

# 地理数字图像机群并行处理试验<sup>\*</sup>

Test of Geographic Digital Image Parallel Processing Based on PC Group

方金云 何建邦 池天河 胡伟武 史岗 张福新

(中科院地理所 北京100101) (中科院计算所 北京100083)

**Abstract** The image processing speed and efficient are a bottle-neck of the geographic image processing. Based on JIAJIA, which is a distributed shared memory system, the parallel image processing is discussed in this paper. The parallel algorithms are realized, and tested by TM and SAR images in Poyang Lake area in particular. The testing result is a linear correlation between speed-up and amount of processor.

**Keywords** Image processing, JIAJIA, Parallel algorithm, Linear speed-up

地理数字图像是指与地理位置有关的数字图像,包括栅格图像、矢量图形等。随着数字地球时代的到来,地理数据采集设备分辨率的提高,地理图像所蕴含的信息量呈海量增长<sup>[2]</sup>,这要求图像处理速度应加快,效率应提高。因此,并行性引入数字图像处理成了必然趋势<sup>[3]</sup>。地理图像处理中一般为几十兆—几百兆字节的矩阵操作,需大内存;分块矩阵计算过程中通信量少,因而地理数字图像处理易于并行化。

图像并行化处理需要基于一定的并行化环境和系统。由于目前较流行的并行处理系统如 IVY、Midway、Munin、CVM、TreadMarks 等大都基于工作站,在此基础上产生了一系列图像处理的并行算法<sup>[4~7]</sup>和系统<sup>[8~9]</sup>。但组建一套这样的系统需几百万元人民币。本试验选择了适于机群并行计算的 JIAJIA 系统作为开发平台<sup>[10]</sup>,以实现低费用高效率的并行图像处理算法。

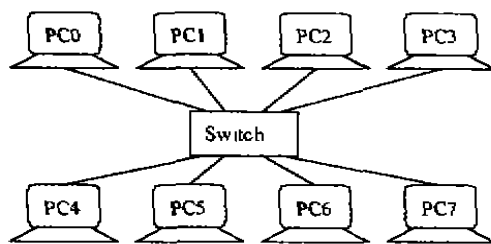


图1 并行机群

## 1 试验环境设计

由8台主频400M、内存256MB的PC机和1台

Switch组成机群以太网,采用 LINUX1.1.2操作系统和基于分布式内存共享的 JIAJIA 开发平台。

机群系统的优点是利用现成计算资源,用较少的投资实现计算、内存、文件等资源共享,进而实现高性能计算。整套并行系统约人民币10万元。

## 2 并行算法

地理图像的处理包括数字图像预处理、图像变换、图像增强、图像分类、信息提取等方面的内容。许多串行处理算法早已提出并付诸使用<sup>[11]</sup>,但这些串行算法存在不少问题:首先是由串行性本身引起的技术问题,如采样精度;其次是速度问题,对于较大图像处理效率很低,其中图像几何精校正和图像分类在处理中最为费时,故将其并行算法的实现作为本试验的重点。

单机串行算法和机群并行算法的根本区别在于:前者由于受内存限制,每次只处理一块数据,而后者将数据一次全部读入内存并分配到各个PC机同时处理。

### 2.1 几何精校正

几何纠正指校正像元传感器由于几何视角和地形变化等原因引起的相对位置误差。一般说来,校正是数据文件坐标到另一种栅格或坐标系统(称作参考坐标系统)的转换。

无论单机串行还是多机并行处理,都包括以下通用步骤:(1)确定地面控制点;(2)计算转换矩阵;(3)像元重采样生成新的输出图像文件。

几何精纠正的并行算法如下:

<sup>\*</sup> 本文得到国家重大自然科学基金69896250-4项目和中国科学院“九五”基础性重大研究项目 KJ951-B1-703-02资助。

```

①分配共享变量 jiatmp 空间;
②分配内存空间给待处理图像 ImageIn 和结果图像 ImageOut;
③由 0 号处理机
{
    读入控制点;
    读入图像大小 (col1, lin1) 等参数;
    计算多项式系数 aa[], bb[] 以及输出图像大小 (col2, lin2) 等;
    把 col, lin, col2, lin2, aa[], bb[], P 等存入临时共享变量 (如 jiatmp[0]=col1; jiatmp[1]=lin1;)
}
④由 1 号处理机把输入图像读入 ImageIn;
⑤将 ImageIn 中的数据分块, 数据行的位置为 (start, end), 并将数据分配到各处理机;
⑥由各处理机同时处理各相应的分块矩阵, 生成结果图像。
for (I=start; I<end; I++)
    for (j=0; j<col2; j++)
        {
            x=xy(P, I, j, aa);
            y=xy(P, I, j, bb);
            ImageOut[I][j]=ImageIn[y][x];
        }
⑦由 0 号处理机把 ImageOut 输出到图像文件;
}

```

①②为初始化处理; ③④为并行计算, 其中③完成步骤(1)(2)中的处理, ⑤⑥为并行计算, 完成步骤(3)中的处理, ⑦串行输出。

### 2.2 非监督分类

非监督分类就是将像元按照它们在性质上亲疏远近的程度进行分类, 其具体算法是:

- (1) 输入类数  $k$  和相似度, 任选  $k$  个像元作为原始聚类中心  $M_i(1), i=1, 2, \dots, k$ ;
- (2) 计算各像元到聚类中心  $M_i(1)$  距离, 并将之归到最近的类, 形成  $k$  类;
- (3) 计算各类均值向量  $M_i(2), i=1, 2, \dots, k$ ;
- (4) 比较两个均值向量, 若  $M_i(2)=M_i(1)$ , 结束。否则, 重新执行(2)~(4)。

非监督分类由于需对多幅图像同时处理, 所以并行算法较复杂, 仅给出主要部分的并行化算法:

```

{
    ... (公共并行, 私有数组初始化);
    计算各处理机所处理数据行的位置 (start, end);
    对所有处理机执行以下操作:
    for (j=start; j<end; j++)
        for (k=0; k<列数; k++)
            for (i=0; i<波段数; i++)
                tm[i]=(unsigned int)band[i][j]*col+k;
        for (g=0; g<类数; g++)
            dd[g]=0.0;
            for (i=0; i<波段数; i++)
                dd[g]+=fabs((float)tm[i]-occ[i][g]);
    }
    .....
}
处理机同步;
jia_lock(0);
for (j=0; j<类数; j++)
    for (i=0; i<波段数; i++)
        jiancc[i][j]+=ncc[i][j];
        jiasc[i][j]+=ss[i][j];
}
jian[j]+=n[j];
}
jia_unlock(0);
}

```

```

处理机同步;
for (j=0; j<类数; j++)
    for (i=0; i<波段数; i++)
        ncc[i][j]=jiancc[i][j];
        ss[i][j]=jiasc[i][j];
}
n[j]=jian[j];
}
.....
}

```

### 3 试验结果

在并行算法实现的基础上, 利用标准 C 语言, 实现了地理图像并行处理的原型系统 (ParGIP), 并用其对鄱阳湖地区的 TM 图像和 SAR 图像进行试验。

滤波采用  $5 \times 5$  中值滤波器剔除噪音, 并有效地防止图像边缘模糊; 几何纠正采用二元三次多项式和二元四次多项式校正图像; 分类采用非监督分类方法, 分别预设四类和五类。

表 1 并行图像处理试验结果

程序	图像大小	所需内存	串行时间(秒)	并行时间(秒)	加速比
滤波	13200 * 15000	378MB	202	12	16.1
三次多项式纠正	13200 * 15000	378MB	1510	91	16.1
四次多项式纠正	13200 * 15000	378MB	6724	174	38.1
分五类	6546 * 5728	286MB	1285	114	11.1
分八类	6546 * 5728	286MB	4081	395	10.1

从表中可看出, 地理图像处理一般需要几百兆的内存空间, 而大多数 PC 机尚不具备这一条件, 导致处理过程中对硬盘的多次读写, 致使处理速度慢。并行处理是解决这一问题的关键, 无论采用何种图像并行处理算法, 均具有非常好的线性加速比。

**结论** 地理图像处理为几十兆—几百兆字节的矩阵操作, 需大内存和较高的处理速度。并行处理技术是解决地理图像处理速度的关键技术, 并行图像处理算法是并行系统的核心。基于分布式内存共享的 JIAJIA 系统在图像处理中与处理器个数具有线性加速比。

在以上的算法试验的基础上, 已开发了并行图像处理的原型系统, 专业化的并行地理图像处理系统预计在年底面世。

### 参考文献

- 1 陈景良. 并行数值方法. 清华大学出版社, 1983
- 2 陈述彭. 地球信息科学及其应用. 卫星应用, 1997, (3): 1~9

服务器对外接口不变,就能顺利地调整系统的功能,而该过程对用户是透明的。

·为传统的信息系统移植到互联网环境中提供了一种办法。现在,许多原有的信息系统将要移植到互联网环境中。若采用三层结构,我们不仅要完成网站页面的设计,还要重新编写大量的程序以实现原有信息系统的应用逻辑。这样不仅做了大量的重复劳动,而且其可靠性也没有保障。采用四层结构,只需在原有信息系统服务器程序上“粘”一块程序,相应地在 Web 服务器方也“粘”一块程序,就可以很快地将原有系统移植到互联网环境中。

当然,互联网应用的四层结构也有不足之处,由于用户的请求被转发了一次,这可能会影响系统运行的效率。但任何事情都具有两面性,笔者认为采用四层结构的好处:可重用性、易开发性、易维护性、易集成性,远远大于其带来的负面影响。另外,必须明确一点,四层结构并不是对于任何互联网应用系统都是适合的。只有对于那些复杂的以应用逻辑为中心的互联网应用系统,四层结构才是比较适合的。

### 3.3 几种结构之间的比较

下面我们从适用对象、交互性、动态性、性能、易维护性和可重用性几方面对典型的 B/S 结构、三层结构和四层结构做一个简单比较(表1)。

**结束语** 随着互联网应用的日益广泛,出现了以应用逻辑为中心的复杂的互联网应用系统,原有的支持以数据库为中心的互联网应用的体系结构已经不适合,必须采用新的体系结构来支持复杂的互联网应用系统。本文在分析复杂互联网应用的基础上,提出了一种针对复杂互联网应用系统的新的互联网应用体系结

构——“浏览器-Web 服务器-应用服务器-数据库”。该结构有利于复杂互联网应用系统的快速开发、维护和重用,也为原有信息系统移植到互联网环境中提供了一个很好的办法。

表1

	适用对象	交互性	动态性	性能	易维护性	可重用性
典型的 B/S 结构	以信息发布为主的静态的网站	差	差	好	差	差
三层结构	以数据库为中心的互联网应用	较好	好	较好	中等	中等
四层结构	以应用逻辑为中心的互联网应用	好	好	稍差	好	好

### 参考文献

- Gellersen H, Gaedke M. Object-Oriented Web Application Development. *Internet Computing*, 1999 (January-February): 60~68
- Gellersen H W, Wiche R, Gaedke M. WebComposition: An Object-Oriented Support System for the Web Engineering Lifecycle. 1996. Available at: <http://proceedings.www6conf.org/HyperNews/get/PAPER232.html>
- 熊忠阳, 张玉芳, 吴中福. 三层结构中的数据库访问技术. *计算机科学*, 2000, 27(4): 95~97
- Fan Xin, Chen Jian. Design for Maintenance: Experience of Developing an ICP Web Site. In: 1<sup>st</sup> ICSE Workshop on Web Engineering, Los Angeles, USA, 1999
- Zhao Wei-quan, Chen Jian. CoOWA: A Component Oriented Web Application Model. In: Proc. of 34<sup>th</sup> Intl. Conf. on TOOLS Asia. IEEE Computer Society, Nanjing, 1999

(上接第100页)

- Akl S G. The Design and Analysis of Parallel Algorithms. Prentice-Hall, Englewood Cliffs, NJ, 1989
- Chellappa, R, Chatterjee S. Classification of Textures Using Gaussian Markov Random Fields. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1985, 33(August): 959~963
- Cohen F S, Cooper D B. Simple parallel hierarchical and relaxation algorithms for segmenting noncausal Markovian fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(March): 195~219
- JaJa J. An Introduction to Parallel Algorithms. Addison-Wesley Publishing Company, New York, 1992
- Bader D A, JaJa J, Chellappa R. Scalable Data Parallel Algorithms for Texture Synthesis and Compression Using Gibbs Random Fields. [Technical Report CS-TR-3123 and

UMIACS-TR-93-80]. UMIACS and Electrical Engineering, University of Maryland, College Park, MD, August 1993

- Johnsson S L, Jacquemin M, Krawitz R L. Communications Efficient Multi-Processor FFT. *Journal of Computational Physics*, 1992, 102: 381~397
- Palmer J, Steele Jr C L. Connection Machine Model CM-5 System Overview. In: *The Fourth Symposium on the Frontiers of Massively Parallel Computation*, IEEE Computer Society Press, Los Alamitos, CA, 1992: 474~483
- Lu H, Dwarkadas S, Cox A, Zwaenepoel W. Quantifying the Performance Differences Between PVM and TreadMarks. *Journal of Parallel and Distributed Computing*, 1997(42)2: 65~78
- Castleman K R. *Digital Image Processing*. Prentice Hall, 1997