

# 面向应用层的遥感数据分块技术及其应用

李新春, 木红旭, 叶发旺, 邱骏挺

核工业北京地质研究院遥感信息与图像分析技术国家级重点实验室, 北京

收稿日期: 2023年6月18日; 录用日期: 2023年7月17日; 发布日期: 2023年7月26日

## 摘要

遥感是一种重要的地表观测技术, 在许多领域有广泛应用。随着技术的不断进步, 遥感技术已从传统低空间分辨率、多光谱向高空间分辨率、高光谱方向发展, 使得遥感数据的体积过于庞大而无法一次性全部载入计算机内存进行处理。这种情况下, 开展数据分块处理就显得尤为重要。当前针对遥感数据分块的模块主要在程序层级整合, 且多数分块模块是商用的, 这对应用层级的软件开发与集成并不友好。针对这一问题, 本文提出一种面向应用层级的分块机制, 并利用IDL语言开发了针对遥感大数据文件自动分块和结果自动整合两个过程。在此基础上, 以新疆吉木萨尔地区的航空高光谱条带数据为数据源, 开展了分块、处理、整合实验, 最终证明该机制和相关过程的可行性与实用性, 不仅为后续开发应用层级的遥感数据处理软件奠定基础, 也可开展数据并行处理和大数据网络传输提供帮助。

## 关键词

遥感大数据, 数据分块, IDL, 编程, 集成软件

# Application Level Oriented Remote Sensing Data Segmentation Technology and Its Application

Xinchun Li, Hongxu Mu, Fawang Ye, Junting Qiu

National Key Laboratory of Science and Technology on Remote Sensing Information and Image Analysis, Beijing  
Research Institute of Uranium Geology, Beijing

Received: Jun. 18<sup>th</sup>, 2023; accepted: Jul. 17<sup>th</sup>, 2023; published: Jul. 26<sup>th</sup>, 2023

## Abstract

Remote sensing is an important surface observation technology that is widely used in various fields. With the development of the technology, the data of remote sensing has evolved from traditional low

spatial and multispectral resolutions to high spatial and hyperspectral resolutions, which has made the volume of the data too large to be fully loaded into computer memory for processing at once. In this situation, the data segmentation is particularly important to data processing. Presently, modules for remote sensing data segmentation are mainly integrated during the programming process, and most of the modules are commercial, which constrains software development and integration at the application level. This article presents a new segmentation mechanism to address the issue of application level data segmentation that has not been achieved by previous segmentation modules. Two procedures, including the big data automatic segmentation procedure and the result automatic integration procedure, are developed using IDL language. The two procedures were used to do data segmentation experiments based on the airborne hyperspectral data of Jimusaer area, and the results demonstrate the feasibility and practicality of the mechanism and related procedures. The achievements in this study not only contribute to the further development of the application-level integration software systems for remote sensing processing, but also help to carry out data parallel processing and big data network transmission.

## Keywords

Remote Sensing Big Data, Data Segmentation, IDL, Programming, Integrated Software

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

遥感技术具有空间覆盖范围广、数据获取速度快、信息丰富、价格便宜、无侵入等特点,在开展大区域尺度,特别是人力难以抵达地区,地表观测方面具有得天独厚的技术优势。因而在地质、环境、农业、军事等多个领域都有广泛应用[1]-[10]。近些年来,随着遥感技术的不断发展,遥感数据已从传统低空间分辨率、多光谱向高空间分辨率、高光谱方向转变[11][12][13]。

由于技术的进步,当前遥感数据所蕴含的信息比以往更为丰富,而与之匹配的数据量也不断增加。以航空高光谱 SASI 数据为例,由于其包含上百个波段,获取单条带的数据量通常在几个 Gb 至十几个 Gb,而对长航带甚至可达几十个 Gb。面对如此庞大的数据量,很多计算机内存无法一次性加载完整的数据,因为这会导致溢出而使计算机系统崩溃。此外,过大的数据量也会给网络传输和在线处理带来困难。

针对数据量过大的问题,前人提出了数据分块处理技术[14][15]。该技术的核心思想是将原先过大的数据拆分为一系列小数据块,保证每一个小数据块加载后不会发生溢出错误,再将小数据块依次加载到内存中进行处理,或依次通过网络进行传输,最后将处理或传输结果进行整合。

虽然数据分块技术在当前遥感数据处理领域已较为成熟,如商业软件 ENVI 就引入了该分块机制。但这种机制和相应的模块仅作用于程序层级,即需要在编程过程中引用和调取,这对单一编程语言开发的软件较为适用,因为不同过程间参数传递是通畅的,内存共享自由;但不适于应用层级的集成软件开发,因为集成软件各模块的开发语言和规范可能不同,一些模块还可能受版权、保密因素制约,无法从程序级别进行调取。不仅如此,由于 ENVI 是商业软件,将分块功能写入自主开发的软件时需额外支付费用,这不利于软件的开发。

因此,本研究提出了一种应用层级的分块机制,并利用 IDL 语言自主开发了超大文件分块功能,实现了对遥感大数据应用层级的分块,最后利用该技术对吉木萨尔地区的航空高光谱条带数据进行了处理

实验, 获得了良好的处理效果。

## 2. 运行环境调研

为了合理选择开发语言、掌握当前遥感数据的处理量级和数据处理使用的操作系统以及硬件设备的情况, 本研究开展了相关调研工作。调研对象主要为在校大学生(约 26 人)和青年遥感工作者(11 人), 这两部分人群为当前主要的遥感数据处理人员。此外还包括工作 3~5 年的职工(3 人)。共计 40 人。

### 2.1. 操作系统

共调研了 6 种操作系统, 分别为 Windows XP、Windows 7、Windows 8、Windows 10、Linux、Mac OS (图 1)。其中有近 80% 的人选择 Windows 系统, 而使用 Windows 10 的用户达到了全部调研人员的一半。主要因为 Windows 系统的稳定性以及与大部分软件完美的兼容性。只有 9 个人选择使用了 MacOS 和 Linux 系统。另外, 当前有 55% 的人使用 64 位操作系统, 45% 的人使用 32 位操作系统。

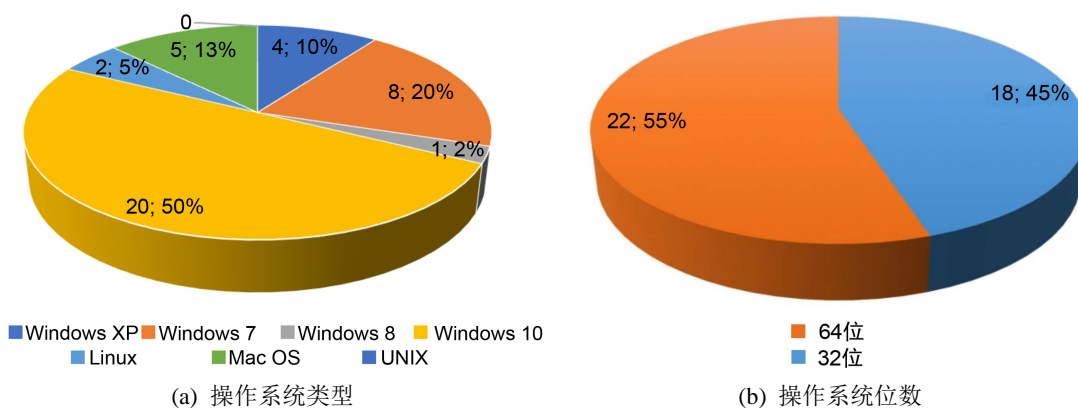


Figure 1. Results of questionnaire on operating system requirements  
图 1. 操作系统需求问卷结果

### 2.2. 硬件设备

硬件设备方面主要涉及 CPU 性能、硬盘容量、内存大小。目前对 CPU 的频率需求主要为大于 2.5 GHz, 8 线程为主(图 2)。硬盘容量大小以 2 T 和 4 T 需求为主(图 3(a))。内存大小的需求主要集中在 8~16 G, 其次为 4~8 G (图 3(b)), 目前小于 4 G 的内存使用已较少。

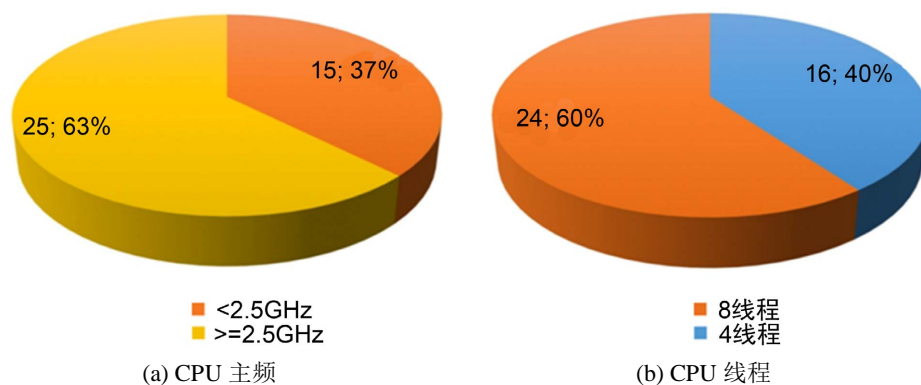


Figure 2. Results of questionnaire on CPU performance  
图 2. CPU 性能问卷结果

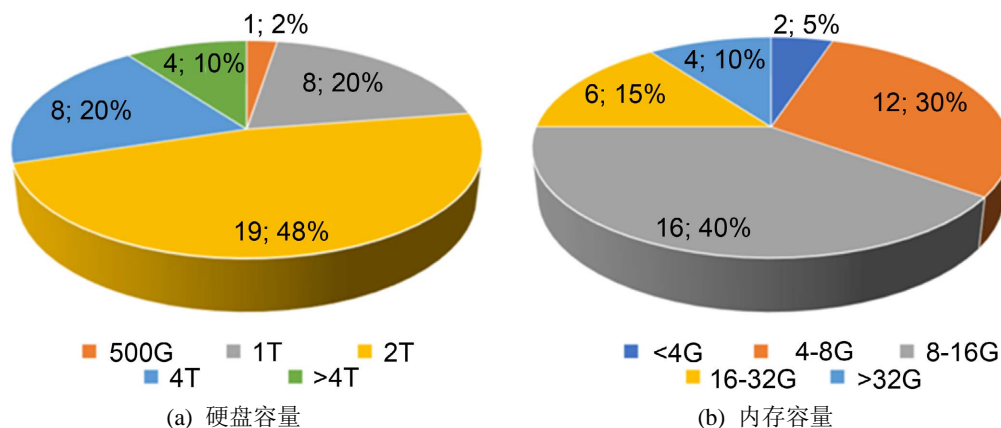


Figure 3. Results of questionnaire survey on hard disk and memory capacity requirements

图 3. 硬盘和内存容量需求问卷调查结果

### 2.3. 单景数据处理量

当前，国内遥感数据主要以国产高分系列卫星、资源系列卫星为主，国外遥感数据主要以 Landsat、Sentinel、ASTER 系列卫星数据为主。其他商业高分卫星也有使用。这些遥感数据大多在 1 Gb 空间左右，部分超过 1 Gb。近些年来，随着航空高光谱技术的发展，对数据量的需求提升巨大。以至于单景文件空间容量超过 1 Gb 变为普遍现象。根据图 4 所示，当前主要的遥感数据空间占用量介于 1~10 G，航空高光谱数据普遍在 10~50 Gb，部分超过 50 Gb。

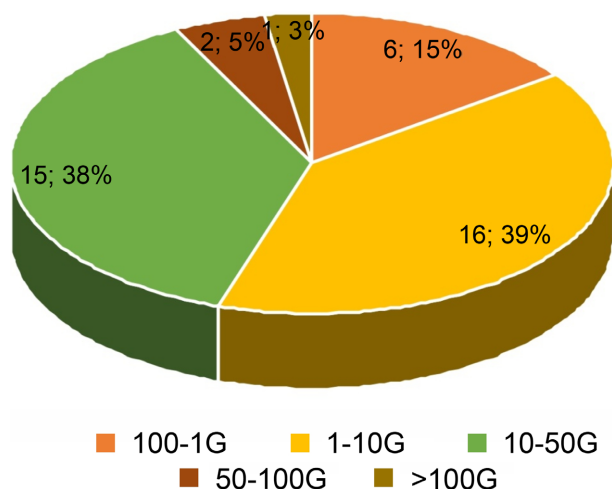


Figure 4. Results of questionnaire survey on processing data volume

图 4. 处理数据量问卷调查结果

## 3. 分块机制和程序设计

### 3.1. 分块机制设计

根据需求调研，当前遥感数据的空间占用量多为 1~50 Gb (图 4)，以核工业北京地质研究院遥感重点实验室 2013 年 10 月 16 日获取的 SASI\_3001 数据文件为例，该文件大小为 13.6 GB。另具调研报告显示，当前计算机的内存多在 8~16 Gb。可见，对于一些计算机，很多遥感数据无法一次性全部载入到内存中进行处理，必须进行数据分块。

考虑到非商用遥感数据处理模块可能使用不同语言、不同编程规范以及可能受版权、保密等多种因素制约，因此，只能在应用层进行整合。为此，设计了如图 5 所示的分块处理机制。该机制首先利用一个分块程序将超大文件拆分为若干小文件，并针对每一个小文件形成一个空间信息文件，指定小文件的维度、数据类型等信息；之后，外部的功能模块程序可直接对小文件进行处理，得到每一个对应小文件的处理结果；最后整合程序将处理结果按照之前记录的空间信息进行恢复，形成一个完整的处理结果。

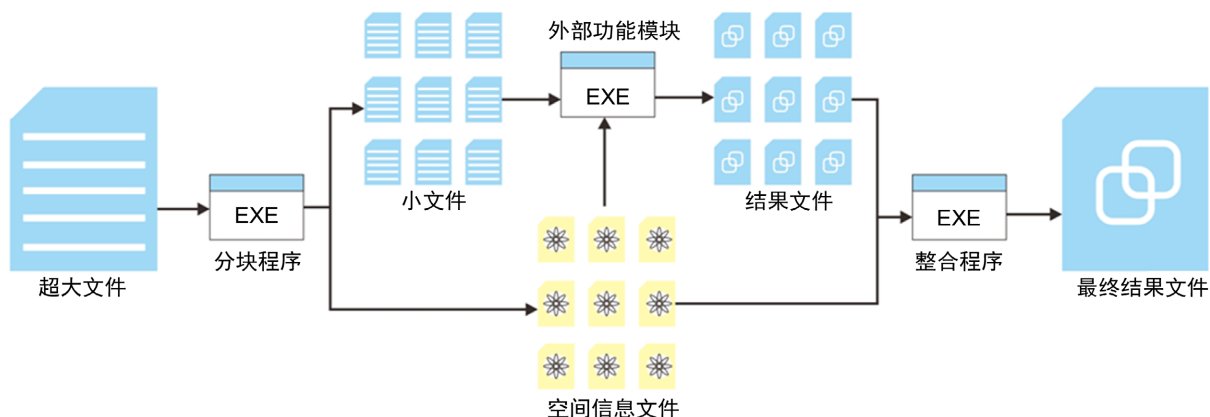


Figure 5. Schematic diagram of application layer block processing mechanism

图 5. 应用层分块处理机制示意图

### 3.2. 分块程序设计

设计的分块程序执行流程如图 6 所示，现将每个流程介绍如下：

第一步，读取遥感影像的头文件信息，通过头文件信息可以获得遥感影像文件行列数、波段数、数据类型(整形、长整形、浮点型、双精度浮点型等)；

第二步，利用行列数、波段数和数据类型计算文件的总大小。文件总大小 = 行数 × 列数 × 波段数 × 数据类型占用字节数；

第三步，计算最小分割单元的大小。为了保证图像分块图像的完整性，需要设定最小的分割单元。本研究以遥感图像的一行像元作为最小分割单元，则最小分割单元的大小 = 1 × 列数 × 波段数 × 数据类型占用字节数；

第四步，计算每一个分块文件可以包含的最小分割单元的数量。每一个分块文件中最小分割单元的数量 = INT 向下取整(分块文件的容量限制/最小分割单元的大小)；

第五步，计算分块数。分块数 = 文件总大小/(每一个分块文件包含的最小分割单元的数量 × 最小分割单元的大小)，如果分块数为整数，则令分块数 = 分块数，若有小数部分则令分块数 = 分块数 + 1；

第六步，计算每一个分块的大小。如果第五步中分块数为整数，则分块的大小 = 每一个分块文件包含的最小分割单元的数量 × 最小分割单元的大小，如果不为整数则整数部分的分块大小 = 每一个分块文件包含的最小分割单元的数量 × 最小分割单元的大小，最后一个分块的大小 = 文件总大小 - (最小分割单元的数量 × 最小分割单元的大小)；

第七步，使用二进制读写函数打开遥感影像；

第八步，按照分块大小从遥感影像中抽提数据；

第九步，使用写函数将抽提的数据写入到硬盘中，并创立空间信息文件；

第十步，反复执行第八步和第九步的操作，直到所有的分块都被执行完后，使用关闭函数关闭所有文件。

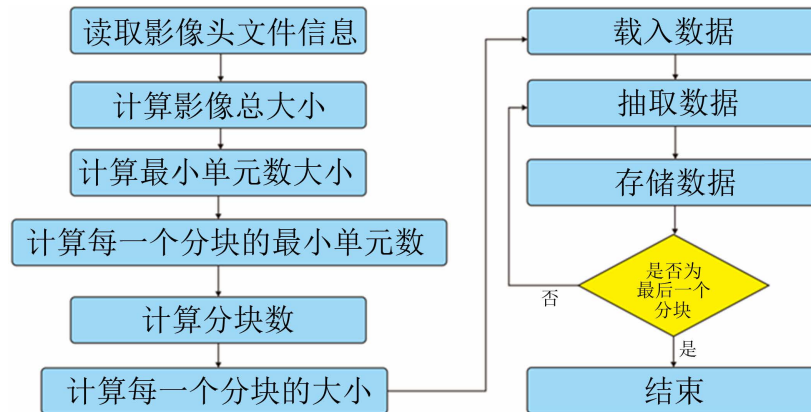


Figure 6. Diagram of data segmentation procedure  
图 6. 数据分块流程图

### 3.3. 整合程序设计

使用分块程序实现了对超大数据的分割保证了内存中的数据不发生溢出。但随之而来的问题是：分块后每一个文件都彼此独立，后续的处理过程都是针对单一分块开展的，其处理结果也只能覆盖每一个分块文件的范围。

为了能够获取整幅遥感影像的处理结果，设计了结果整合程序，该程序的工作流程如图 7 所示，整个程序的执行步骤介绍如下：

第一步，获取结果文件列表，计算结果文件的数量；

第二步，计算建立两个临时变量 TYS 和 TIMG；

第三步，载入一个结果文件，如果为第一次载入文件，则将结果文件的大小保存在 TYS 中；将结果文件的数据保存在 TIMG 中；如果不是第一次载入文件，则执行  $TYS = TYS + \text{文件大小}$ ，同时将文件数据追加到 TIMG 之后；

第四步，判断循环是否结束，如未结束则跳至第三步继续执行，如结束则令  $YS = TYS$ ， $OUTIMG = TIMG$ ，并输出 OUTIMG 作为最终结果。

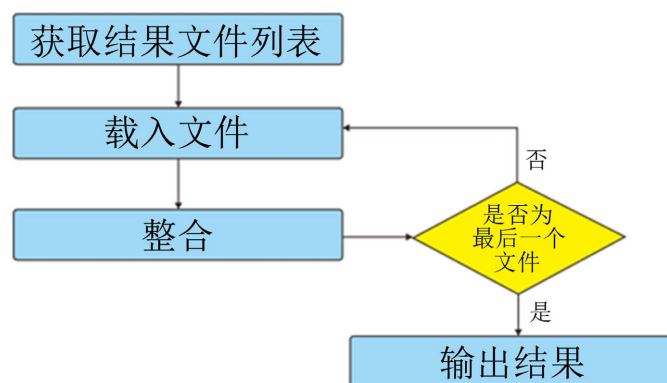


Figure 7. Diagram of result combination procedure  
图 7. 结果整合流程图

## 4. 分块和整合程序开发

数据分块和整合程序利用 IDL (Interactive Data Language) 语言实现。该语言是一种数据分析和图像化

应用程序及编程语言。语言集开放性、高维分析能力、科学计算能力、实用性和可视化分析为一体，可在多种硬件和操作系统平台上运行，并内置数学库函数可大大地减少图像处理算法开发的工作量。

本次程序开发，主要使用了 IDL 语言的 OPENR 和 CLOSE 两个函数来操作文件的打开和关闭，利用 READU、WRITEU 两个函数进行二进制文件的读写，利用 PRINTF、READF 两个函数进行文本文件的读写，同时利用数组运算符 ARRAY=[[ARRAY1], [ARRAY2]]完成数据的追加。通过调用上述函数，连同其它语句，共同封装于两个过程中，名称分别为：IDL\_CREATE\_TILES 和 IDL\_COMBINE\_TILES，其中前者负责完成分块，后者负责将分块结果整合，两个过程的形式如下：

```
PRO IDL_CREATE_TILES,INFILE,OUTDIR,TILE_LIST=TILE_LIST,SIZE_LIMITE
```

```
PRO IDL_COMBINE_TILES, TILE_LIST, OUTFILE
```

其中参数的含义为：

INFILE-超大文件的存储路径；

OUTDIR-分块文件的临时存储位置；

TILE\_LIST-分块文件的名称列表；

SIZE\_LIMITE-分块文件的大小限制；

OUTFILE-最终结果输出文件的路径。

IDL\_CREATE\_TILES 过程主要包含一个主控 FOR 循环(进行循环分块)、一个文件指针移位器 OFFSET(用来移位到每一个分块的起始数据位置)、一个输入函数(READU 负责从 unit1 中读二进制文件)、一个输出函数(WRITEU 负责向 unit2 中写二进制文件)和一个文本输出函数(PRINTF 负责向 unit3 中写文本)。主要的代码如下：

```
...
BAND_BYTE=XS*YS*DATA_BYTE;计算一个波段的字节数
LINE_BYTE=XS*BANDSUM*DATA_BYTE;计算图像中每一行的点的全部光谱所占用的字节数
...
TILE_LINE_NUM=SIZE_LIMITE/LINE_BYTE;计算每一个分块最多有多少行
TILE_BAND_BYTE=XS*TILE_LINE_NUM*DATA_BYTE;计算一个分块的一个波段的字节数
TILE_NUM=YS/TILE_LINE_NUM;计算分块数
...
make_custom_array,xs,TILE_LINE_NUM,bandsum,data_type,array=tile_array;创建一个用来保存数字的数组
...
openr,unit1, INFILE,/get_lun;打开待分块的文件
FOR i=1,TILE_NUM+ADDITION_TILE DO BEGIN;第一个循环负责建立分块文件
    openw,unit2,OUTDIR+"\ "+file_basename(INFILE)+".tile"+STRTRIM(string(i),1),/get_lun;建立一个分块文件，unit2 是分块文件的逻辑号
    openw,unit3,OUTDIR+"\ "+file_basename(INFILE)+".tile"+STRTRIM(string(i),1)+".hdr",/get_lun;建立一个分块文件头文件，unit3 是分块文件的逻辑号
    IF i EQ 1 THEN BEGIN
        TILE_LIST=OUTDIR+"\ "+file_basename(INFILE)+".tile"+STRTRIM(string(i),1)
    ENDIF ELSE BEGIN
        TILE_LIST=[TILE_LIST,OUTDIR+"\ "+file_basename(INFILE)+".tile"+STRTRIM(string(i),1)]
    
```

```

ENDELSE
FOR j=0,BANDSUM-1 DO BEGIN;第二个循环负责从大的数据中逐个波段的读取数据，再放入到
新的分块文件中
    point_lun,unit1,OFFSET+ulong64(BAND_BYTE)*j+(i-1)*ulong64(TILE_BAND_BYTE);移动偏离
值
    IF i LE TILE_NUM THEN BEGIN
        readu,unit1,tile_band_array
        tile_array[*,*j]=tile_band_array
    ENDIF ELSE BEGIN
        readu,unit1,addition_band_array
        addition_array[*,*j]=addition_band_array
    ENDELSE
ENDFOR
...
writeu,unit2,tile_array;写入分块内容
printf,unit3, header_info;写入头文件
...
CLOSE,unit2 & FREE_LUN, unit2;关闭分块文件
CLOSE,unit3 & FREE_LUN, unit3;关闭头文件
ENDFOR
CLOSE,unit1 & FREE_LUN, unit1;关闭原始文件
...

```

IDL\_COMBINE\_TILES 过程主要包含一个主控 FOR 循环(进行循环分块)、一个临时存储数组 (ARRAY)、一个文本输入函数(READF)、以及一个输出函数(WRITEU)。

调用 IDL\_CREATE\_TILES 过程时，指定超大文件在计算机上的存储路径 INFILE，设置好分块文件的临时存储位置 OUTDIR 以及每一个分块的限制大小 SIZE\_LIMITE，程序运行后会自动返回一个分块文件的列表 TILE\_LIST。

调用 IDL\_COMBINE\_TILES 过程时，只需将 TILE\_LIST 传入过程，并设置好最终结果的输出路径 OUTFILE，再执行过程就能获得最终的结果文件。

## 5. 应用实验

为了验证分块机制的有效性，本研究以核工业北京地质研究院遥感重点实验室采集的吉木萨尔地区 SASI\_3002 号航带的航空高光谱文件为例，对本文提出的分块机制进行了测试研究。

测试时，首先执行 IDL\_CREATE\_TILES 过程，设置每个分块的大小不超过 512 Mb。由于 SASI\_3002 号文件的大小为 4 Gb，因此最终分块数为 8。分块前与分块后的遥感图像如图 8 所示。

之后，调用核工业北京地质研究院自主开发的岩心高光谱蚀变矿物信息自动提取软件，采用光谱角匹配方法(SAM)，对分块结果进行处理。由于岩心数据相比航空高光谱数据小，因此该软件在设计之初并未采取分块机制，导致该软件无法直接对航空高光谱数据进行处理。但经分块后，原先的大文件转变为多个小文件，使得软件可以进行批量信息的提取。本轮测试主要针对高岭石蚀变，提取结果如图 9 所示。



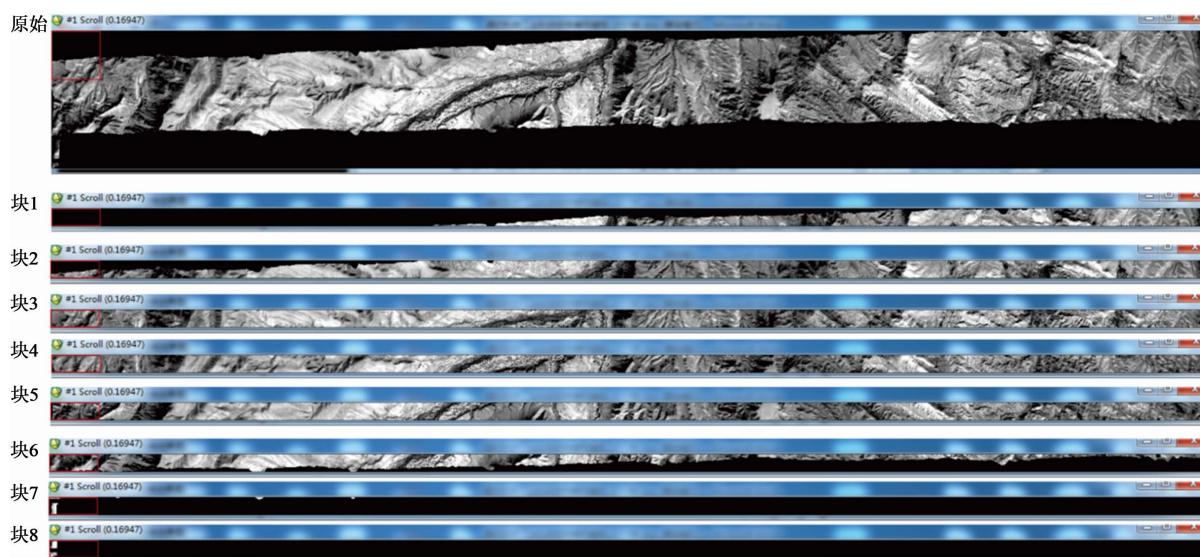
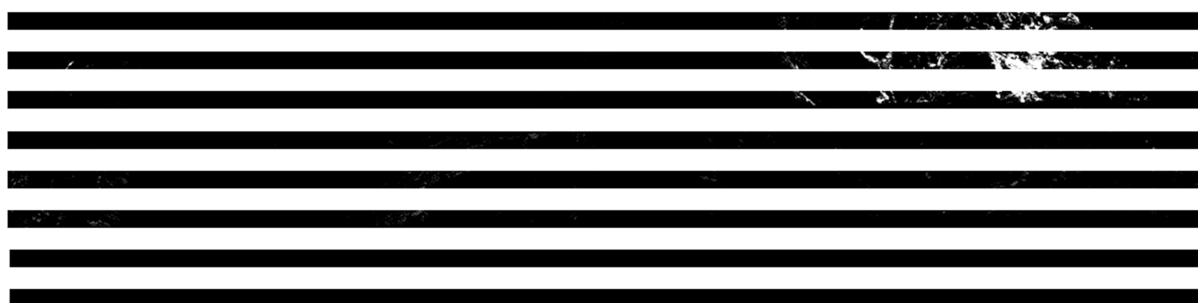


Figure 8. Data segmentation for SASI\_3002 file of Jimusaer region

图 8. 吉木萨尔地区 SASI\_3002 号航带数据分块情况



(白色部分示高岭石)

Figure 9. Kaolinite alteration extracted from segmentation data of SASI\_3002 file of Jimusaer region

图 9. 吉木萨尔地区 SASI\_3002 号航带分块结果提取的高岭石分布图

最后，执行 IDL\_CREATE\_TILES 过程，将提取结果进行整合，最终得到完整的高岭石蚀变提取结果(图 10)。



(白色部分示高岭石)

Figure 10. Kaolinite alteration result for SASI\_3002 file of Jimusaer region

图 10. 吉木萨尔地区 SASI\_3002 号航带高岭石提取结果图

此外，利用 ENVI 软件的光谱角匹配算法，选用相同的输入参数进行高岭石蚀变提取。最终获得的结果与图 10 结果一致，说明采用本分块机制的确可以从应用层级实现遥感数据的分块处理，其最终处理结果与程序级分块获得的结果一致。相比 ENVI，本程序为免费程序，可打包为独立 exe 运行，仅需提供 3 个输入参数就能完成分块，且无需调用 ENVI 环境和进行代码引用。这种特点为后续开展应用层级别的集成软件系统开发奠定了基础，也可为并行处理和大数据网络传输提供便利的前处理工具。

## 6. 结论

本文针对大体积遥感数据无法直接载入内存进行处理的问题,提出了一种应用层级别的数据分块处理机制,并采用IDL语言编写了IDL\_CREATE\_TILES和IDL\_COMBINE\_TILES两个过程,通过调用第一个过程将超大文件拆分成若干小文件,之后调用处理模块进行批处理获得每一个小文件对应的处理结果,再调用第二个过程将处理结果整合为最终结果。通过开展应用测试,证明利用IDL语言可以实现应用级别的数据分块,从而为遥感大数据集成软件系统的开发奠定了基础。

## 基金项目

本研究受国防科工局核能开发项目(YH2001-5)资助。

## 参考文献

- [1] 木红旭, 马天碧, 郭强, 邱骏挺, 芮歆旻. 松辽盆地南部局部排泄构造遥感识别与铀找矿方向探讨[J]. 地质与勘探, 2023, 59(1): 134-144.
- [2] 刘德长, 邱骏挺, 闫柏琨, 田丰. 高光谱热红外遥感技术在地质找矿中的应用[J]. 地质论评, 2018, 64(5): 1190-1200.
- [3] 刘德长, 田丰, 邱骏挺, 叶发旺, 闫珀琨, 孙雨, 王子涛. 柳园-方山口地区航空高光谱遥感固体矿产探测及找矿效果[J]. 地质学报, 2017, 91(12): 2781-2795.
- [4] 刘德长, 赵英俊, 叶发旺, 田丰, 邱骏挺. 航空高光谱遥感区域成矿背景研究——以甘肃柳园-方山口地区为例[J]. 遥感学报, 2017, 21(1): 136-148.
- [5] 刘德长, 闫柏琨, 邱骏挺. 航空高光谱遥感固体矿产预测方法与示范应用[J]. 地球学报, 2016, 37(3): 349-358.
- [6] 刘德长, 叶发旺, 赵英俊, 田丰, 邱骏挺. 航空高光谱遥感金矿床定位模型及找矿应用——以甘肃北山柳园-方山口地区为例[J]. 地球信息科学学报, 2015, 17(12): 1545-1553.
- [7] 刘德长, 邱骏挺, 童勤龙, 谢涛, 杨燕杰, 王子涛. 新疆吉木萨尔地区油气渗漏与盐碱化相关性的航空高光谱遥感探测研究[J]. 中国地质, 2018, 45(5): 1062-1073.
- [8] 高明琪, 肖龙龙, 王煜博. 星载高光谱成像仪的发展及军事应用[J]. 红外, 2022, 43(3): 22-30+52.
- [9] 孟曦婷, 计璐艳, 赵永超, 杨炜曦. 基于深度学习的多尺度导弹发射井目标检测[J]. 中国科学院大学学报, 2021, 38(6): 800-808.
- [10] 孙博学, 杨景嘉, 付若曦, 马玥. 农作物种植结构遥感识别研究进展[J]. 科技创新与应用, 2023, 13(15): 76-79+84.
- [11] 叶发旺, 张川, 李瀚波, 余长发, 刘洪成, 孟树, 邱骏挺, 童勤龙, 田丰, 秦凯, 武鼎, 郭帮杰, 裴承凯, 朱黎江, 车永飞, 任梦如, 田青林. “空天”高分辨率遥感技术及其在铀资源勘查中的应用进展与发展建议[J]. 铀矿地质, 2021, 37(3): 313-328.
- [12] 苏红军. 高光谱遥感影像降维: 进展、挑战与展望[J]. 遥感学报, 2022, 26(8): 1504-1529.
- [13] 李华玉, 陈永富, 陈巧, 王娟, 张超. 基于遥感技术的森林树种识别研究进展[J]. 西北林学院学报, 2021, 36(6): 220-229.
- [14] 丁翠, 杨云涛, 张建. 遥感影像并行分割的无缝拼接算法[J]. 测绘科学, 2017, 42(4): 34-38.
- [15] 吴尚竹, 程承旗. 一种遥感数据自适应剖分分块渐进传输方法[J]. 地理信息世界, 2016, 23(2): 1-8.