Geophysical Research Abstracts
Vol. 15, EGU2013-7138, 2013
EGU General Assembly 2013

Fast variogram analysis of remotely sensed images in HPC environment

Lluís Pesquer (1), Anna Cortés (2), Joan Masó (1), and Xavier Pons (3)

(1) CREAF, Cerdanyola del Vallès 08193, Spain (l.pesquer@creaf.uab.cat), (2) Computer Architecture and Operating Systems Department. Edifici Q. Universitat Autònoma de Barcelona, Cerdanyola del Vallès 08193, Spain (ana.cortes@uab.es), (3) Geography Department. Edifici B. Universitat Autònoma de Barcelona, Cerdanyola del Vallès, 08193,Spain (xavier.pons@uab.cat)

Exploring and describing spatial variation of images is one of the main applications of geostatistics to remote sensing. The variogram is a very suitable tool to carry out this spatial pattern analysis. Variogram analysis is composed of two steps: empirical variogram generation and fitting a variogram model. The empirical variogram generation is a very quick procedure for most analyses of irregularly distributed samples, but time consuming increases quite significantly for remotely sensed images, because number of samples (pixels) involved is usually huge (more than 30 million for a Landsat TM scene), basically depending on extension and spatial resolution of images. In several remote sensing applications this type of analysis is repeated for each image, sometimes hundreds of scenes and sometimes for each radiometric band (high number in the case of hyperspectral images) so that there is a need for a fast implementation.

In order to reduce this high execution time, we carried out a parallel solution of the variogram analyses. The solution adopted is the master/worker programming paradigm in which the master process distributes and coordinates the tasks executed by the worker processes. The code is written in ANSI-C language, including MPI (Message Passing Interface) as a message-passing library in order to communicate the master with the workers. This solution (ANSI-C + MPI) guarantees portability between different computer platforms.

The High Performance Computing (HPC) environment is formed by 32 nodes, each with two Dual Core Intel(R) Xeon (R) 3.0 GHz processors with 12 Gb of RAM, communicated with integrated dual gigabit Ethernet. This IBM cluster is located in the research laboratory of the Computer Architecture and Operating Systems Department of the Universitat Autònoma de Barcelona.

The performance results for a 15km x 15km subcene of 198-31 path-row Landsat TM image are shown in table 1. The proximity between empirical speedup behaviour and theoretical linear speedup confirms a suitable parallel design and implementation applied.

N Workers	Time (s)	Speedup
0	2975.03	
2	2112.33	1.41
4	1067.45	2.79
8	534.18	5.57
12	357.54	8.32
16	269.00	11.06
20	216.24	13.76
24	186.31	15.97

Furthermore, very similar performance results are obtained for CASI images (hyperspectral and finer spatial resolution than Landsat), showed in table 2, and demonstrating that the distributed load design is not specifically defined and optimized for unique type of images, but it is a flexible design that maintains a good balance and scalability suitable for different range of image dimensions.

N Workers	Time (s)	Speedup
0	5485.03	
2	3847.47	1.43
4	1921.62	2.85
8	965.55	5.68
12	644.26	8.51
16	483.40	11.35
20	393.67	13.93
24	347.15	15.80

28 306.33 17.91
32 304.39 18.02

Finally, we conclude that this significant time reduction underlines the utility of distributed environments for processing large amount of data as remotely sensed images.