

# 遥感图像波谱角并行分类算法

刘晓云<sup>1</sup> 康一梅<sup>1</sup> 齐同军<sup>2</sup> 方金云<sup>2</sup>

(北京航空航天大学软件学院 北京 100191)<sup>1</sup> (中科院计算技术研究所 北京 100190)<sup>2</sup>

**摘要** 针对海量遥感数据处理提出了一种遥感图像波谱角并行分类算法。该算法是基于单机波谱角算法。针对其处理遥感图像数据速度慢、效果差、内存不足等缺点,提出了在集群环境下遥感图像波谱角并行分类算法的同步、互斥和负载均衡等策略,以及在中台设备上并行处理的方法。通过对算法的时间复杂度、加速比进行分析,并在集群环境下进行实例验证,将分类图像与知名软件 ENVI 进行对比,验证了算法的优越性和有效性。

**关键词** 并行处理,波谱角分类,遥感图像分类

中图法分类号 TP751.1 文献标识码 A

## Algorithm of Spectral Angle Parallel Classification on Remote Sensing Image

LIU Xiao-yun<sup>1</sup> KANG Yi-mei<sup>1</sup> QI Tong-jun<sup>2</sup> FANG Jin-yun<sup>2</sup>

(School of Software Engineer, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)<sup>1</sup>

(Instruction of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)<sup>2</sup>

**Abstract** This paper proposed a spectral angle parallel classification algorithm on remote sensing image against mass remote sensing data processing. Based on single-machine spectral angle algorithm, aiming at defects in processing remote sensing image data, such as slow speed, poor effect, and insufficient internal memory, this algorithm proposed some synchronized, exclusive and load balancing strategies against spectral angle parallel classification algorithm on remote sensing image, as well as method of parallel processing on multiple equipments under cluster environment. The paper compared classified image with famous software ENVI through analyzing time complexity and speedup ratio of the algorithm and doing verification by actual examples under the cluster environment to inspect and verify the superiority and validity of the algorithm.

**Keywords** Parallel processing, Spectral angle classification, Remote sensing image classification

在遥感应用中,通过遥感图像处理和判读识别各种地物是一个主要的目的。无论是地物信息提取、土地动态变化监测,还是专题地图制作和遥感图像库的建立等,都离不开分类。分类的目的是从图像中识别实际地物,进而提取地物信息。对各种原始遥感图像或者具有各种变化的图像使用统计方法达到分类的目的。

遥感图像分类处理主要有两大类:监督分类和非监督分类。非监督分类是在完全没有先验知识的基础上完全由计算机自行分类的方法,所以这种分类方法一定要后期通过实地勘测把每一类赋予真实意义,只在在地物没有太多了解的基础上使用。监督分类是在有先验知识的基础上进行分类,主要有 6 种比较成熟的方法:平行六面体算法简单,但它是按照均值和标准差划分平行六面体,与实际地物数据点分布的种群形态不一致,因此效果不好;最小距离分类方法,如果波段等方差分布,则分类效果较好,算法也相对简单,否则就会出现错误;最大似然分类,由于它是基于 bayes 准则的一种非线性分类方法,如果参加分类的波段相关性比较强,则方差、协

方差的逆矩阵不存在或不稳定,那么分类结果不准确;马氏距离分类,虽然排除了变量之间的相关性的干扰,但夸大了变化微小的变量的作用;二值编码分类,在编码过程中忽略了细节,地物光谱比较接近的很难区分,只能粗略地对图像分类;波谱角分类处理,除可以利用自选的样本外,还可以利用波谱库中的值作为监督样本进行分类,更能反映实际地物,计算波谱角也不涉及波段间的相关性,所以这种分类方法是一种既算法简单又分类正确的算法。波谱角分类算法实际应用中,对波谱比较接近的农田和林区,分类结果非常好;对于江心,只有这种算法分类比较合理<sup>[7]</sup>。

波谱角分类需要对图像的每个像元求波谱角。一幅遥感影像所含的像元数目非常大,而且每分钟传回地面的图像能达到几十 TB,这样在处理速度上就要有更高的要求。如果处理速度不能达到一定的要求,就难以实现应用的真实意义。例如,对洪水的淹没面积的监测、对火灾面积变化的监测等,如果不能及时地处理,就不能采取及时的措施来降低损失。

本文基于单机波谱角算法,在集群环境下解决了并行计

到稿日期:2008-10-30 返修日期:2009-01-07

刘晓云(1980-),女,硕士,主要研究方向为并行遥感图像处理,E-mail: xiao\_yun\_liu@yahoo.com.cn;康一梅(1968-),女,博士,副教授,主要研究方向为并行算法、无线传感器网络、嵌入式系统设计等;齐同军(1974-),男,博士生,助理研究员,主要研究方向为分布式计算、并行处理、空间数据库;方金云(1968-),男,博士,副研究员,主要研究方向为海量空间数据处理关键支撑技术、网格 GIS 等。

算的同步、共享变量的互斥和负载均衡等技术问题,提出了遥感图像波谱角并行算法,实现了多台处理器并行处理遥感图像的分类。并通过试验验证了并行波谱角算法在处理遥感图像时时间明显减少、处理效果较好以及弥补了内存不足这一缺陷。

## 1 波谱角算法描述

### 1.1 算法介绍

波谱角分类是一种波谱匹配技术,通过计算图像元光谱与样本参考光谱之间的夹角来划分类别。

波谱角分类方法的原理是:把光谱作为向量投影到  $N$  维空间上,其中维数  $N$  是所进行计算的波段数。 $N$  维空间中,光谱曲线被看作有方向、有长度的向量,各个光谱之间形成的夹角叫波谱角。

波谱角的数学表达式<sup>[4]</sup>为

$$\alpha = \cos^{-1} \frac{\sum xy}{\sqrt{\sum(x)^2 \sum(y)^2}} \quad (1)$$

其中,  $\alpha$  为影像元光谱与参考光谱之间的夹角(波谱角);  $x$  为影像元光谱曲线向量;  $y$  为参考光谱曲线向量。

$\cos \alpha$  的值接近 1 时,有最好的估计光谱值和类别分类结果。

### 1.2 波谱角分类的流程

(1) 计算模板各类中各个波段的均值,并且计算分类像素值与模板每个类中各个波段均值的积,求出所得乘积各个波段的和。

(2) 分别求出分类像素值各个波段平方和以及每类中各个波段的平方和,最后求出这两个平方和的积的平方根。

(3) 将(1)和对应每一类的(2)结果相除,求该结果的反余弦函数。将该像素点与所有类做(1)–(3)的计算。

(4) 求出角度最小的归类。

(5) 判断是否计算完所有的分类像素。如果完毕,结束分类;如果未完,继续(1)–(4)计算。

波谱角计算流程图如图 1 所示。

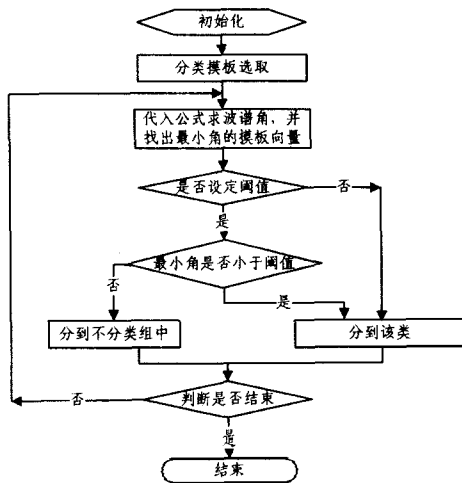


图 1 波谱角计算流程图

## 2 并行处理分析

### 2.1 算法分析

从算法流程上看,波谱角算法比较简单,单次执行时间很

短,几乎可以忽略。但对于海量的遥感,图像每分钟都会有几十 TB 的数据,甚至更多,分类就是要对每个像素计算一次,需要的时间已经不能用多少个小时计算。就算计算结束,可能已经错过了应用的最佳时期,有些已经不具备现实意义了。而且对于一个 PC 内存很难满足需求,计算数据无数次 I/O 操作有可能丢失数据,那么就将图像中的信息丢失,这样处理效果则不好。应用并行波谱角分类算法则能很好地解决以上问题。

### 2.2 能够分布式并行处理的特征

对于各种数据的分析,应用并行处理的效果较好的数据应具备以下特点:

1) 计算量大。如果小计算量的工作用并行处理,体现不出并行处理的优势,效果不明显。

2) 每个数据点上的计算形式相同,有利于将大的数据划分为小的区域,在不同的机器上操作。

3) 每个数据比较独立,数据间的关联性比较小,有利于处理后数据还原成原始数据<sup>[1]</sup>。

具有了以上特征的数据,就可以利用分布式并行处理技术进行处理,遥感图像分类处理就具有这样的特征。图像本身数据量很大,计算量也很大,分类时对每个数据进行相同的计算,并且每个像元之间基本是独立的。

### 2.3 解决波谱角并行算法关键问题的策略

波谱角算法完全具备并行化的特征,但在并行化处理中需要解决一些并行计算时会出现的问题。下面对这些问题进行分析并给出解决策略。

1) 共享变量。在波谱角分类中,对于共享变量(如均值、方差等值),各个处理器的进程对这些变量操作时为了保证数据的正确性和完整性,要对这些变量进行互斥操作。采用加锁机制,遵循域一致性模型,即对一个共享变量操作必须通过对该变量加锁和解锁来实现。

2) 全局同步。在波谱角处理中,每台处理器上的分类计算必须同时完成。为了使各个机器上达到进程同步,在程序设置了栅障处理,作用是暂停当前进程,等待所有进程达到这个栅障处,直到域内所有数据同步为止。

3) 负载均衡。在并行处理中有效地分配计算量,实现负载均衡,在同构的集群中可以减少同步时间。波谱角计算为了有效利于负载均衡,采取了按图像的行平均分割数据的方法,因为图像数据中数据间几乎没有联系,这样处理完全不影响处理结果。

### 2.4 波谱角分类并行算法原理及流程

根据对单机波谱角算法的分析,考虑到并行波谱角算法要解决的问题及对这些问题所采取的策略,得到波谱角并行算法流程图,如图 2 所示。

算法步骤如下:

1) 将原始图像按照行划分成  $N$  块,  $N$  依据处理器的台数确定。通过配置文件来定义处理器的数目,并对处理器编号。  
\* start 表示每一台处理器数据块的开始坐标, \* end 表示每台处理器的结束坐标, jiahost 变量存储处理器数目, jiapid 变量标识每个处理器的编号。根据图像的行数对整幅图像进行分块:

$$* \text{start} = \text{lin} / \text{jiahosts} * \text{jiapid};$$

$$* \text{end} = * \text{start} + \text{lin} / \text{jiahosts}.$$

然后将每一块数据和波谱角分类算法代码通过网络传输到各个处理器上等待处理。

2) 在每台处理器上运行波谱角分类算法, 这是计算量最大的部分。每一台处理器对相应的数据进行分类, 并把结果存到相应坐标变量中, 实现同步操作, 保证数据的一致性。计算结束后恢复成原始图像。

3) 结束分类。

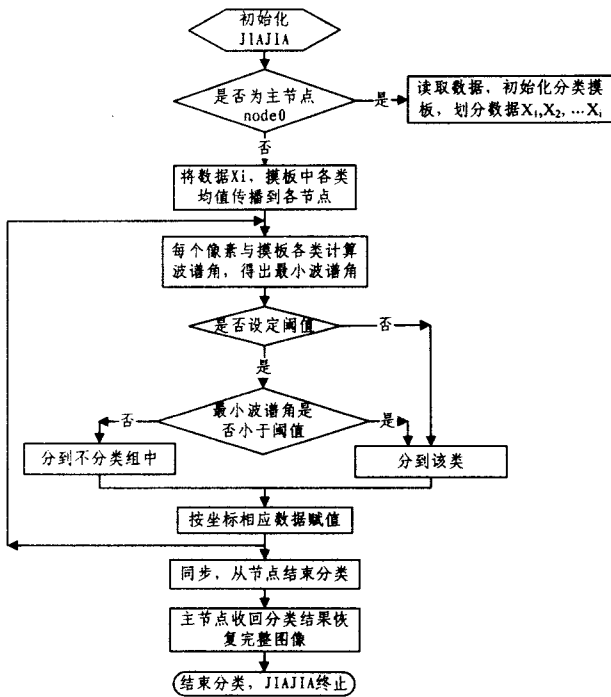


图2 并行波谱角算法流程图

### 3 并行波谱角分类算法时间复杂度分析

根据程序的执行过程计算出算法的时间复杂度。并行波谱角分类算法处理过程是在每台机器上单独进行的, 对于每台处理器分配的图像的像素数  $a$  是一个常数, 从而得出时间复杂度为  $O(an)$ , 其中  $n$  是分类语句执行的时间规模。如果不是用并行算法, 时间复杂度大约是并行算法的  $jiahosts$  倍, 具体由参与处理的处理器台数决定。

## 4 实验验证

### 4.1 多台 PC 处理结果分析

实验环境介绍: JIAJIA 采用一种基于锁的新型 Cache 一致性协议来实现内存存储一致性模型。与传统的基于目录的协议相比, 基于锁的协议使处理机通过访问附带有锁上的 write-notice(用来记录一个页面是否被修改过)来维护一致性, 从而避免由目录引起的存储开销和系统复杂度。为了降低系统开销, JIAJIA 只支持一种存储一致性模型和一种写传播策略(写无效策略), 同时采用多写(multiple writer)协议来避免假共享。JIAJIA 系统的另一个特征是它能够把多个机器的存储器组织起来, 形成一个更大的存储空间, 其最大容量可达每个机器的局部存储器之和<sup>[2]</sup>。基于 JIAJIA 的集群分布式结构图如图 3 所示。

实验 1 本试验基于 linux 平台, 使用 C 语言和 JIAJIA7 为软件并行开发环境, 通过局域网将多台机器组织成集群环

境。使用 3 个波段、255 个灰度级的遥感图像, 图像大小为  $2667 * 2667$  个像素。

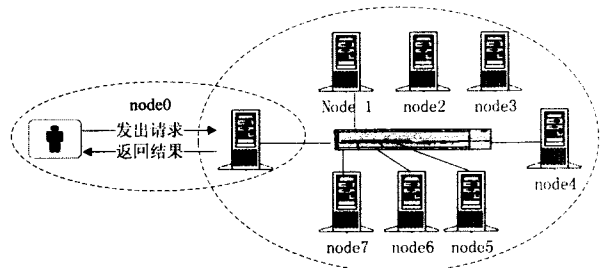


图3 基于 JIAJIA 的集群分布式结构图

用不同台数的处理机对相同的图像进行处理, 得到加速比。加速比是相对于处理中只用一个节点时的执行时间, 如表 1 所列。

表 1 不同处理器运行统计

处理器数目	耗时(秒)	加速比
1	10.33	1.000
2	5.45	1.900
4	3.01	3.431
8	1.36	7.595

结果分析: 图像波谱角处理算法获得了接近线性的加速比, 这样的加速比决定了图像处理速度的优势。但随着处理器的增加, 通信时间就随着增加。而当分类处理所用时间减少的速度小于通信时间增加的速度时, 加速比会偏离线性加速比。这种现象在处理器达到一定数目会出现, 所以处理器不是无限增加的, 在本实验中未达到。

### 4.2 对比结果分析

实验 2 对相同图像用 ENVI 中波谱角分类和并行波谱角分类(2 台机器)进行处理的时间对比。ENVI 是当前比较流行的遥感图像处理软件, 通过与 ENVI 对比来验证并行波谱角分类的正确性, 如表 2 所列。

表 2 ENVI 与并行处理时间对比

编号	大小(M)	像素数目	ENVI 耗时(秒)	并行耗时(秒)
1	101	13280 * 2604	120	69
2	203	26560 * 2560	235	102

结果分析: 两台机器并行处理的速度比 ENVI 速度快将近 1 倍。大部分的遥感图像数据都在几十个 GB, 甚至更大, 所以从以上处理来看, 并行波谱角分类算法的优势更明显。

### 4.3 两个软件处理后的效果图对比

从效果图(如图 4 所示)看, 由并行波谱角计算的方式得到的图像与 ENVI 软件得到的图像几乎相近, 但其在某些区域要比 ENVI 分类结果好, ENVI 分类的图中有很零星的不能连片的单个像素区域。仔细对比发现, 并行分类结果更能反映原始图像, 如表 3 所列。

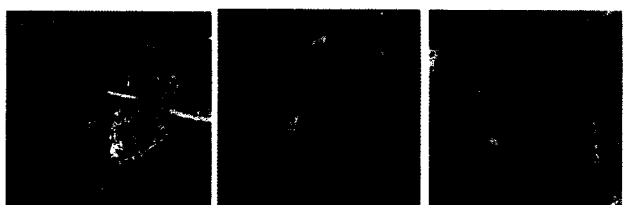


图 4 效果图

表3 ENVI与并行波谱角算法的像素分类百分比

类号	并行处理分类(%)	ENVI波谱角分类(%)
0	46.86	45.962
1	15.16	15.258
2	17.43	16.753
3	19.45	19.558
4	1.09	0.254

结果分析:从分类的百分比来看几乎接近,在并行处理中有99.99%的像素点参与分类,ENVI中有96.185%的像素参与分类。分类精度几乎接近,可以体现分类的正确性。

结束语 本文提出了分布式并行遥感图像的波谱角分类处理并行算法。经过多次不同图像的分类测试,验证了用此方法处理图像时,在计算量小的情况下,处理速度的优势几乎不能显现;但处理数据量大的情况下,大大减少了计算时间,即该算法对于处理海量的遥感图像优势很明显。其次,通过与ENVI分类结果进行比较,验证了并行波谱角处理的正确性,并且本文提出的并行波谱角分类算法对高光谱分类处理更为有效。

### 参考文献

[1] 黄国满,郭建峰.分布式并行遥感图像处理中的数据划分[J].遥

感信息,2001(2):9-10

[2] 方金云,池天河,黄国满,等.基于机群的地理数据并行处理试验[C]//中国地理信息系统协会第本局年论文集(2).2001:158-162

[3] 曾丽芳,杨学军,黄春,等.基于JIAJIA统的消息传递和共享存储编程模式比较[J].计算机工程,2002(10)

[4] 张银辉,罗毅,赵庚星.垦利县耕地利用动态监测研究[J].自然资源学报,2003,18(6):655

[5] 胡明昌,史岗,胡伟武,等.PC机群上JIAJIA与MPI的比较[J].软件学报,2003,14(7)

[6] 何月,邢立新,许军强.毛乌素沙地典型地段沙漠化土地格局分析[C]//第二届全国国土资源遥感技术交流会.2006-07-21

[7] 韦玉春,汤国安,杨昕,等.遥感数字图像处理教程[M].北京:科学出版社,2007:231-233

[8] 方金云,何建邦,池天河,等.可扩展的并行地理图像处理系统[J].地球信息科学,2000(12):62

[9] 吴洪桥,池天河,方金云,等.基于SVM的遥感数据并行服务体系研究[J].南京理工大学学报,2004,28(4):352-353

(上接第261页)

从表1中可以看出,未经优化的H.264编码算法在第一次再编码时会有15%到20%左右的视频畸变。与原始clip算法相比较,改进的clip算法实现了严格意义上的视频无再损编码。图4显示了不同量化因子下使用改进clip算法与原始clip算法在编码300帧时的平均亮度峰值性噪比(PSNR)。从表1与图4中不难看出,随着PSNR的提高,视频畸变逐渐减小,严格无再损视频编码对应的PSNR为100。

图5画出了量化因子为34时不同算法下测试序列平均亮度峰值性噪比与对应的视频再编码次数的曲线。从图5(a)可以看出,采用现有算法时,第一次再编码平均亮度峰值性噪比下降最明显,为0.6db。随后峰值性噪比下降变缓,并且随着再编码次数的增加而逐渐达到稳定(类似MPEG多次再编码中的“饱和”特性<sup>[2]</sup>)。类似现象也出现在图5(b)与5(c)中。而采用改进的无再损帧内编码算法时,亮度峰值性噪比一直保持为100,并且不随再编码次数的增加而变化。

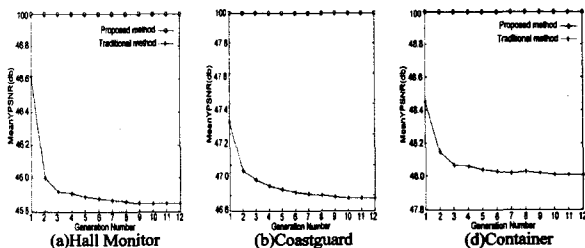


图5 亮度PSNR与视频再编码次数曲线

结束语 本文提出了新的基于统筹学理论的H.264视频严格无再损优化算法。实验结果证明这种方法在数据溢出时可以有效消除由原有clip算法导致的视频畸变,并且与改进的预测算法结合能实现严格无再损H.264视频帧内编码。今后的研究方向为如何在开启环路滤波器的情况下实现严格无再损H.264视频帧内编码。

### 参考文献

[1] Horne C, Naveen T, Tabatabai A, et al. Study of the Characteristics of the MPEG2 4: 2: 2 Profile-Application of MPEG2 in Studio Environment[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1996, 6(3): 251-272

[2] Erdem A T, Sezan M I. Multi-generation Characteristics of the MPEG Video Compression Standards[C]// IEEE Proceedings International Conference on Image Processing. vol. 2, 1994: 933-937

[3] Sorial H, Lynch W. Multigeneration of transform coded images[J]. Proceedings of SPIE, the International Society for Optical Engineering, 1997, 3024(2): 1394-1405

[4] Hda L, Rosilesb J. Achieving idempotence in near-lossless JPEG-LS[J]. Image and Video Communications and Processing, 2000, 3974: 685-690

[5] Joshi R, Rabbani M, Lepley M. Comparison of multiple compression cycle performance for JPEG and JPEG2000[C]// Proceedings of SPIE, the International Society for Optical Engineering. vol. 4115, 2000: 492-501

[6] Lin T. Achieving Re-Loss-Free Compression in H.264 Like Hybrid Video Coding[C]// IEEE Proceedings of 2008 International Conference on Signals, Circuits & Systems. Hammamet, Tunisia, 2008: 1-6

[7] Lin T. Achieving Re-Loss-Free Video Coding[C]// IEEE Transactions on Signal Processing Letters (to appear)

[8] ITU-T. Advanced video coding for generic audiovisual services [OL]. <http://www.itu.int/rec/T-REC-H.264-200711-I/en>, Nov. 2007

[9] <http://trace.eas.asu.edu/yuv/index.html>

[10] Vanderbei R J. Linear Programming: Foundations and Extensions. second edition. Springer, May 2001

[11] Videolan.org. X264code[OL]. <ftp://ftp.videolan.org/pub/videolan/x264/snapshots/>

[12] Suehring K. H.264/AVC Software Coordination[OL]. <http://iphome.hhi.de/suehring/tml/>