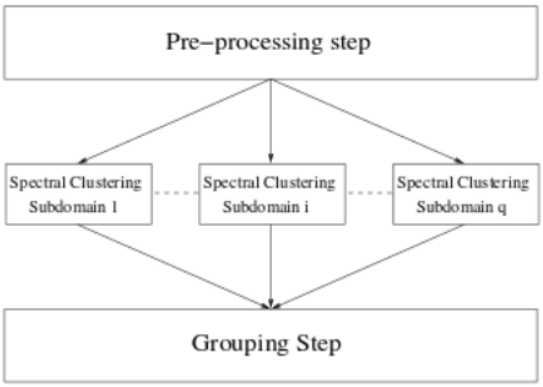

Figure 1:

The objective is first to add two methods : Kernel K-means and Mean shift in such structure then to refactore/clean the existing code.

2 Environment

2.1 Hardware environment

As it is a parallel computing program. The environment must enable the team to test on single and on multiple entity the code produced. In order to do so the code will be tested on lab machines, which are the most likely to represent the IRIT environment.

2.2 Software environment

This project implies the use of multiple technologies: Matlab, Fortran, MPI, ssh, Doxygen. The test must run on the client machine configuration.

3 Test definition

The following test must show the client that the code fit the specifications. It include non-regression test : the original code will be tested on the team configuration as a performance reference and will be compared to the new implemented algorithm result in term of efficiency and quality.

Id: T001 **Description**: This is the initial test, all the initial function of the program must run.
bouquet, 3blocSimples, Cible, arthus will be tested
Result : The program must run on 2D, 3D geometrical and color images.
Bouquet: It must produce 4 clusters : leaf, oranges, tree, backgroung.
Cible: It must produce 1 cluster per ring.
3blocSimples: 3 clusters
Arthus: It must run, find clusters in a limited time.

Id: T002 **Description**: We must test the matlab implmentation of Mean Shift
bouquet, 3blocSimples, Cible will be tested
Result: The program must run on 2D, 3D geometrical and color images.
Bouquet: It must produce 4 clusters : leaf, oranges, tree, backgroung.
3blocSimples: 3 clusters

Cible: It must produce 1 cluster per ring.

Id: T003 **Description**: We must test the matlab implmentation of Kernel K-Means
bouquet, 3blocSimples, Cible will be tested
Result: The program must run on 2D, 3D geometrical and color images.
Bouquet: It must produce 4 clusters : leaf, oranges, tree, backgroung.
Cible: It must produce 1 cluster per ring.
3blocSimples: 3 clusters

Id: T004 **Description**: We must test the Fortran implmentation of Mean Shift
bouquet, 3blocSimples, Cible will be tested
Result: The program must run on 2D, 3D geometrical and color images.
Bouquet: It must produce 4 clusters : leaf, oranges, tree, backgroung.
Cible: It must produce 1 cluster per ring.
3blocSimples: 3 clusters

Id: T005 **Description**: We must test the Fortran implmentation of Mean Shift
Arthus1 will be tested with different bandwidth
Result: The program must run on 2D, 3D geometrical and color images.
Arthus 1 must produce an image segmentation result in limited time

Id: T006 **Description**: We must test the Fortran implmentation of Kernel K-means
bouquet, 3blocSimples, Cible will be tested:

1. using polynomial kernel function
2. using gaussian kernel function

Result: The program must run on 2D, 3D geometrical and color images.
Bouquet: It must produce 4 clusters : leaf, oranges, tree, backgroung.
Cible: It must produce 1 cluster per ring.

3blocSimples: 3 clusters

Id: T007 **Description**: We must test the initial implmentation of Spectral clustering using different Kernel function
bouquet, 3blocSimples, Cible will be tested:

1. using polynomial kernel function
2. using gaussian kernel function

Result: The program must run on 2D, 3D geometrical and color images.
Bouquet: It must produce 4 clusters : leaf, oranges, tree, backgroung.
Cible: It must produce 1 cluster per ring.
3blocSimples: 3 clusters

The test T005 o T007 will be duplicated for parallel computing
Thoose test will cover:

1. Matlab implementation of Kernek K-means
2. Matlab implementation of Mean shift
3. Fortran implementation of Kernek K-means
4. Fortran implementation of Mean shift
5. Fortran implementation of Kernel methods
6. Integration of the methods into a parallele code

Id: T008 **Description**: We must show that the code refactoring has een succesull
The code will e satical analysed to show the improvement

Result: The code readability, syntax must have been improved and standards must have applied : one language, one naming standard etc