The Eighteenth International Symposium on Artificial Life and Robotics 2013 (AROB 18th '13),
Daejeon Convention Center, Daejeon, Korea, January 30-February 1, 2013

Study on the Efficient Use of Satellite Image Data Analysis System

Hayao Mori[1], Seiji Tsukamoto[1],
Takashi Yamaguchi[2], Masaki Hanada[2], Jong Geol Park[2], Eiji Nunohiro[2],

[1] Graduate School of Tokyo University of Information Sciences, Japan
[2] Tokyo University of Information Sciences, Japan
(Tel: 81-43-236-1292, Fax: 81-43-236-1292)

[1]{g10022hm, g12008st}@edu.tuis.ac.jp, [2]{tyamagu, mhanada, amon, nunohiro}@rsch.tuis.ac.jp

Abstract: Tokyo University of Information Sciences receives MODIS (Moderate Resolution Imaging Spectroradiometer) data, one of the sensors equipped by NASA's Terra and Aqua satellites, and researches of the analysis on change of environment as part of the academic frontier project. For the information infrastructure of this frontier research, we are developing a satellite image data analysis system to support of web system, a parallel distributed system configuration using multiple PC clusters, database for MODIS data to open the research results and MODIS data for public use. This paper presents the overview of satellite data analysis system and new feature which are the composite of multiple satellite data, and LSM with virtualization of storage system for satellite data, scheduling for multiple cluster nodes and performance evaluation.

Keywords: Scheduling, Virtualization, MODIS

1 INTRODUCTION

Tokyo University of Information Sciences (TUIS) receives MODIS (Moderate Resolution Imaging Spectroradiometer) data from one of the sensors equipped by NASA's Terra and Aqua satellites, and provides processed data to universities and research institutes as a part of academic frontier project organized by TUIS. The main research theme of this project is on "the sustainable development of economic and social structure dependent on the environment in Eastern Asia". One of the major fields of the research using MODIS data is to analyze environmental change. In the frontier project, we are developing satellite image data analysis system (SIDAS) to support the frontier research such as to forecast weather change and environmental changes as well as processing and maintaining the satellite data.

2 SATELLITE IMAGE DATA ANALYSIS SYS TEM (SIDAS)

2.1. Overview

This system works as the main platform of the total satellite image data analysis service, maintains the satellite image data, and implements satellite image data analysis applications. The user can access the system over the network via Web browser or other web applications. Users can request for data analysis, check the results and download satellite image data through the system.

The satellite image data is extremely large, and direct manipulation of the large data leads to a high load on both the processor and network. In order to improve the throughput and turnaround time of data processing, in the proposed system the computer nodes are configured from high performance servers and multiple PC clusters, implemented as a parallel distributed system. The various analysis applications and the preprocessing of satellite image data are run on this distributed computer nodes.

2.2. System configuration

SIDAS consists of a web server, an application server consisting of multiple PC clusters, and a database server. Figure 1 shows the overview of SIDAS. The application server schedules the execution of applications, manages the satellite data, and controls PC clusters which execute applications. The user can access the system, request data analysis and check the results over the network via a Web browser or other web applications, and the user can also download satellite image data through the system.

The Web server accepts request of the user and submits processing of the request to the application server. The application server transmits the MODIS data used by the processing from the MODIS database to the local disk in the application server. The application server executes application for the request and returns the results to Web server.

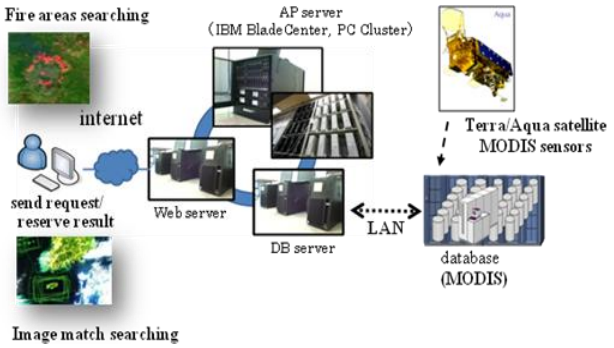

Fig. 1. Overview of SIDAS

3 SATELLITE IMAGE DATA ANALYSIS SYS TEM (SIDAS)

3.1. MODIS data using

MODIS is a multiple band radiometric sensor instrument aboard Terra and Aqua satellites to monitor the environment of a large region of the earth covering a width of 2,330 km. It measures 36 spectral bands between 0.405 and 14.385 μm with three spatial resolutions (250m, 500m, 1,000m). Table 1 shows the band information of MODIS data.

Tokyo University of Information Sciences (TUIS) is receiving MODIS data from three places in Hokkaido, Tokyo and Okinawa, and received the data 3 or 4 times in the daytime and 3 or 4 times in the nighttime from each place. The observing areas from the three places are from 80 degrees to 180 degrees of east longitude and from 0 degrees to 80 degrees of north latitude.

- MOD02 (spatial resolution: 500m) is used to create visualized data from satellite data and to calculate NDVI, and we use 7 spectral bands between visible range and short wave infrared range. BAND1 (red in visible region), BAND4 (green in visible region) and BAND3 (blue in visible region) are used to create the visual information from the satellite data.BAND2 (near-infrared region) are mainly used to measure the vegetation index (The correlation is high in the photosynthesis activity).BAND5, 6, 7 (short wave infrared) has the feature of temperature and reflection factors.
- MOD03 (spatial resolution: 1KM) has geological information containing geodetic latitude, longitude, surface height above sea level, solar zenith, solar azimuth angles, satellite zenith, satellite azimuth angles, and a land/sea mask. Mod03 is used to make a map information from satellite data (MOD02, MOD09, MOD11, MOD28 and MOD35) according to map projection.
- MOD11 (spatial resolution: 1KM) is used to estimate the land surface temperature by using reflectance of land cover. MOD11 is effective data for the heat island phenomenon analysis of urban areas, soil moisture conditions analysis and the heat flux analysis.
- MOD13 (spatial resolution: 500m) is vegetation index data after atmospheric correction.
- MOD14 (spatial resolution: 1KM) is thermal Anomalies data includes fire occurrence (day/night).
- MOD28 (spatial resolution: 1KM) is Sea Surface Temperature data. This data is mainly used to analyse the change of sea surface temperature and the flow of sea.

Table 1. Band information

BAND	WAVELENGTH (nm)	SPATIAL RESOLUTIONS	KEY USE
1	620 ~ 670	250m	Absolute Land Cover change, Vegetation Chlorophyll
2	841 ~ 876	250m	Cloud Amount, Vegetation Land Cover change
3	459 ~ 479	500m	Soil/ Vegetation Differences
4	545 ~ 565	500m	Green Vegetation
5	1230 ~ 1250	500m	Leaf/Canopy Differences
6	1628 ~ 1652	500m	Snow/Cloud Differences
7	2105 ~ 2155	500m	Cloud Properties, Land Properties

3.2. Data format

MODIS data is converted into Raw data for every band by the database creation processing, so that it may be easy to use public and processing of the various analysis.

Raw data is a data format near a general graphics format. In the data part of Raw data, photography data at equal intervals are stored continuously.

Additional information such as position information and time of MODIS data are stored in another file (header file).

3.3. Preprocessing of satellite image data

3.3.1 Mosaic processing

Figure 2 shows the example of mosaic processing. The mosaic process aggregates the converted Raw data sets into one data set per one orbit. The aggregated data set is generated from three data sets respectively received from the three stations and is maintained consistency because the three received data sets are the same data at the same point.

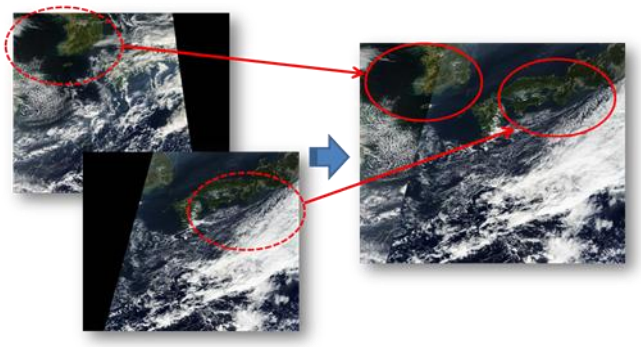

Fig. 2. Example of mosaic processing

3.3.2 Composite processing

When analyzing large area satellite data, one common problem is noise caused by cloud interference. In order to remove the interference of clouds, the standard solution is to create a composite data of the same regions during a selected time span, and to patch together data not covered by clouds to create a clear image.

3.4. Example of the use of MODIS data

One of the purposes of SIDAS is to support researches which use the MODIS data such as environmental perturbation analysis and earth surface temperature analysis.

Main researches under development are as follows.

1) To search for fire regions in forests and fields.

2) To search for a set of data, from the satellite data in TUIS, which is similar image data analyzed by the system.

3) To provide the spatiotemporal analysis of land cover changes of East Asia.

4) To create the disaster map of East Asian.

5) To forecast NDVI by using auto-regression analysis.

6) To provide statistical and soft computing analysis for land cover estimation.

4 EXECUTION CONTROL OF APPLICATION

The analysis processing which uses MODIS data uses large-scale satellite data.

In order to process at the high speed of analysis processing, efficient employment of a system resource is required.

For this reason, we developed LSM (Logical Storage Manager) which manages the execution status of PC cluster and satellite data stored in each cluster.

LSM includes the scheduling function of PC cluster, and the reuse function of satellite imagery data, and determines the PC cluster configuration dynamically which AP (application program) uses.

4.1. Scheduling function of PC cluster

The purpose of the scheduling function of PC cluster is to reduce the execution time of APs by effective use of resources such as CPU and local disk of each PC cluster node. Figure 3 shows an example of the scheduling function. LSM manages the system status such as the operation status of all PC cluster nodes and status of the data stored on the local disk of each PC cluster node. LSM dynamically determines the configuration of the PC cluster nodes (i.e., the number and location of PC cluster nodes) to execute the AP by considering both the operation status and the data which the AP uses. Dynamic resource allocation can be performed with dynamically combining the AP and the data which the AP uses. The procedure of the scheduling function is shown as follows.

① The AP requires the number of PC cluster nodes and the data which the AP uses.
② From ①, the assigned number of PC cluster nodes is determined.
③ LSM determines the configuration of the PC cluster nodes to execute the AP with comparing the data stored on the local disk of each PC cluster node to the data which the AP uses.
④ The AP and the data which the AP uses are combined.

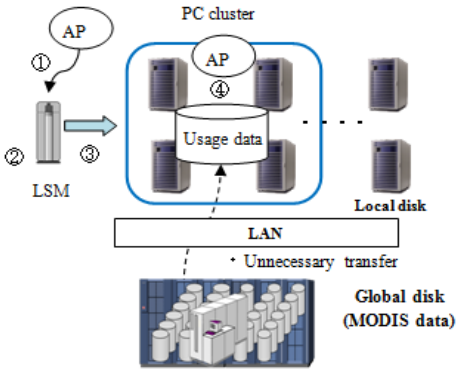

Fig. 3. PC cluster scheduling

4.2 Data reuse function

The purpose of the data reuse function is to reduce an amount of transferred data between PC cluster node and the global disk by reuse of the data stored on the local disk. Figure 4 shows an example of data reuse. The procedure of the data reuse function is shown as follows.

① LSM assigns the AP to the PC cluster node in which the data which it uses is stored.
② When the data which the AP uses is stored on the local disk of a running PC cluster node, the data is transferred from the running PC cluster node to a PC cluster node to execute the AP.
③ When the data which the AP uses is not stored on the local disk of any PC cluster nodes, the data is transferred from the global disk to a PC cluster node to execute the AP.

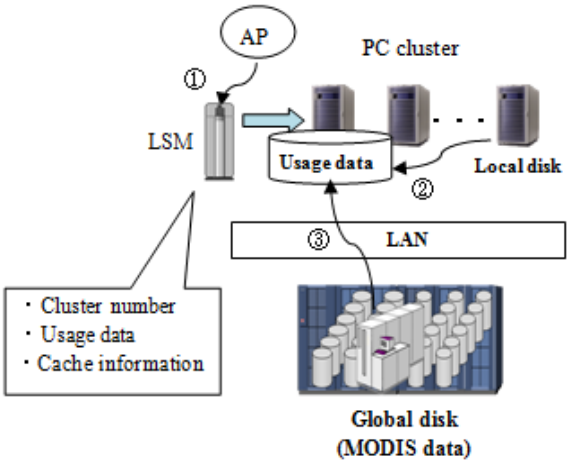

Fig. 4. Data reuse

5 PERFORMANCE EVALUATION

5.1 Evaluation method

The following conditions was used for evaluation of the data reuse function.

- Assess changes reusability of satellite image data (hit rate)
- Multiple tasks are register to the queue of LSM, and are performed in order.
- Number of clusters and the data used for a single task is determined by the random number
  - ➢ The following types random number was used
    - ✧ Uniform random number
    - ✧ Normal random number
- Evaluation condition
  - ➢ Number of PC cluster : 32 units
  - ➢ local data retention limit of PC cluster : 4
  - ➢ Number of data in the database : 128 RAW data
  - ➢ the number of tasks to be performed on PC cluster : 200

5.2 Evaluation result

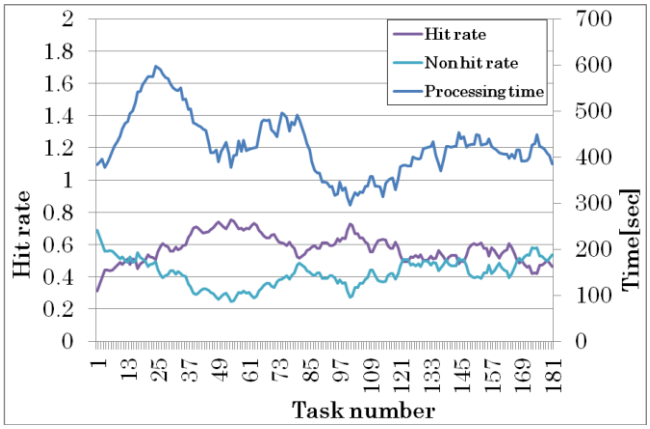

Fig. 5. Hit rate versus time

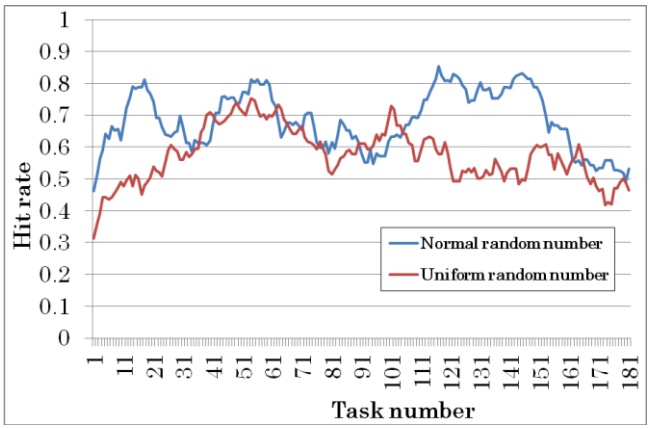

Fig. 6. Comparison of hit rate

Figure 5 and 6 show the results of the evaluation of the proposed functions. In the evaluation, 200 tasks are generated and the proposed functions are evaluated in terms of hit rate. Here hit rate is a rate of the number of data found on the local disk to the number of the generated tasks. The processing time is reduced as the hit rate increases. Figure 5 shows the hit rate and task processing time. The hit rate under normal distribution is higher than uniform distribution. Figure 6 shows the hit rate under normal distribution and uniform distribution.

5.3 CONCLUSIONS

In the satellite image data analysis system, we have proposed and implemented LSM to improve the processing performance and the data reusability. We measured the hit rate of the data that is stored on the local disk of PC clusters to evaluate the scheduling function and the data reuse function of LSM. From the evaluation results, we confirmed that the processing time was reduced as the hit rate increases.

In the future, we aim the efficient use of the system resources by more improving the scheduling function and the data reuse function.

Acknowledgments

This research was supported by the Research project of Tokyo University of Information Sciences for the sustainable development of economic and social structure dependent on the environment in eastern Asia.

REFERENCES

[1] Akihiro Nakamura, Jong Geol Park, Kenneth J. Mackin, Eiji Nunohiro, et al, Development and Evaluation of Satellite Image Data Analysis Infrastructure, The Sixteenth International Symposium on Artificial Life and Robotics, Proceeding Index OS18-3, (AROB 16th' 2011)

[2] Eiji Nunohiro, Jong Geol Park, Kenneth J. Mackin, et al, Development of satellite data analysis system, Proceedings of the 5th International Conference on Information (2009), pp.143-146

[3] Eiji Nunohiro, Kei Katayama, Kenneth J. Mackin, Jong Geol Park, Forest and Field Fire Search System using MODIS Data, Journal of Advanced Computational Intelligence and Intelligent Informatics (JAC 3, 2007), vol.11 No.8, pp.1043-1048

[4] Kenneth J. Mackin, Eiji Nunohiro, Development of Satellite Image Searching using Distributed Genetic Algorithm with Normalized Correlation, The twelfth International Symposium on Artificial Life and Robotics, Proceeding Index GS12-7, (AROB 12th' 2007)

[5] National Aeronautics and Space Administration (NASA), MODIS Web, http://modis.gsfc.nasa.gov/