**International Journal of Decision Science**
VOLUME 8 • NUMBER 1 • JUNE 2017

# Research on Performance of Parallel Unsupervised Classification of Remote Sensing Image based on Message Passing Interface

**Hong Liu[1,2*], Yuan Ren, Xiaoliang Ye[1]**
[1] *School of Electrical Information, Shanghai Dianji University, Shanghai 200240, China*
[2] *Department of Physics, Physical Engineering and Optics, University Laval, Sainte-Foy, Quebec, Canada GIK 7P4*
[*] *Corresponding author: liuhong@sdju.edu.cn*

**ABSTRACT**

The requirement on rapid processing of massive data in measurements of remote sensing put challenges on capabilities of computers. Under the parallel programming environment based on the message passing interface (MPI), parallel K-means unsupervised classification of remote sensing images were performed by the choice of different numbers of computers. Furthermore, the efficiency of the parallel computation together with the effect of message communications on the so-called parallel unsupervised classification is also analyzed, respectively. Results show that the velocity of the classification while dealing with massive computational data has been greatly improved by using our method. Besides, the parallel efficiency is also influenced by the parallel unsupervised classification. It is shown that during the parallel programming, message communications should be minimized as far as possible in order to improve the computation efficiency.

***Keywords:*** Remote sensing image, Parallel computation, message passing interface

## 1. INTRODUCTION

With the constant development of remote sensing technology, spatial resolution, spectral resolution and temporal resolution of sensor are significantly improved. The means to collect data also increased, so the obtained remote sensing image data amount rapidly increased. On the other hand, real-time requirements of society on remote sensing information also continuously increased, such as disaster monitoring, fishing condition prediction, precision agriculture and other requirements of synchronous, quasi-simultaneous completion of data acquisition, analysis and processing, quickly provide continuous and practical information. Therefore, the rapid processing of mass remote sensing data put challenges on computer's processing capability. With increase of remote sensing data, it is difficult for traditional stand-alone sequential processing mode to significantly improve processing speed and solve the problem of large memory capacity solving. Distributed parallel computation was widely used in remote sensing image processing. There are many researches on parallel remote image processing data partition [1], remote sensing image parallel processing algorithms, parallel model [2] and etc. An image compression algorithm based on wavelet transform was proposed to achieve idea effect on optical image based on Mallat tower-type wavelet decomposition [3]. Also the image enhancement and implementation of the methods for the digital image enhancement were studied in a detailed manner [4]. In order to aim at high complexity and computationally of scale invariant feature transform operator, an image feature matching algorithm based on fixed scale group scale invariant feature transform operator was brought out [5]. The paper researched on relationships among parallel speedup, parallel efficiency and data amount of remote sensing image of parallel remote image unsupervised classification in computer parallel environment of different number as well as effect on message communication on classification. The paper is organized as follows. Section 2 introduces parallel remote sensing image classification method based on message passing interface (MPI). Section 3 verifies performance of parallel classification method based on specific K-means unsupervised classification and section 4 concludes our work.

## 2. PARALLEL REMOTE SENSING IMAGE CLASSIFICATION BASED ON MESSAGE PASSING INTERFACE

### 2.1 Introduction to message massing interface

The basic idea of our paper is based upon the parallel computation that allocates operation task on stand-alone computer to multiple machines for operation in parallel [6]. It is a kind of parallel programming technique to speedup computation and solve large amount solving problem to divide a traditional serial processing task and allocate it to multiple processing units for processing. Message passing parallel programming mode is one of the most important parallel programming modes, which has lower level and provides flexibility to the programmers. It can solve problems that can not be address by other parallel programming models. Each parallel execution component exchange message, coordinate pace and control operation through message passing.

The MPI was firstly reported in May 1994, which is a standard explain of message passing function library and one of most popular parallel programming environment currently. MPI absorbed multiple advantages of parallel programming environment, namely portability and ease of use. It has also complete asynchronous communications and provides hundreds of function call interfaces, which can be called by variety of programming software.

### 2.2 Parallel remote sensing image classification

The overall objective of remote sensing image classification process is to automatically classify land cover types or land cover classification on all pixels in the image. It is application of statistical pattern recognition technology in the field of remote sensing to conduct remote sensing image classification with computer. In general, remote sensing image processing can be divided into two types of point processing and area processing. In view of nature of remote sensing image classification, it is process of point processing. There are not many connections among pixels of remote sensing image in the classification process. Parallel computation process does not need remote data. Each process unit can avoid too much message communication in the parallel, so as to greatly improve the parallel efficiency.

The basic idea of parallel remote sensing image classification is to divide image data into several sub-blocks and sent to different processing units. Each processing unit perform clustering on corresponding data block. Then final classification results were aggregated to access final result. Parallel image classification based on message passing can be completed with collaboration between the various processes. It needs message communication between them to progress and to coordinate their exchange of messages between cells. In the MPI-based parallel programming, each processing unit sends and receives messages by calling the MPI library functions. Message can be instructions, data and synchronization signal or interrupt signal.

### 2.3 Partitioning of remote image data

Parallel remote sensing image classification divide image data into several sub-blocks and sent to different processing units. Each processing unit perform clustering on corresponding data block. In the process, data partitioning a very important part [2], which can affect efficiency of the parallel image classification.

Partition of remote image data has four types, namely horizontal bands, vertical bands, rectangular blocks and irregular partition [1], where vertical and horizontal bands classification are the most commonly used methods. Data partition requires that the final image data can be merged according to partition order. The ordinary partition is to divide image data in average so that processors can achieve load balance as possible and to obtain best parallel efficiency, and try to reduce mutual communication among processors. The paper used method of horizontal bands data partition.

Assume that *Lines* is number of rows of image and the *ranksize* is total number of processes. To further simplify our computation and let imaging processing procedures be more explicit, we employ the minimize number of lines allocated to each process which was defined in [3]:

$$AverageLines = \frac{Lines}{ranksize} \tag{1}$$

If the *Lines* can not be evenly allocated, there will be *LeftNumber* processes allocated *Average Line*+1 lines:

$$LeftNumber = Lines \bmod ranksize \quad (2)$$

Therefore, the number of process whose number is *rank* was allocated can be expressed as follows [4]

$$LineNumber = \begin{cases} AverageLines + 1 & rank < LeftNumber \\ AverageLines & rank \geq LeftNumber \end{cases} \quad (3)$$

In the following discussion of the paper, we will adopt the algorithm rules such as Eqs. (1)-(3) proposed by Li [3, 4] to design the parallel programming design mode, and further list some typical examples to illustrate the applicability of Li's algorithms fitting to our simulation results.

### 2.4 Parallel programming design mode

Two basic MPI parallel programming design modes are peer to peer mode and master-slave mode. Majority of MPI programs are of one mode or combination of these two. Peer to peer mode is a single program multiple data (SPMD) mode. The status, function and code of this mode are basically same, but different data objects. Master-slave mode is another parallel programming model. In this mode, there are tow kinds of processes, one is mater process and others are slave processes. The master process maintain global data structures and responsible for data partition and allocation. It also keeps message communications with slave processes. The slave process obtains data from messages. It is responsible for data computing and return result to master process. If necessary, slave processes also communicate with each other.

Seen from data characteristics of remote sensing image and processing manner, these above parallel modes are all suitable for parallel image classification. Remote sensing image has different classification modes and methods. In specific program design process, we can adapt different parallel mode according to method itself. The paper used K-means clustering method as example of parallel remote sensing image classification. In the parallel computation of this paper, a master process was set, and Hilbert and singular value truncating algorithms in [6] will be used not only responsible for data partition and allocation, but also keeps communication with slave processes and image data classification. The slave mainly classifies image data, as shown in Fig. 1.

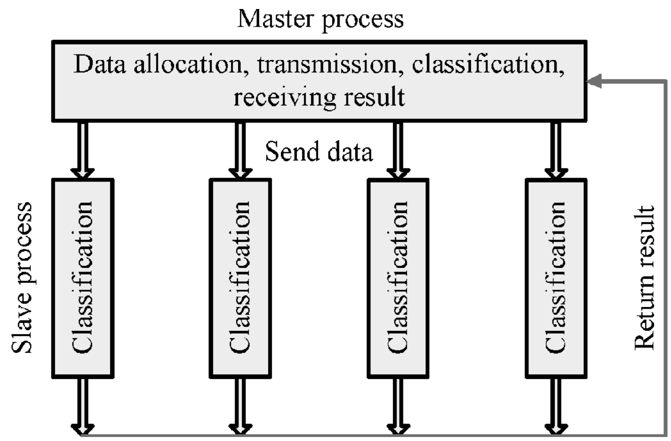


**Figure 1:** Parallel classification mode.

## 3. PARALLEL K-MEANS UNSUPERVISED CLASSIFICATION EXAMPLE

### 3.1 Parallel K-means unsupervised classification algorithm

In the remote sensing image classification, we utilize the method of feature point screening based on fixed scale group operator, which was first defined by Li in [5]. By using Li's method to further process our image data, we are able to find that all image pixels can be divided into supervised classification and unsupervised classification based on whether there is prior knowledge. K-means is the most simple and effective unsupervised classification algorithm. It continuously performs cycle classification on samples to achieve classification. The flow of parallel K-means unsupervised classification algorithm was shown in Fig. 2.

### 3.2 Parallel speedup and efficiency analysis

The main advantage of parallel algorithm is the gain can be achieved compared with serial algorithm, which was also proposed by Li [7]. The basic mechanism of this algorithm is to use parallel K-means unsupervised classification of remote sensing image with different sizes, and we will based upon this algorithm to perform parallel environment with different computer numbers. Two parameters were introduced to evaluate performance of computation,

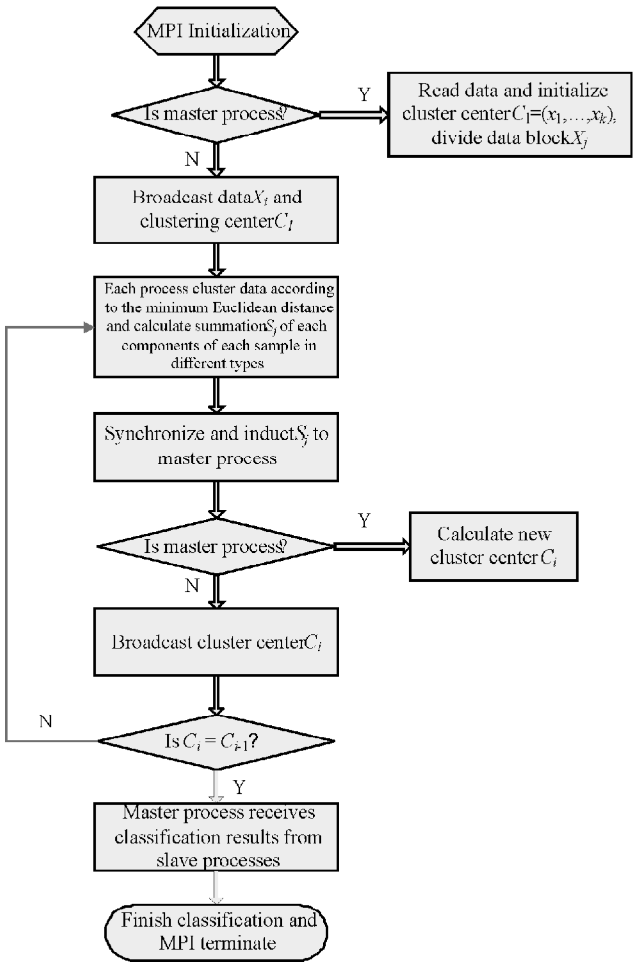


**Figure 2:** Parallel K-means unsupervised classification algorithm

namely speedup $S_p$ and parallel efficiency $f_p$. The speedup is most commonly used evaluation index and used to represent a given application. It is index to measure parallel computation speed that how many times of execution speed of parallel to that of serial algorithm. To best optimize the parallel arrangement process and obtain imaging data more efficiently and accurately, we use the parallel efficiency to measure the utilization of processor [7]

$$S_p = T_1 / T_p;\ f_p = S_p / p \quad (4)$$

Where, $T_1$ is process time of serial algorithm on stand-alone machine and $T_p$ is time needed with parallel algorithm on $P$ processors.

The experiment used Visual C++ .NET 2005 and MPICH. NT. 1. 2. 5 as the development environment of parallel computations. In a LAN, the parallel computing environment of 2 machines and 4 machines were respectively arranged [5]. The settings of machine are exactly same. To facilitate comparison, data come from simulated remote sensing data. 3 bands and 256 gray-level remote sensing images were simulated [5]. The image size includes 20 levels that pixel number from 1,000 to 2000,000. K-means unsupervised classification was conducted under conditions of 2 machines and 4 machines, so as to achieve $T_2$, $T_1$ as well as $S_p$ and $f_p$, as shown in Fig. 3 and Fig. 4.

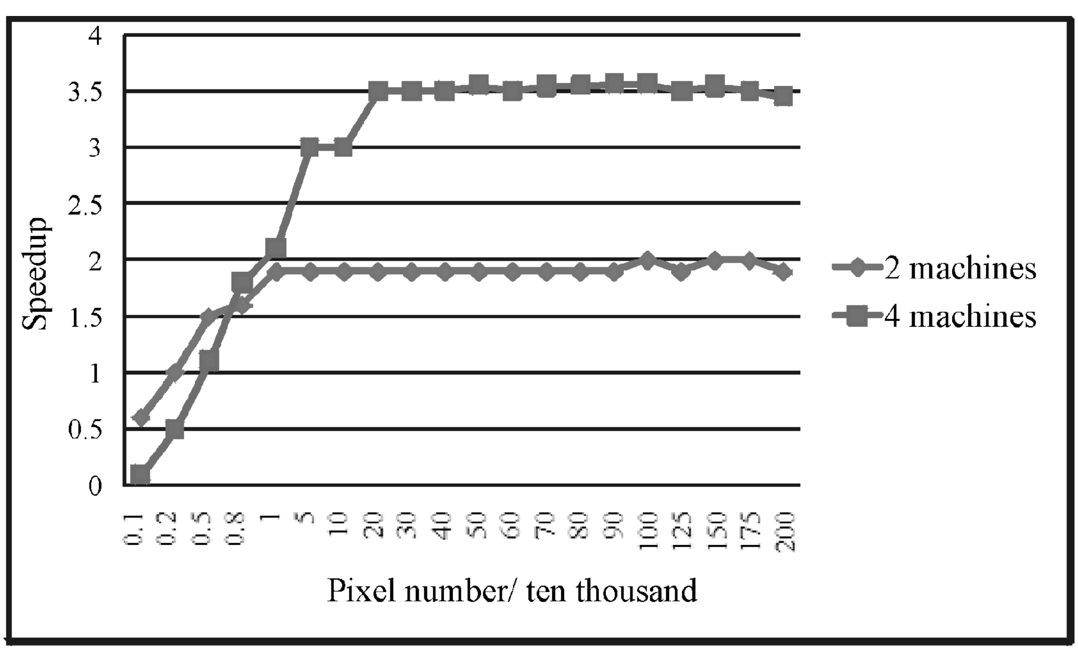


**Figure 3:** Parallel speedup.

From Fig. 3 we can know that when the data amount is less than 2000 pixels, the speedup is less than 1 and processing speed of stand-alone machine is faster than parallel computation. As data amount increase, the speed of parallel computing is significantly faster than that of stand-alone computing, and four computers faster than two. In the parallel computing process, message communication between the processes consumes some running time, so in the case of small data amounts, parallel computing can not show its advantages. Only in computations of large data amount, we are able to verify that the parallel computation algorithm proposed by Li [7] can achieve better computing speed than most other imaging computation algorithms.

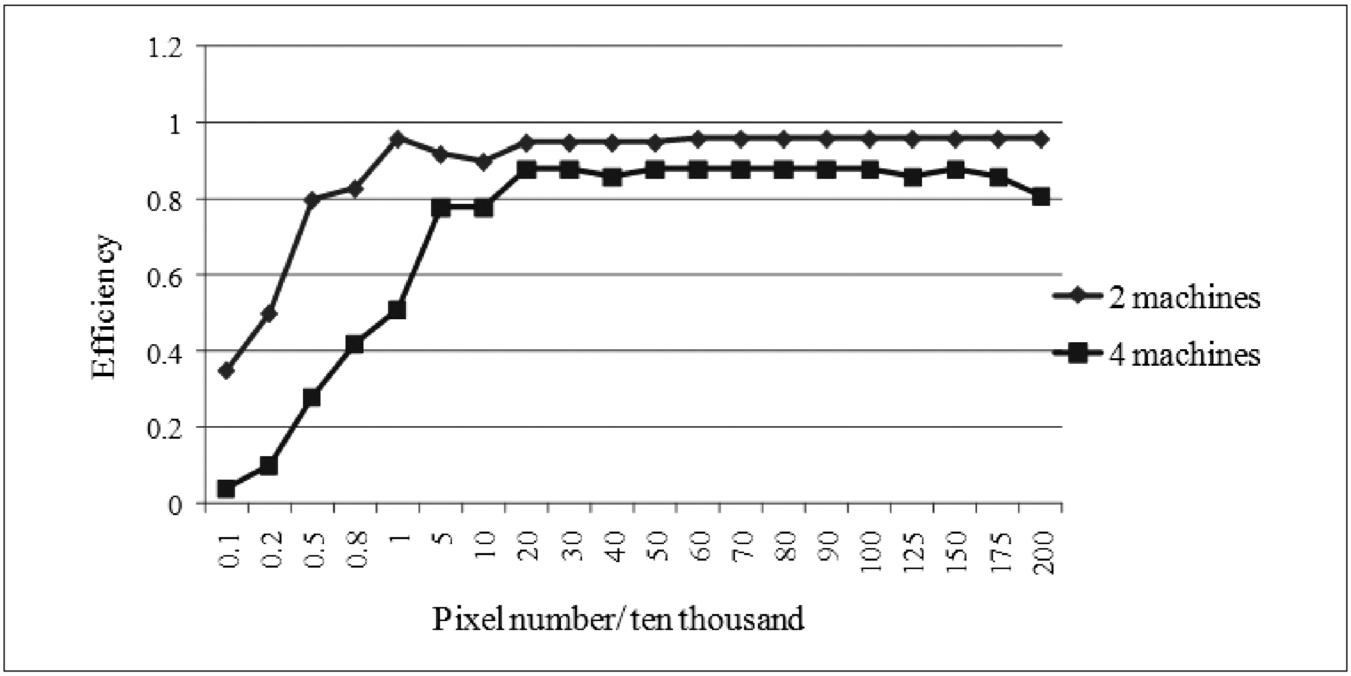


**Figure 4:** Parallel efficiency.

Fig. 4 shows the parallel efficiency $f_p$. The running time of parallel program is consisted of computation time and communication time that can not be overlapped [9]. The parameter $f_p$ reflects the proportion $P_t$ of communication time account for overall computation time to some extent in the parallel computation. When the data amount is relatively small, $P_t$ is larger. As data amount increases, $P_t$ gradually decreased. From above comparison, the parallel computation introduced in [6, 7] can be regarded as very perfect and convenient tool, which has essential applications to large amount computation work nowadays. On the other hand, the parallel efficiency $f_p$ of 2 machines is higher than that of 4 machines. When the parallel efficiency of these two reach stable, the average efficiency of 2 machines is 0.946 and 4 machines is 0.897. This is because as number of machines increases, the communication overhead also increases, which reduce the parallel efficiency to some extent [10].

## 4. CONCLUSIONS

The parallel programming based on MPI can realize unsupervised classification of remote sensing image, especially in the classification process of large amount remote sensing images, the classification speed can be greatly improved. In the parallel K-means unsupervised classification, repeated clustering needs constant communication among master process and slave processes, so the parallel algorithm has large communication amount, which reduce parallel efficiency to some extent. The core problem of MPI parallel program design is task allocation and communication. in the parallel computing process, it is inevitably to introduce communication time among processes. In the parallel program design, we should try to reduce message communication as possible and take optimization measures to improve operation efficiency.

## ACKNOWLEDGMENTS

This work is supported by the National High Technology Research and Development Program of China (2009AA032708), the National Natural Science Foundation of China (NSFC 61205121, 61304124), the computer application subject in Shanghai DianJi University (13XKJ01).

## REFERENCES

[1] Guo-man Huang, Jian-feng Guo. Data partition for distributed-parallel processing of remote sensing imagery. Remote Sensing Information, 2001, 2: 9-12.

[2] Jun Li, De-ren Li. Key techniques for distributed processes of remote sensing imagery. Journal of Wuhan Technical University of Surveying and Mapping (STUSM), 1999, 24: 15-19.

[3] Jia Li, Liping Chang. A SAR compression algorithm based on Mallat tower-type wavelet decomposition. Optik, 2015, 126: 3982-3986.

[4] Jia Li. Application of image enhancement method for digital images based on Retinex theory. Optik, 2013, 124: 5986-5988.

[5] Jia Li. An image feature point matching algorithm based on fixed scale feature transformation. Optik, 2013, 124: 1620-1623.

[6] Jia Li. An improved wavelet image lossless compression algorithm. Optik, 2013, 124: 1041-1044.

[7] Jia Li, Yali Qin, Hongliang Ren. Research on parallel unsupervised classification performance of remote sensing image based on MPI. Optik, 2012, 123: 1985-1987.

[8] Ying Jiang, Yong-mei Lei. Some algorithms' parallel programming universal modules in MPI. Computer Engineering and Applications, 2003, 39: 139-141.

[9] Ming-chang Hu, Gang Shi, Wei-wu Hu. How does communication influence clusters parallel computing performance. Mini-micro Systems, 2003, 24: 1569-1573.

[10] Li-jun Lu, Ming-sheng Liao, Lu Zhang. Application of parallel computing techniques of distributed memory multiprocessors to remote sensing image processing. *Journal of Geomatics*, 2005, 30: 1-3.