A Parallel Algorithm for Geoprocessing of Vegetation Indices

Albert G. Saribekyan

Institute for Informatics and Automation Problems of NAS RA
e-mail: albert_saribekyan@ipia.sci.am

Abstract

Geographic information systems (GIS) play a vital role in environment-related issues, from which well-known is the calculation of vegetation indices (VI) using the satellite images. In this paper a parallel algorithm for geoprocessing of VI's is introduced with appropriate benchmarkings, which were performed using high-performance computing (HPC) resources.

Keywords: GIS, GRASS, Vegetation Indices, Chunk, Region, NDVI, GARI, GVI.

1. Introduction

Earth remote sensing plays an important role in our reality which is immediately associated with satellite images. Satellite images contain large amount of useful information. The process of extracting this information and understanding it may require huge computing power and processing time. Distributed computing can reduce the processing time by providing more computational power. Geographic Resources Analysis Support System (GRASS) GIS [1] is an open source software/tool, which has been used to process satellite images. Inside GRASS, different modules have been developed for processing satellite images. There are some operations and modules that are widely used in the GRASS GIS, for instance, the calculation of VI [2] is a very important and frequently used operation, so it was used as a test example for this study. In order to make the calculation of VI more convenient with GRASS-GIS a special module i.vi [3] has been developed to calculate 13 different VI from raw spatial data. GRASS GIS module i.vi works with raster images (rows x columns). Different band raster [4] images are required for calculation of different indices:

- NDVI (Normalized Difference VI), RVI (Ratio VI) - red and nir (near infrared)
- ARVI (Atmospherically Resistant VI) - red, nir and blue
- GVI (Green VI) - red, nir infrared, blue, green, band5 and band7

- GARI (Green Atmospherically Resistant VI) - red, nir, blue, green
- RVI (Ratio VI) - red and nir
- IPVI (Infrared Percentage VI) - red and nir
- DVI (Difference VI) - red and nir
- PVI (Perpendicular VI) - band5 and band7
- WDVI (Weighted Difference VI) - red and nir
- SAVI (Soil Adjusted VI) - red and nir
- MSAVI (Modified Soil Adjusted VI) - red and nir
- MSAVI2 (Second Modified Soil Adjusted VI) - red and nir
- GEMI (Global Environmental Monitoring VI) - red and nir

GRASS functions are used to extract row wise data from the specific band images and store them in buffers, and then each column value is extracted sequentially from the buffers and sent them for generating the specific VI values. Thus, after completing the VI from row buffers, the row wise VI values are put back into an output image and this process is applying recursively for each row.

2. Parallelization Algorithm

A parallelization algorithm that splits the geographic region of the target location into specified number of chunks has been implemented after series of experiments using several parallelization approaches. The algorithm allows to calculate each chunk separately in a parallel way and then it combines all the parts together when it finishes the calculations.

The parallelization approach consists of three main stages:

- splitting
- calculation
- merging

It is known that in GRASS GIS, a region refers to a geographic area with some defined boundaries, based on a specific map coordinate system (WGS 84[5]) and map projection. One of the features of GRASS GIS is to perform calculations only on that part of raster data which is defined by the current region. So by changing the coordinates of the current region it is possible to change the part on the raster image on which the calculation will take place. This feature is used to perform the first stage of parallelization – splitting. A shell script has been developed, which in the first step reads the northern and southern boundaries of the image and after dividing the area between those boundaries into specified number of pieces, defines new northern and southern boundaries for each so called chunk, these new defined boundaries are then saved into the named regions. The region's boundaries are given as the northernmost, southernmost, easternmost, and westernmost points that define its extent (cell edges). The north and south boundaries are commonly called northings, while the east and west boundaries are called eastings. Each mapset has a current geographic region, which defines the geographic area for raster analyses. Raster data, in case of necessity is resembled, in order to meet the cell resolutions of the current geographic region setting. Each GRASS GIS location has a fixed geographic region, called the default geographic region that defines the extent of the database. The current region can easily be reset to the default region. Each MAPSET may contain any number of pre-defined and named geographic regions, called saved regions. Any of these pre-defined geographic regions is possible to select by name in order to define it as a current geographic region.

In the next stage, the script runs a specified number of parallel processes; each of this process sets current region of the default mapset PERMANENT to one of the saved regions and does the

calculation on its part of image. After the parallel calculation is completed, there comes a time for merging and exporting the processed chunks of data.

In the third stage of parallelization approach the merging process is being started and begins to combine the split regions on which the calculation has already been done. After that the complete image will be obtained. This operation has been done using gdal_merge.py [6] utility of Geospatial Data Abstraction Library (GDAL) [7]. The main important aspect is that all input images have to be in the same coordinate system. The gdal_merge.py operates outside GRASS so the chunks of the image need to be exported and merged outside GRASS. So, when all the data is processed all the chunks saved as a raster pictures exported from GRASS and they will be combined with gdal_merge.py utility (see fig. 1).
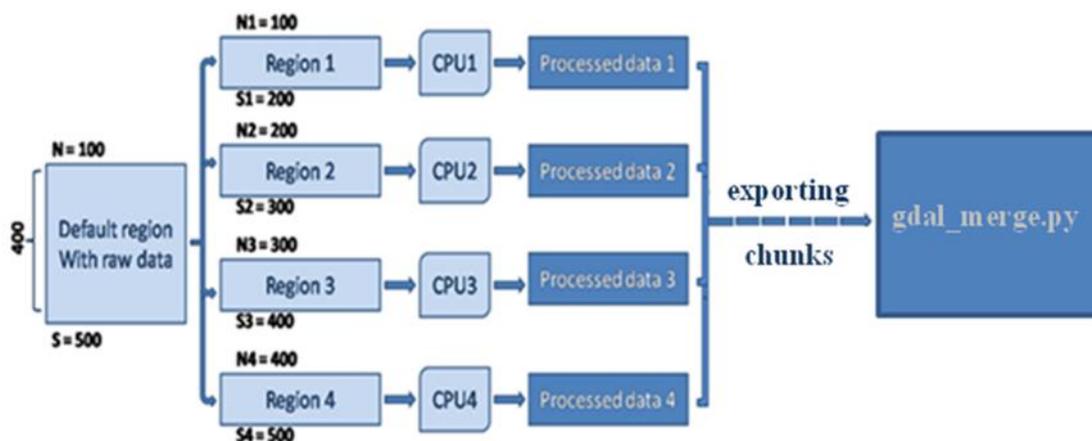

Fig. 1. The structure of parallelization algorithm

3. Benchmarkings

In order to find out a general parallelization approach for operations with different complexity, a set of benchmarkings have been carried out. As a target data for the experiments the Landsat Thematic Mapper (TM) image is used [8], which consists of 7 bands and 30 m (98 ft) spatial resolution for bands used in VI calculations (bands 1-5, 7). Every band consists of 7130 rows and 7844 Columns, so the total number of cells is 55927720. The experiments have been carried out using GRASS GIS 7.0.svn and some computational resources (1 node with the following parameters - 8 cores, Intel E5420, 8 GB RAM) of the Armenian National Grid Infrastructure [9]. As there are at least two ways to split the image into chunks (horizontal and vertical), the both ways have been implemented to find out the optimal option. Since horizontal splitting of chunks is faster (speedup of vertical efficiency is 26.5% instead of horizontal – 35.2%) than those with vertical one, then for the future experiments the horizontal splitting is used.

From technical point of view, the calculation of NDVI index is the simplest one among 13 VIs of i.vi. It uses only 2 bands, redchan and nirchan (3-rd and 4-th bands of Landsat TM/ETM+ images). The algorithm of calculation consists of 3 operations: *NDVI = (nirchan - redchan) / (nirchan + redchan)*. In case of GARI, it uses 4 bands, redchan, nirchan, bluechan and greenchan (3-rd, 4-th, 1-st and 2-nd bands of Landsat TM/ETM+ images). The algorithm of calculation of GARI is more complex: *GARI = nirchan - (greenchan - (bluechan - redchan)) / (nirchan- (greenchan- (bluechan - redchan)))*. The GVI is the last studied and the most studied complex VI. It uses 6 bands of spatial image bluechan, greenchan, redchan, nirchan, chan5chan and chan7chan. The calculation algorithm is also more complex than the previous examples: GVI = *(-0.2848 * bluechan - 0.2435 **

*greenchan - 0.5436 * redchan + 0.7243 * nirchan + 0.0840 * chan5chan-0.1800 * chan7chan).* Fig. 2 presents the parallel calculation speedup of each of these indices.
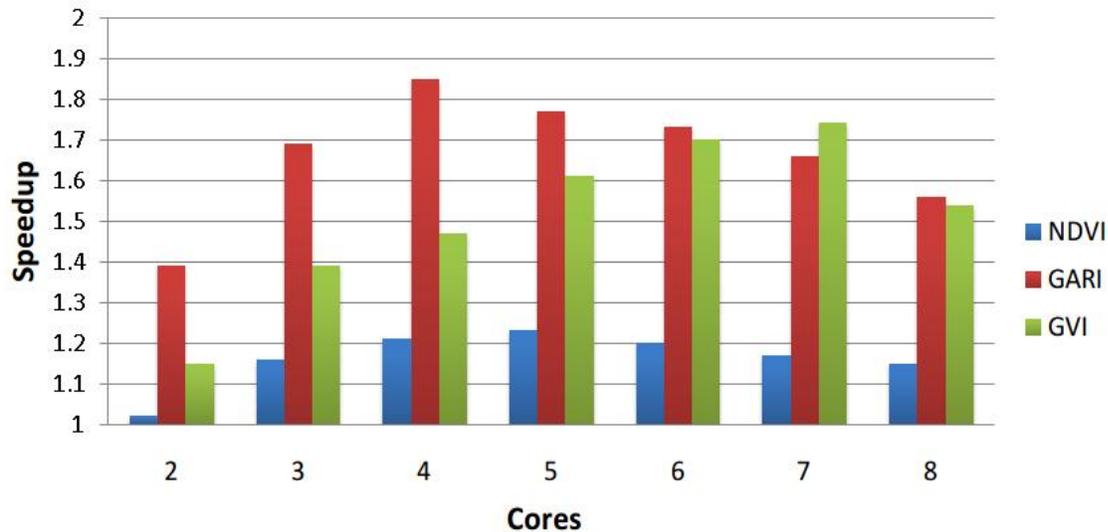

Fig. 2. Parallelization of VI Calculations

4. Conclusion

The series of experiments show that the suggested parallel algorithm significantly decreases the calculation time using the standards VI modules of GRASS GIS. The i.vi module of GRASS GIS is used for the experiments mainly concentrated on 3 from 13 Vis, which differ from each other from technical point of view. The results show that the eficiency of the parallelization depends on the complexity of the algorithm of the calculation of indices. It is planned to continue the experiments with other GRASS GIS modules, too.

Acknowledgements

References

[1] M. Neteler and H. Mitasova, *Open Source GiS: A GRASS GIS Approach*,
   Second Edition, Kluwer Academic Publishers, 2003.

[2] C. Ünsalan and K. L. Boyer, "Linearized vegetation indices, multispectral satellite image understanding", *Advances in Computer Vision and Pattern Recognition*, pp 19-39, 2011.

[3] i.vi calculates different types of vegetation indices, [Online]. Available:

http://grass.osgeo.org/grass70/manuals/i.vi.html

[4] Raster bands, [Online]. Available:
http://edndoc.esri.com/arcsde/9.2/concepts/rasters/entities/rasterbands.htm

[5] Defense Mapping Agency, *"Geodesy for the Layman"*, chapter 8.

[6] Description of gdal_merge.py utility. [Online]. Available:
http://www.gdal.org/gdal_merge.html

[7] G. Brent Hall, Michael G. Leahy, *Open Source Approaches in Spatial Data Handling*, Springer Berlin Heidelberg, 2008.

[8] National Aeronautics and Space Administration, *"Landsat 7 Science Data Users Handbook",* [Online]. Available: http://landsathandbook.gsfc.nasa.gov/pdfs/Landsat7_Handbook.pdf

[9] H. Astsatryan, V. Sahakyan and Yu. Shoukourian, "Brief introduction of Armenian National Grid Initiative", *Book of Abstracts of 4th International Conference "Distributed Computing and Grid-technologies in Science and Education"* (Grid'2010), pp. 30-31, June 28 - July 3, Dubna, Russia, 2010.

Ջուգահեռ ալգորիթմ բուսականության ցուցիչների աշխարհագրական հաշվարկների համար

Ա.Սարիբեկյան

Ամփոփում

Երկրատեղեկատվական համակարգերը մեծ նշանակություն ունեն շրջակա միջավայրի հետազոտման խնդիրներում, որոնցից առավել հայտնի է բուսականության ցուցիչների ստացումը տիեզերական պատկերների միջոցով: Տվյալ աշխատանքում ներկայացված է բուսականության ցուցիչների հաշվարկման համար զուգահեռ ալգորիթմ՝ համապատասխան վերլուծություններով, որոնք իրականացվել են բարձր արտադրողականության հաշվողական համակարգերի վրա:

.

,
.

,
.