Development of Satellite Image Searching using Distributed Genetic Algorithm with Normalized Correlation

Kotaro Matsushita[1], Kei Katayama[2], Kenneth J.Mackin[1], Eiji Nunohiro[1]
[1]Tokyo University of Information Sciences
[2]Gracuate School of Tokyo University of Information Sciences
1200-2 Yatoh-cho, Wakaba-ku, Chiba 265-8501, Japan
{matusita, mackin, nunohiro}@rsch.tuis.ac.jp, g05004kk@edu.tuis.ac.jp

Abstract

Tokyo University of Information Sciences maintains and distributes MODIS (Moderate Resolution Imaging Spectroradiometer) satellite data as part of the research output for Frontier project. An intelligent image search system is being developed as part of the project, in order to retrieve requested images such as matching images patterns or forest and field fires extraction. The intelligent image search system applies GA (Genetic Algorithm) in the search algorithm. When searching for a target image area within the MODIS image database, it is possible that the search algorithm cannot match the optimal location when the brightness of the search image data and MODIS data image are very different. In order to solve this problem, we applied normalized correction to the GA fitness function to improve the matching accuracy. Further, we implemented the image search as distributed genetic algorithm search over a PC cluster network, in order to increase the search speed within the satellite image database. We tested the proposed system and verified the effectiveness of distributed genetic algorithm for the distributed MODIS satellite database search process.

Keyword: Satellite image data, Image search system, Brightness, Distributed genetic algorithm, nornalized correction

1 Introduction

Tokyo University of Information Sciences receives MODIS image data from NASA satellites Terra and Aqua, and distributes collected data to universities and research institutions. MODIS sensors are the key sensors aboard the Terra and Aqua satellites. Moderate resolution remote sensing allows the quantifying of land surface type and extent, which can be used to monitor changes in land cover and land use for extended periods of time.

In order to effectively utilize the large scale satellite image database, an efficient search algorithm allowing quick retrieval of the required information from the large database is required. An intelligent image search system is being developed as part of the research output for the Frontier project, in order to retrieve requested images such as matching images patterns or forest and field fires extraction. The intelligent image search system applies GA in the search algorithm.

In our previous work, we have proposed applying GA to match partial images from MODIS image data [2]. But when searching for a target image area within the MODIS image database, it is possible that the search algorithm cannot match the optimal location when the brightness of the search image and MODIS data image are very different due to differences in time of day of the satellite image retrieval. In order to solve this problem, we applied normalized correction, which is effective in matching images with different relative brightness, to the GA fitness function to improve the matching accuracy. Further, we implemented the image search as distributed genetic algorithm search over a PC cluster network, in order to increase the search speed within the satellite image database.

We applied our proposed method in our MODIS data image matching system, and verified that the proposed method improved matching accuracy for images with different relative brightness.

2 Outline of Image Match Search Processing

We develop an image search system which retrieves matching images with similar features and characteristics to the search image, from the MODIS image database. When searching for a matching image area within the MODIS image database, it is possible that the search algorithm cannot match the optimal location for the following reasons.

(1) The MODIS image data for the correct location of the search may be masked by clouds, or changed from land surface alterations.

(2) The range of brightness for the requested search image and compared MODIS image data may be different due to difference in data retrieval time of the satellite data.

For the solution to problem (1), we proposed using multiple time-series data for the same search space, and applying GA within the time-series data to search for the requested image. For the solution to problem (2), we propose applying normalized correction, which is effective in matching images with different relative brightness, to the GA fitness function to improve the matching accuracy. We describe the basic flow of the proposed search process below. Figure 1 shows the image match search system outline.

(1) At the server, the requested search image, search area, and search time span is specified.

(2) The server distributes the search image, search area, and search time span for each individual client machine.

(3) The requested search image and MODIS search space image data are converted from color images to grayscale images by each client. Each client applies distributed genetic algorithm to search for the requested image pattern among the distributed search space. Brightness of the gray-scale images is adjusted by the distributed genetic algorithm.

(4) Clients which finished the search relays the search result to other clients.

(5) When the search location is determined, each client analyzes the image data for the same location, and evaluates the land cover change for the time span.

3 Image Search Process using Distributed Genetic Algorithm

Recently, GA has attracted much attention as an effective method for solving large scale complex problems [1]. GA searches for the optimum solution by applying genetic operations of 'mutation', 'crossover', 'evaluation' and 'selection' to the population. In the distributed GA model, the main population is divided into several subsets, and genetic operations are repeated among the distributed sub-population.

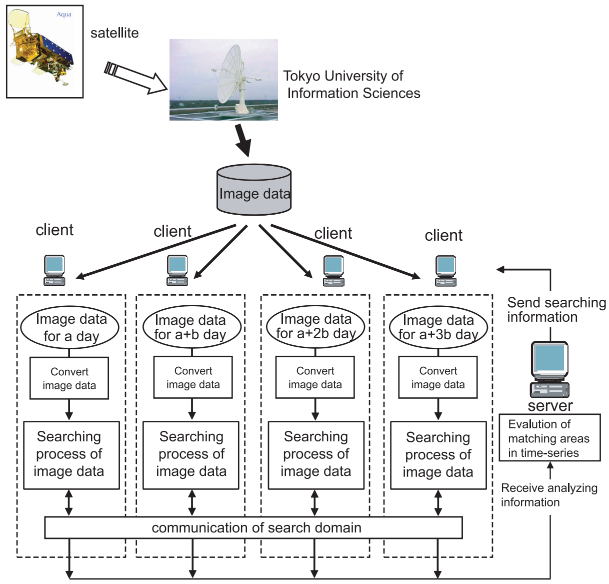

Figure 1: Outline of the image match search system

3.1 Distibuted Genetic Algorithm

DGA (Distributed Genetic Algorithm) is a parallel model of GA. In DGA, the main population is divided into sub-populations. DGA is also called the 'island model' from to this feature.

The divided sub-population is distributed into several clusters for parallel processing, and each cluster executes the GA search on the sub-population received. At each predefined interval or number of generations, each cluster exchanges individual solutions or chromosomes. This is defined as a 'migration' operation.

For this research, time-series image data are distributed to PC clusters, and in each cluster the GA search is processed in parallel. The parallel search process is completed when any one of the clusters finds a strong match to the requested image pattern.

Figure 2 shows the outline of DGA model, executed that preprocessing brightness confronts.

3.2 Brightness Adjustment Method

MODIS data is the value of the strength of reflected wavelengths for specified bands. This means that the value of MODIS data (i.e. brightness) for

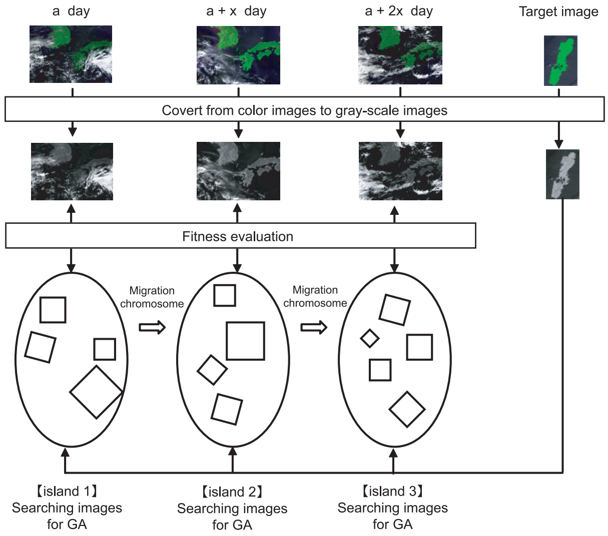

Figure 2: Outline of DGA model

the same location vary depending on the time of day. When searching for a matching image area within the MODIS image database, it is possible that the search algorithm cannot match the optimal location when the brightness of the search image and MODIS data image are very different.

MODIS data is the value of the strength of reflected wavelengths for specified bands. This means that the value of MODIS data (i.e. brightness) for the same location vary depending on the time of day. When searching for a matching image area within the MODIS image database, it is possible that the search algorithm cannot match the optimal location when the brightness of the search image and MODIS data image are very different. The pattern matching algorithm, such as GA, must be able to match similar location images with different brightness. In this research, we applied normalized correction, which is effective in matching images with different relative brightness, to the GA fitness function to improve the matching accuracy. Details of the fitness function is described in section 3.6. In order to shorten the computation time to acceptable limits for actual use, we first convert the RGB color image data for the search image to grayscale images using the following equation [3].

$$p = 0.299 * R + 0.587 * G + 0.144 * B \qquad (1)$$

where p is the image pixel of the gray-scale image, R,G,B are the respective red, green, and blue pixel data of the color image at the same pixel location.

x	y	rate	angle

Figure 3: Contents of the chromosome

3.3 Chromosome Expression

The image matching problem is defined as an optimization problem of finding the highest matching rate image by modifying the 4 parameters: 1)x-axis location, 2)y-axis location for the center of the matching image, 3)the magnification rate, and 4)the rotation angle of the matching image. The GA chromosome is designed to encode the 4 parameters, and the fitness is defined as the matching rate of the compared image pixels. Figure 3 shows contents of the chromosome.

3.4 Selection, Crossover, Mutation Operations

In this paper, we evaluated the following genetic operation methods.

(a) Selection Operation
The elite selection strategy was applied, in which the individuals with the highest fitness will be selected for the next generation.

(b) Crossover Operation
For the selection operation, a single locus method is applied, in which a random locus or break point is selected and two parents chromosomes swap one side of the chromosome up to the break point. This creates two child chromosomes of the same length combining information of both parents.

(c) Mutation Operation
For the mutation operation, 2 points in a single chromosome are selected at random, and the bit information between these 2 points are replaced in reverse order.

3.5 Distributed Genetic Algorithm Procedure

In this research, DGA is applied to the image match search system. The procedure for each client is the following.

(1) Receive the search image, search area and search time span, and read the specified date and location MODIS data from the MODIS database.

(2) Convert the color search image and color MODIS image data to gray-scale images.

(3) Create initial population
Create an initial population of chromosomes with random parameters, with coordinates (x,y) for center of search region, magnification m, and rotation r.

(4) Calculate the fitness of the chromosomes.

(5) Evaluate the search completion condition.

(6) If the search completion condition is met, notify the result to other clients.

(7) If the completion condition is not met, continue with (8).

(8) Selection
Select individuals based on fitness, and create m pairs randomly.

(9) Crossover
Create new child chromosomes from the selected pairs as parents, applying the specified crossover operation strategy.

(10) Mutation
Applying the specified mutation strategy, randomly modify a part of the selected chromosome.

(11) Migration
At each specified number of generations, migrate selected chromosomes according to the specified migration strategy.

(12) repeat from (3).

3.6 Fitness Evaluation

The fitness is evaluated by calculating the matching rate of the selected image pixel data that is specified by the search parameters encoded in the chromosome. The fitness function (C) is shown in equation (2), where the compared image size is M(pixel)*N(pixel), the brightness value (color value) for the template (MODIS) image and compared search image at coordinates(x,y) are T(x,y) and I(x,y), respectively.

$$C = \frac{\sum_{x=1}^{N}\sum_{y=1}^{M}\{I(x,y)-\mu_I\}\cdot\{T(x,y)-\mu_T\}}{\sqrt{\sum_{x=1}^{N}\sum_{y=1}^{M}\{I(x,y)-\mu_I\}^2\cdot\sum_{x=1}^{N}\sum_{y=1}^{M}\{T(x,y)-\mu_T\}^2}} \tag{2}$$

where

$$\mu_I = \frac{1}{NM}\sum_{x=1}^{N}\sum_{y=1}^{M} I(x,y) \tag{3}$$

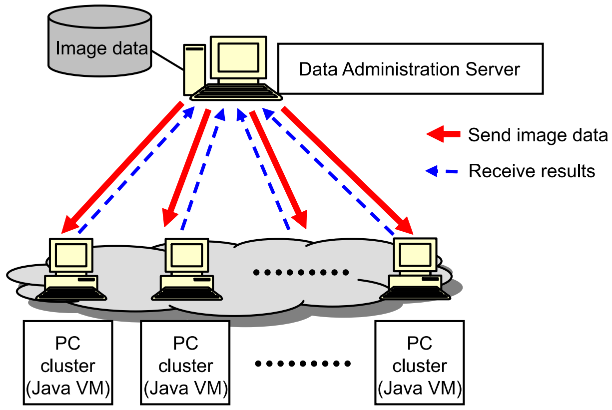

Figure 4: Outline of Sysem Configuration

$$\mu_T = \frac{1}{NM}\sum_{x=1}^{N}\sum_{y=1}^{M} T(x,y) \tag{4}$$

The fitness function is evaluated, and when the fitness result exceeds a predefined value, the specified regions is selected as an optimal match of the search.

4 System Configuration

Figure 4 shows the image searching system configuration. The responsibilities of the data management server are 1)Managing the MODIS Image database, 2)Transmission of search information (MODIS satellite images and search space), 3)Configuration and transmission of DGA parameters. Each PC clustering client receives the search information and GA parameters from the server, and processes the image data search. Taking into account the ease of portability for a multiplatform cluster environment, the system was implemented using Java programming language. Java RMI (Remote Method Invocation) architecture was used to implement the network messaging and control features.

5 Expreiment Results

In this experiment, the effectiveness of the proposed brightness correction method in the image search system is evaluated. The system configuration of the experiment is shown in table 1. Table 2 shows the DGA parameters applied. Figure 5 shows the image of chromosome migration between islands (clients).

The mean retrieval times of 100 independent trials with different random initial populations were used

Table 1: Specification of PC Cluster system

Item	Contents
Number of PC	16
Specifications of PC	CPU: Intel Celeron D 310(2.13GHz) Memory: PC2700 DDRSDRAM 512MB

Table 2: DGA's parameter

Parameters		Contents
Basic GA	Population	200
Basic GA Parameters	Number of generations	2500
	Selection	Comparing fitness value of each generation
	Crossover type	1 point
	Crossover rate	0.5
	Mutation rate	0.05
	Maximum fitness value	0.6
	Strategy of selection	Elite
	Number of island	16
	Interval of migrations	20
DGA Parameter	Number of migrations	1 (population of maximum fitness value)
	Method of migrations	Neighbor (Migration to neighbor island)
	Strategy of migrations	The best chromosome is migration to the worst chromosome of neighbor island.

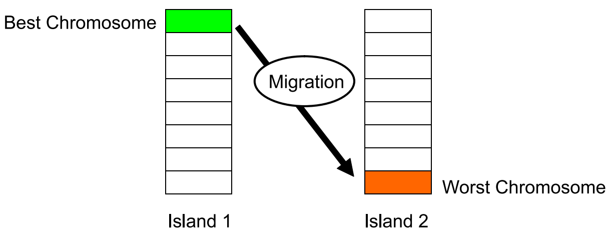

Figure 5: Example of migration between islands

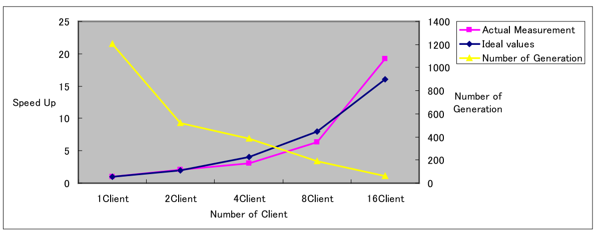

Figure 6: Evaluation results

to evaluate the performance of the proposed system. Figure 6 shows the results.

The measurement of each trial uses the elapsed time of when the performance measurement started searching and found the similarity image data. We were able to verify that by applying normalized correction, the matching accuracy for images with different relative brightness had improved compared to previous research[2]. Parallel efficiency is not enough in comparison with cluster number. However, the searching time of DGA decrease as increasing number of cluster and we confirmed the requested search image that have range of brightness is difference compared with MODIS image data can be searched.

6 Conclusion

We developed a distributed parallel processing system which searches for requested image patterns within a large-scale satellite image database. We applied normalized correction to the DGA fitness function used for the image search algorithm.

The tested results showed that the developed system with the proposed normalized correction method improved the matching accuracy for images with different relative brightness, compared to previous research. From this result we were able to verify that the proposed method was effective for matching images within templates of different relative brightness.

For future works, we plan to research algorithms to effectively search for complex formations and features, as well as investigate different methods to evaluate the time series change in land formations.

References

[1] Akio YAMAMOTO, Hiroshi HASHIMOTO, Application of the Distributed Genetic Algorithm for IN-Core Fuel Optimization Problems under Parallel Computation Environment, Journal of NUCLEAR SCIENCE and TECHNOLOGY, Vol.39, No.12, P.1281-1288, December 2002.

[2] Eiji Nunohiro, Kei Katayama, Kenneth J. Mackin, Masanori Ohshiro, Kazuko Yamasaki, Image Match Search System using Distributed Genetic Algorithm, SCIS&ISIS 2006, pp.1536-1541, September 2006.

[3] Kunio Kashino, Takayuki Kurozumi, Hiroshi Murase, Quick AND/OR Search for Multimedia Signals Based on Histogram Features, The transactions of the institute of electronics, information and communication engineers D-II Vol.J83-D-II No.12 pp.2735-2744, December 2000.