



SATELLITE IMAGE PROCESSING ON A GRID-BASED PLATFORM

Dana Petcu ¹⁾, Dorian Gorgan ²⁾, Florin Pop ³⁾, Dacian Tudor ⁴⁾, Daniela Zaharie ¹⁾

- ¹⁾ Computer Science Department, Western University of Timisoara, B-dul Vasile Parvan 4, 300223 Timisoara, Romania, {petcu,dzaharie}@info.uvt.ro, <http://web.info.uvt.ro/~{petcu,dzaharie}>
- ²⁾ Computer Science Department, Technical University of Cluj Napoca, Str. G.Baritiu 26, 400027 Cluj-Napoca, Romania, Dorian.Gorgan@cs.utcluj.ro, <http://users.utcluj.ro/~gorgan>
- ³⁾ Computers Department, University Politehnica of Bucharest, Splaiul Independentei 313, 060032 Bucharest, Romania, florinpop@cs.pub.ro, <http://www.florinpop.ro>
- ⁴⁾ Computer Science Department, Politehnica University of Timisoara, B-dul Vasile Parvan 2, 300223 Timisoara, Romania, dacian@cs.utt.ro <http://www.cs.utt.ro/~dacian>

Abstract: *Satellite image processing is both data and computing intensive, and, therefore, it raises several difficulties or even impossibilities while being using one single computer. Moreover, the analysis and sharing of the huge amount of data provided daily by the space satellites is a major challenge for the remote sensing community. Recently, Grid-based platforms were built to address these issues. This paper presents a specialized Grid-based platform developed to enable remote sensing image processing for environmental problems, like preventing river floods or forest fires. Moreover, it exposes the novelty elements that distinguish it from other similar approaches.*

Keywords: *Satellite images, Grid services, Meteorology, Environment supervising, Parallel algorithms.*

1. INTRODUCTION

The Grid technologies promised to make feasible the creation of a computational environment handling hundreds of distributed databases, heterogeneous computing resources, and simultaneous users. There are at least three reasons for using Grid computing for satellite image processing: *location* – the required computing performance is not available locally, the solution being the remote computing; *co-operation* – the required computing performance is not available in one location, the solution being cooperative computing; and *specialization* – the required computing services are only available at specialized centers, the solution being application specific computing.

The Romanian national collaborative project MedioGrid (<http://mediogrid.utcluj.ro>) proposes several approaches to tackle with the above mentioned reasons. It aims to accomplish a pilot program to process acquired images from meteorological and resource satellites, in order to extract the meteorological and environmental parameters that characterize the atmospheric and terrestrial state. More concretely, the main objectives of the MedioGrid are:

- to develop a Grid structure to support the parallel and distributed processing of geographical and environment data;
- to develop algorithms for Grid-based processing

of satellite images;

- to develop the MedioGrid's Software Platform with enabled support for Grid services;
- to develop environment supervising applications with data extracted from satellite images;
- to model/visualize the virtual geographical space

The aim of the paper is to present the most recent achievements obtained in the project and to extend the results presented in [23]. MedioGrid is not the first initiative to use Grids for satellite image processing. After the presentation in Section 2 and 3 of the architecture and the main components of the infrastructure, Section 4 discusses the originality of the new platform by comparing with the other former or existing ones. Finally, some conclusions are drawn.

2. ARCHITECTURE AND COMPONENTS OF MEDI GRID PLATFORM

A Grid infrastructure based on a virtual organization and more than fifty workstations owned by seven different partner institutions was built as MedioGrid test-bed. Figure 1 refers to the connection scheme – labels indicate each partner's resources. Figure 2 suggests the way a client request is handled by the platform. Several experiments were run in the last two years to test the platform basic functionality, simultaneously with the analysis and processing of

raw data.

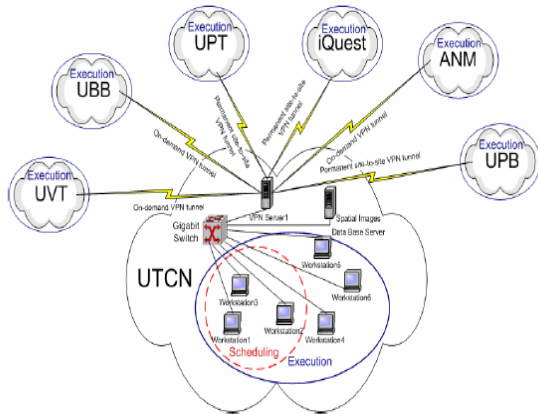


Fig. 1 – MedioGrid’s environment [5].

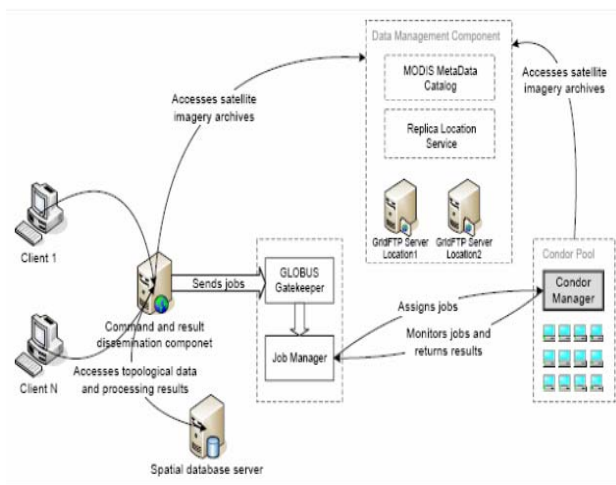


Fig. 2. – MedioGrid’s functionality [19].

The platform kernel contains the basic software components and services [11] which are used by most of the environment oriented Grid applications:

1. MedioGridService provides services for creating, executing and scheduling jobs;
2. MedioGridFactory is a stateful Web service which creates entities that manage job information;
3. MedioGridResource contains information about the job state, the starting and the ending time of the job;
4. MedioGridOGSADAI provides access to the image database;
5. MedioGridUrlCopy supports file transfer using the GridFTP protocol – by using this service the worker node accesses the image files and processes the data;
6. MedioGridRLS returns a list of physical images in order to allow the worker to decide, based on its location, which is the most appropriate GridFTP server for downloading the required data.
7. Condor is used to schedule, manage and monitor every job created by the MedioGridService.

The platform kernel has been empowered with software components implementing fundamental algorithms for image segmentation, as well as applying parallel and distributed data processing. Currently, MedioGrid components are performing basic image processing of MODIS and Landsat images, or more sophisticated operations like vegetation classification, water detection, cloud detection, or unsupervised image classification. Modeling and visualization of the virtual geographical space, GIS and LBS (Location Based Services) were also considered. Grid services were developed to publicly expose into Web interfaces the basic operations that can be performed on satellite images by executing general or specialized legacy codes. Parallel processing has been used as an effective solution in the case of time-consuming algorithms like the classification ones. Figures 3, 4 and 5 present some snapshots of the main components.

3. DETAILS AND PARTICULARITIES OF THE PLATFORM

In the following we discuss some details about the data management, special applications, extension of the algorithms for Grid implementation, and user interfacing through workflows.

3.1. DATA DISTRIBUTION

One research direction of the project focused on the data model and execution environment.

The main objective what concerns the data model was to define a proper data access interface and an efficient replication scheme. Data access patterns have been analyzed both statically and by using monitoring techniques and tools [28]. An open source solution for the Grid environment, based on the Debian Linux distribution and community provided tools, has been chosen. Data distribution and a generic data model based on OGSADAI interfaces have been proposed in [29] where basic computation orchestrations possibilities have been presented and analyzed.

Next step was to address the data distribution problem together with data life cycle. The initial proposal was to make use of a static replication scheme. Such a scheme wastes storage, especially for seldom used data. A dynamic policy based replication scheme was proposed, giving the possibility to adapt replicated application data distribution based on usage patterns [30]. The basic idea is that each of the grid nodes that are able to host data is advertising their available quota. At the top level, the application is making use transparently of the Grid Data Access layer which provides the interface towards the data storage. The organization of the system is depicted in Figure 6.

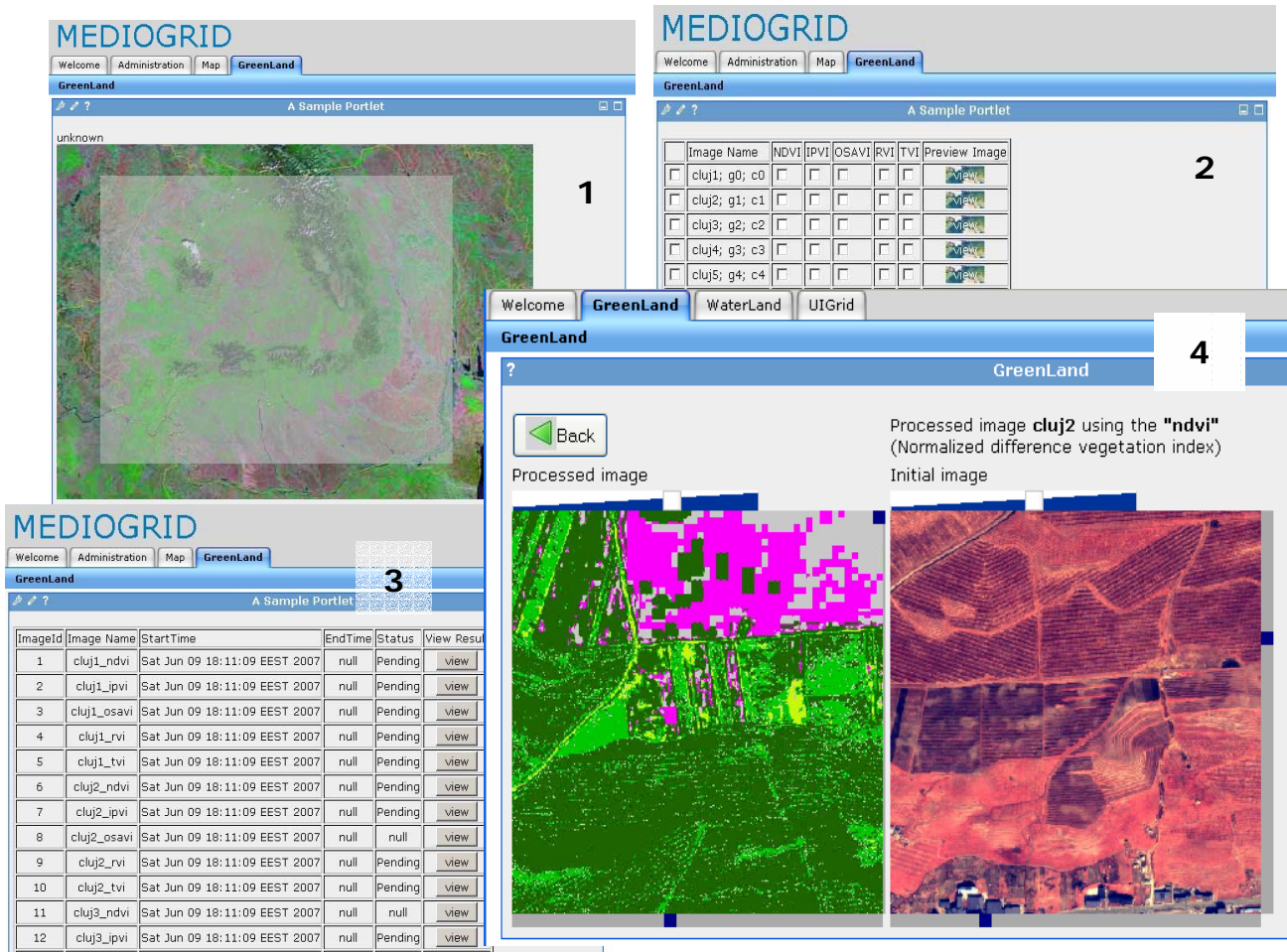


Fig. 3. – Greenland application – vegetation indices based satellite image classification.

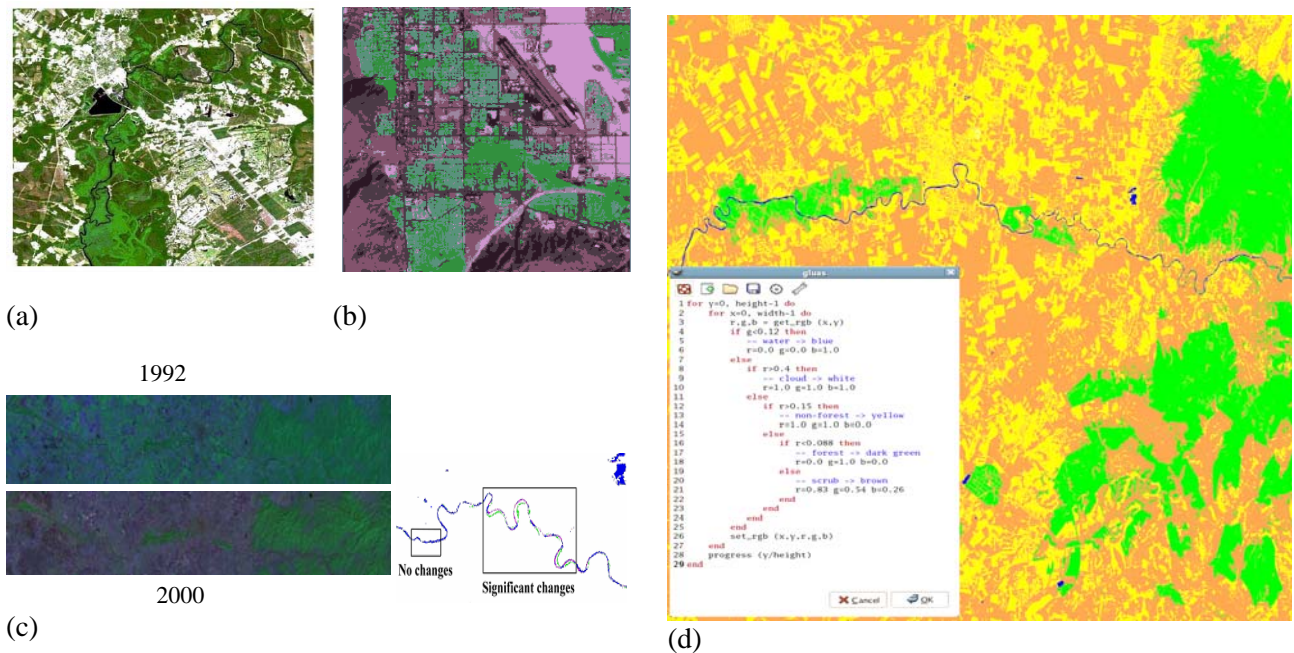
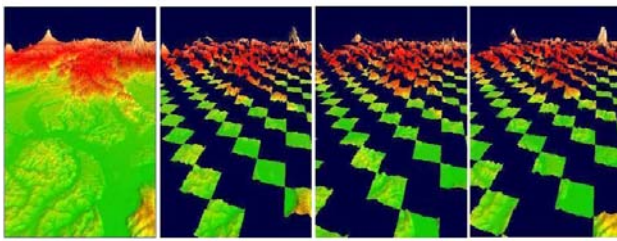
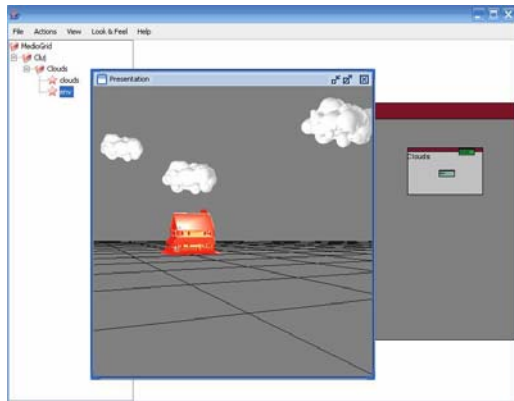


Fig. 4. – MedioGrid's image results: (a) Vegetation indices [2,18]; (b) Unsupervised segmentation by K-means; (c) Detection of changes in river bad [20]; (d) Decision tree-based classification by a GIMP service [21].



(a)



(b)

Fig. 5. – Virtual geographical space modeling: (a) Parallel terrain rendering [25]; (b) Clouds modeling [4].

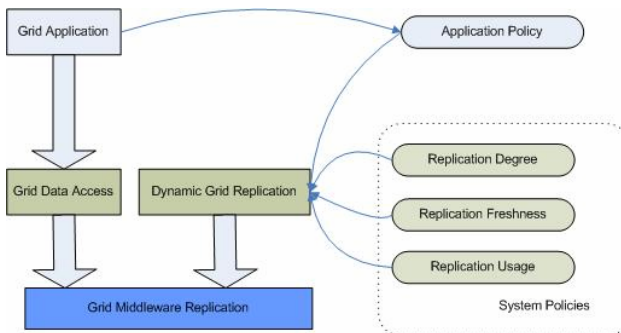


Fig. 6. – Replication Policies.

At the higher level, the system specifies only a collection of replication policies which are taken care of by the policy replication component. The dynamic policy based grid replication framework offers different kinds of replication policies for certain data. Its architecture is depicted in Figure 7. We considered to support application defined replication policies to offer more flexibility towards grid data life cycle customization. The build-in replication policies that we supply are based on data replication ratio, data freshness criteria and data usage. The replication degree policy aims to ensure that critical grid data is available throughout the entire life cycle of a grid application. The replication freshness degree regulates the replication of newly produced data and assures better data distribution in terms of data age. Last but not least, replication usage degree is based on the credit concept and ensures that frequently used data items are replicated with a higher

dynamic rate during their life cycle.

Research activities towards the execution environment are ongoing and they aim to address the problem of providing uniform and proper tools to ease the development and deployment of MedioGrid applications. We have started to analyze both performance and usability of existing Grid programming tools and aim to provide alternative paradigms based on shared data models.

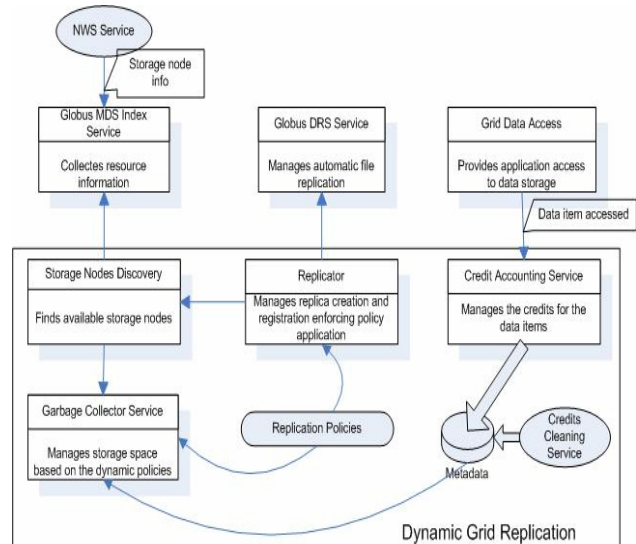


Fig. 7. – Dynamic policy based replication architecture [30].

3.2. SPECIALIZED APPLICATIONS

One distinct application of MedioGrid is Greenland [2]. It studies the extent and structure of the vegetation cover for a certain geographic area. Great attention is paid to the difference in reflectance between the visible and the near infrared domains. Each geographic area has a specific spectral signature, depending on the structure of vegetation that is expressed by the wavelengths emitted or reflected by that surface. The Greenland application classifies the vegetation areas based on vegetation indices. In the first phase the user specifies the time interval and the geographical area the satellite images address to, and the vegetation based classification operation. The system displays a list of images selected from the database according with the previous criteria. Next, the user can choose for classification a subset of them and the corresponding vegetation indices (Figure 3). The user may visualize both initial and classified images, and the status of the on going processing of the jobs. A simpler classification service based on a decision tree was reported also in [21].

Another distinct application is Waterland [3]. It performs the water detection by using short-wave infrared, red and near infrared spectral bands of

Landsat images. It returns a GeoTIFF image that highlights the water boundaries. A preliminary test service for detecting changes in river beds, based on wrapping an open-source image processing tool, was reported in [20] (Figure 4c).

Radio based satellite image processing can also reveal soil composition and the existing of the minerals. The recent developed Minerals application describes a methodology for satellite image based mineral deposit detection and experiments the detection of the hydrothermally-altered areas, iron-oxides and hydroxylbearings minerals in Romanian mountains.

3.3. ALGORITHM EXTENSIONS

A distributed algorithm for multispectral satellite image segmentation was recently developed and presented in [24]. Multispectral image segmentation is used in remote sensing for land cover and land use classification and change detection. Regions of the image are clustered separately and then the results are combined (Figure 8).

The algorithm employs two types of clustering techniques, each specialized to its task and steered towards obtaining a final meaningful segmentation. The results show good spatial coherency in segments and coherent borders between regions that were segmented separately. The algorithm seeks to find more spectrally and spatially coherent clusters by employing spatial information and spectral knowledge, and seeks to distribute the computation to be able to overcome the temporal complexity of the algorithms used as well as the considerable memory requirements. Another approach for unsupervised image segmentation is based on distributed versions of K-means and fuzzy C-means algorithms – see a result in Figure 4b.

While MedioGrid platform is built on the latest version of standard Grid middleware that is implementing Web service concepts (Service Grid), tests were performed also on a classical Grid infrastructure (Computational Grid), SEE-Grid (South-Eastern European Grid infrastructure II), to study the potential of using a joint infrastructure. In this scenario, the satellite images and the client's codes should be available at the client side as well as some minimal facilities to access the computational Grid infrastructure. The Grid middleware allows the execution of client codes on client's data at remote computing nodes. The case study refers to the implementation of a simple classification algorithm based on the binary decision tree. The classification process involves translating the pixel values in a satellite image into meaningful categories. The inputs are two MODIS bands (the red and the infrared ones). The detected land-cover classes include: water, cloud, non-forest, forest, and scrub. Since the classification

algorithm is applied on the pixel level, the splitting of the computational effort into similar tasks acting on parts of the image is straightforward: the bands are split into equal sub-images. The code prepared for the computational Grid consists of three components: the splitter that takes the two bands and splits them into a number of sub-images; the classifier that receiving two images (pieces from the red and infrared bands) applies the binary decision tree and produces the sub-image storing for each pixel the color of the associated land cover class; and the composer that gathers the colored sub-images. Parameter studies were performed to detect the best choice of the number of tasks (sub-images) depending on the image's dimension. The test results reported in [22] encourage the further investigation of this approach.

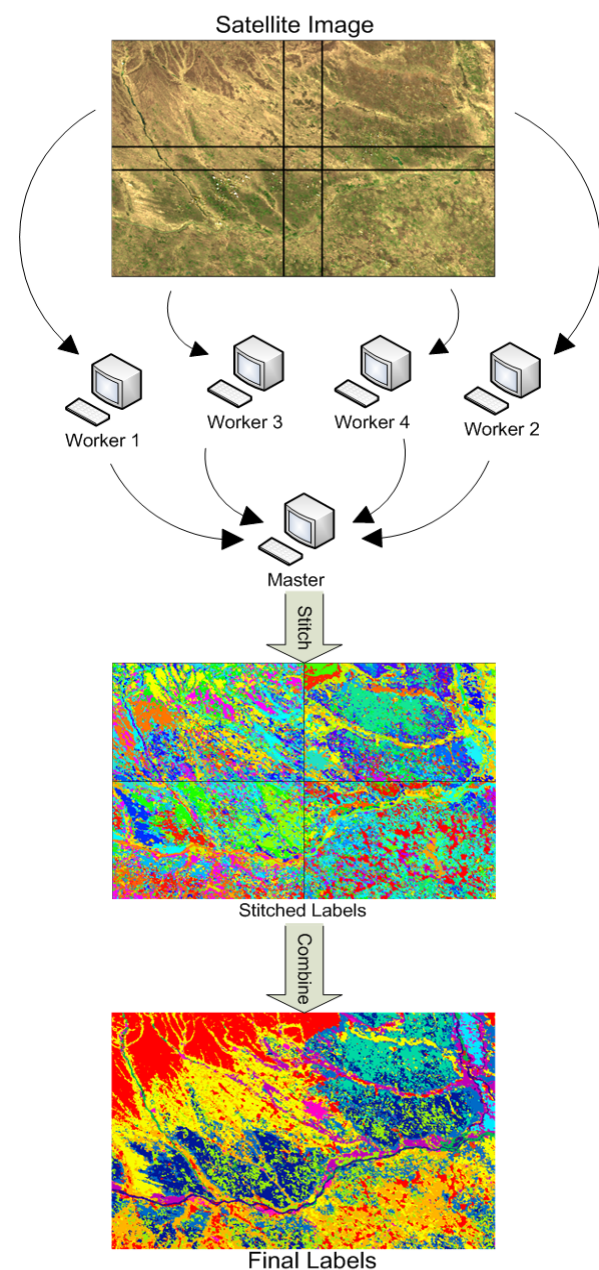


Fig. 8 – Multispectral image segmentation integrated into the platform [24]

3.4. PROCESS DESCRIPTION

Multiple experiments on Grid based satellite imagery classification require flexible descriptions of the processing workflow [14]. The flexible description of the processing workflows is accomplished in MedioGrid by the gProcess toolset by using Process Description Graphs (PDG), see Figure 9. Initially, the PDG description is a pattern concerning with the spectral bands of a specific satellite image type (e.g. MODIS, Landsat). To be scheduled and executed across the Grid, it must be mapped onto a physical satellite image (i.e. type, location, time). The PDG description is instantiated onto a specific satellite image and becomes an Instantiated Process Description Graph (IPDG). The research has concerned as well with experimental evaluation of the optimal mapping of the logical workflow onto the physical level. The execution time has been evaluated for various grouping techniques of the PDG's operational nodes.

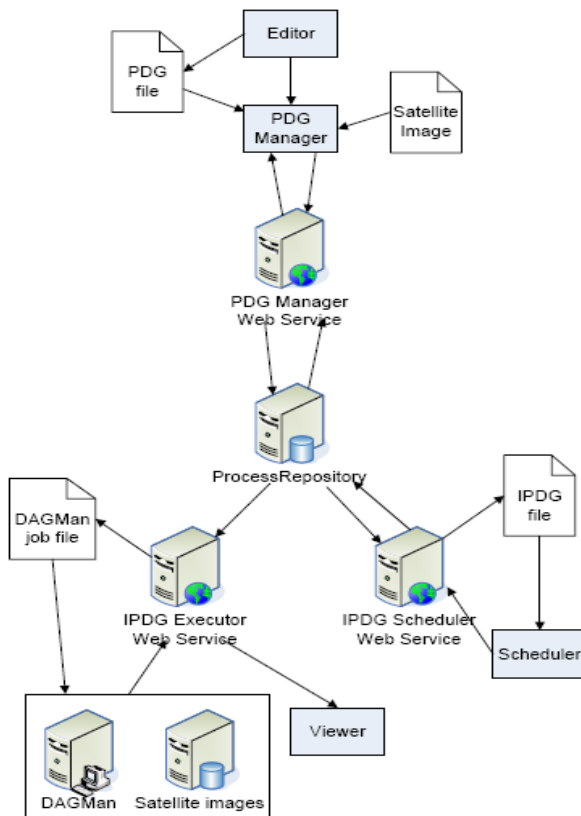


Fig. 9. – gProcess execution workflow.

4. RELATED WORK

Realizing the potential of the Grid computing for the satellite image processing, several projects were launched to make the Grid usage idea a reality. Within the DataGrid project funded by the European Union, an experiment aiming to demonstrate the use of Grid technology for remote sensing applications has been carried out; the results can be found for ex-

ample in the paper [10]. The EuroGrid project's Meteorological Portal, Meteo-GRID [16] developed a solution which allows anyone to run a numerical weather prediction model on demand. Several other international Grid projects were focused on spatial information, like SpaceGrid [26], Earth Observation Grid [15], or Genesis [34]. The MediGrid project [17] aimed to integrate and homogenize data and techniques for managing multiple natural hazards; its services are based on the latest Grid middleware. The authors of paper [1] present an overview of SARA Digital Puglia, a remote sensing environment that shows how Grid technologies and high performance computing can be efficiently used to build dynamic Earth observation systems for the management of space mission data and for their on-demand processing and delivering to final users. The recent G-POD initiative (Grid Processing on Demand [13]) aims to offer a Grid-based platform for remote processing the satellite images provided by European Space Agency. The European DEGREE project [9] delivered a study on the challenges that the Earth Sciences are imposing on Grid infrastructure, as well as several case studies in which Grid are useful. Furthermore the recently started European project D4Science [8] will study the data management of satellite images on Grid infrastructures.

The MedioGrid platform is trying to follow and apply the achievements of the above mentioned projects (based on international platforms) at a national scale and to respond to the needs of National Agency for Meteorology involved as partner in the project, as well as to the needs of common users interested in extracting environment characteristics from satellite images.

In what concerns the algorithms, we should mention that a frequent approach is to use the Grid as a high performance computing facility for CPU intensive operations. The paper [27] for example focuses on the parallelization of the computation-intensive satellite image geo-rectification problem on an Alice-based Grid. Classification is another important field in digital remote sensing image processing; large amount of remote sensing data classification will take a too long time. The aim of the proposed classification middleware on Grid from [31] is to divide jobs into several assignments and submit them to a computing pool.

MedioGrid's platform is encapsulating implementations of new parallel algorithms, e.g. parallel terrain rendering or multispectral segmentation.

Another approach undertaken by several initiatives is to wrap legacy codes as Grid services. For example, the parallel remote-sensing image processing software PRIPS (Parallel Remote-sensing Image Processing System) was encapsulated into a Grid service in [32]. The paper [6] outlines the design and

implementation of Grid-HSI, a service-oriented architecture-based Grid application to enable hyper-spectral imaging analysis. In the paper [33] is discussed the architecture of SIG computing environment (Spatial Information Grid project), based on Globus Toolkit, OpenPBS, and Condor-G; a model of the image division is proposed, which can compute the most appropriate image pieces and make the processing time short.

MedioGrid platform is service-oriented. It is based on the latest generation Grid middleware (Globus Toolkit 4) and uses most recent tools to deploy and access Grid services. It offers an original solution for data replication and distribution based on OGSA-DAI in order to ensure fault tolerance [7]. MedioGrid's platform is encapsulating also Grid-services wrapping legacy codes, e.g. GIMP. In a field dominated by proprietary sophisticated codes, public-domain software packages are used as basic components in order to make possible the public access to the platform services.

5. CONCLUSIONS

Current Grid technologies provide powerful tools for remote sensing data sharing and processing. Realizing this fact, a platform based on open-source software was built to serve the requests coming from domain experts as well as novice users. The MedioGrid platform was populated with several complex applications and distributed implementation of computational and data intensive algorithms that were shortly described in this paper. Further developments of the platform are related to the issue of real-time processing, the development of new applications, as well as the usage of the knowledge and experience accumulated by the partners in the frame of the follower projects, e.g. [12].

5. ACKNOWLEDGMENT

The research reported in this paper was partially supported by the Grant 19CEEX-I03 MedioGrid of the Romanian Ministry of Education and Research.

6. REFERENCES

- [1] G. Aloisio, M. Cafaro. A dynamic Earth observation system. *Parallel Computing* 29-10 (2003), p. 1357-1362.
- [2] V. Bacu, O. Muresan, D.Gorgan. MODIS image based computation of vegetation indices in MedioGRID architecture. *Procs. SYNASC06*, IEEE Computer Society (2006), p. 267-273.
- [3] V. Bacu, D. Gorgan. Resource measurements for water detection algorithm in MedioGrid architecture. *Procs. ISPDC07*, IEEE Computer Press (2007), p. 107-114.
- [4] R. Barbantan, D. Gorgan. Active objects based application over Grid environment, *Procs. SYNASC06*, IEEE Computer Society (2006), p. 289-295.
- [5] M.S. Boboila, G.V. Iordache, F. Pop. V. Cristea. A framework for scheduling image processing applications in MedioGRID, *Procs. SYNASC06*, IEEE Computer Society (2006), p. 303-308.
- [6] C.L. Carvajal-Jimenez, W. Lugo-Beauchamp, W. Rivera. Grid-HSI: using Grid computing to enable hyper-spectral imaging analysis. *Procs. 3th Intern. Conf. Communications, Internet & Information Tech.* (2004), p. 583-588.
- [7] A. Colesa, I. Ignat, R. Opris. Providing high data availability in MedioGRID. *Procs. SYNASC 06*, IEEE Computer Society (2006), p. 296-302.
- [8] D4Science (2008), <http://www.d4science.org/>.
- [9] DEGREE (2007), <http://www.eu-degree.eu/>.
- [10] N.A. Giovanni, F.B. Luigi, J. Linford. Grid technology for the storage and processing of remote sensing data: description of an application. *SPIE* 4881 (2003), p. 677-685.
- [11] D. Gorgan, O. Muresan, V. Bacu, C. Melenti, D. Safta. Satellite image processing by Medio-Grid platform kernel. *Procs. CSCS16*, vol. 2 (2007), p. 142-147.
- [12] GISHEO (2008), <http://gisheo.info.uvt.ro>.
- [13] G-POD (2007), <http://eogrid.esrin.esa.int/>.
- [14] A. Radu, V. Bacu, D. Gorgan. Diagrammatic description of satellite image processing work-flow. *Procs. SYNASC'07*, IEEE Computer Press (2007), p. 341-348.
- [15] Earth Observation Grid, <http://www.e-science.clrc.ac.uk/web/projects/earthobservation>
- [16] C.J. Lenz, D. Majewski. EUROGRID and MeteorGRID (2004). <http://www.eurogrid.org>.
- [17] MediGrid (2006), <http://www.eu-medi-grid.org/>.
- [18] C. Melenti, D. Safta, D. Gorgan. PIMS – multi-spectral image processing tool for semantic information detection based on vegetation indices. *Procs. SYNASC06*, IEEE Computer Society (2006), p. 279-288.
- [19] O. Muresan. F. Pop. D. Gorgan. V. Cristea. Satellite image processing applications in MedioGRID, *Procs. ISPDC06*, IEEE Computer Press (2006), p. 253-260.
- [20] D. Petcu, V. Jordan. Grid service based on GIMP for processing remote sensing images. *Procs. SYNASC06*, IEEE Computer Society (2006), p. 251-258.
- [21] D. Petcu. A solution for satellite image processing on Grids. *Procs. DNCOCO06* (2006), p. 75-80.
- [22] D. Petcu, S. Panica, A. Eckstein. Satellite image processing on computational Grids. *Procs. ACMOS 2007* (2007), p. 216-221.
- [23] D. Petcu, D. Zaharie, D. Gorgan, F. Pop, D. Tudor. MedioGrid: a Grid-based platform for satellite images. *Procs. IDAACS'2007*, IEEE Computer Press (2007), Paper no. i07-056.

- [24] F. Pop, C. Gruia, V. Cristea. Distributed algorithm for change detection in satellite images for Grid environments, *Procs. ISPDC07*, IEEE Computer Press (2007) p. 41-49.
- [25] N. Somosi, D. Petcu. A parallel algorithm for rendering huge terrain surfaces. *Procs. SYNASC 06*, IEEE Computer Society (2006), p. 274-278.
- [26] SpaceGrid, <http://www.spacegrid.org/>.
- [27] Y.M. Teo, S.C. Tay, J.P. Gozali. Distributed georectification of satellite images using Grid computing. *Procs. Int. Parallel and Distributed Processing Symposium*, IEEE Computer Society Press (2003), p. 152-157.
- [28] D. Tudor, F. Pop, V. Cristea, V. Cretu. Towards an IO intensive Grid application instrumentation in MedioGrid. *Procs. CSCS16* (2007), p.130-135.
- [29] D. Tudor, V. Cretu. A data model view on MedioGrid. *Procs. CSCS16* (2007), p. 150-154.
- [30] D. Tudor, G. Macariu, V. Cretu. Dynamic policy based replication on the Grid. *Procs. SACI07* (2007), p. 77-82.
- [31] J. Wang, X. Sun, Y. Xue, Y. Hu, Y. Luo, Y. Wang, S. Zhong, A. Zhang, J. Tang, G. Cai. Preliminary study on unsupervised classification of remotely sensed images on the Grid, *LNCS 3039* (2004), p. 981-988.
- [32] X.J. Yang, Z.M. Chang, H. Zhou, X. Qu, C.J. Li. Services for parallel remote-sensing image processing based on computational Grid, *LNCS 3252* (2004), p. 689-696.
- [33] C. Yang, D. Guo, Y. Ren, X. Luo, J. Men. The architecture of SIG computing environment and its application to image processing, *LNCS 3795* (2005), p. 566-572.
- [34] T. Yunck, B. Wilson, A. Braverman, E. Dobinson, E. Fetzer. GENESIS: The general Earth science investigation suite. *Procs. ESTC04* (2004).



Grid computing.

Dana Petcu, Holds a MSc in Computer Science and a PhD in numerical analysis. She is Professor and Director of Computer Science Department of Western University of Timisoara, as well as Researcher and Director of Institute e-Austria, Timisoara. Her current topics of interests are parallel, distributed and



Dorian Gorgan, is professor and PhD supervisor in Computer Science at the Technical University of Cluj-Napoca. He is the coordinator of the MedioGrid project, and chair of the international GridCAD workshop. His work concerns with computer graphics, distributed interactive applications, Grid computing, grid visualization, virtual geographical space modelling, and semantic Grid and Web services.



Florin Pop, Hold a MSc in Computer Science in 2006 at Computer Science department in University Politehnica of Bucharest. Now he is PhD student and his main work refer to scheduling in Grid environments, distributed systems and parallel computation, communication protocols, numerical methods. He is member of RoGrid consortium and developer in the national projects (e.g. GridMOSI, MedioGRID) and international projects (SEE-GRID, EU-NCIT).



Dacian Tudor, Holds a MSc in Computer Science at the "Politehnica" University of Timisoara, Computer Science Department. Currently he is a PhD student and focuses on creating a grid service layer for shared data programming tailored for wide scale distributed systems. His research topics include distributed and Grid computing as well as mobile communication systems.



Daniela Zaharie, Holds a MSc in Computer Science and a PhD in Probability and Statistics. She is associate professor at the Department of Computer Science from the West University of Timisoara. Her current research interests include evolutionary computing, data mining and image processing.