



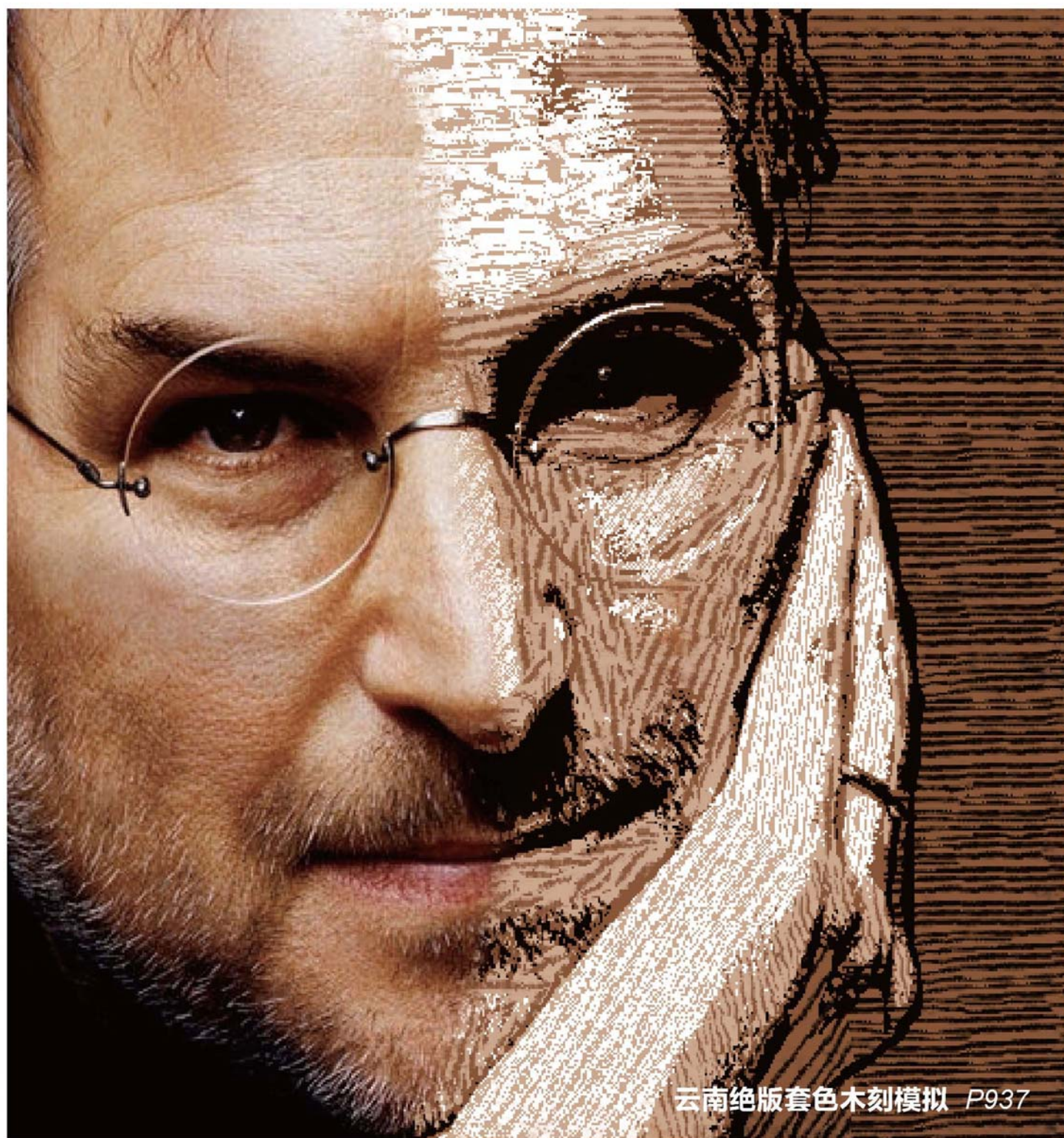
JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2015
07
VOL.20

ISSN1006-8961
CN11-3758/TB



云南绝版套色木刻模拟 P937



第20卷第7期（总第230期）

2015年7月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院遥感与数字地球研究所

中国图象图形学学会

北京应用物理与计算数学研究所

主 编 李小文

编辑出版《中国图象图形学报》编辑出版委员会

邮政信箱 北京9718信箱

邮 编 100101

电子信箱 jig@radi.ac.cn

电 话 010-64807995

网 址 www.cjig.cn

广告经营许可证 京朝工商广字第0361号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

（邮政信箱：北京399信箱 邮编：100048）

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Institute of Remote Sensing and Digital Earth, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief LI Xiaowen

Editor, Publisher Editorial and Publishing Board of Journal of

Image and Graphics

P.O.Box 9718, Beijing, P.R.China

Zip code 100101

E-mail jig@radi.ac.cn

Telephone 010-64807995

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation

(P.O.Box 399, Beijing 100048, P.R.China))

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

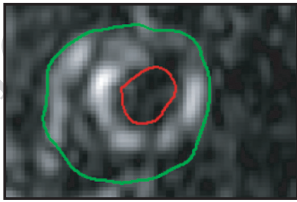
ISSN 1006-8961

CODEN ZTTXFZ

国外发行代号 M1406

国内邮发代号 82-831

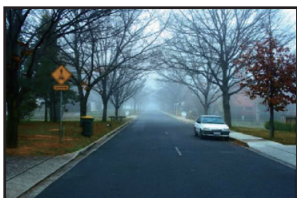
国内定价 50.00元



基于颈动脉粥样硬化斑块的多
序列MR图像去噪与配准
(第0871页)



光滑拟合模型的IIPBDR图像
修复算法(第0895页)



基于暗原色先验的图像快速去
雾(第0914页)



个性化定制的虚拟健身系统设
计与实现(第0953页)

图像处理和编码

融合视觉感知特性的HEVC率失真优化

费马燕, 彭宗举, 李持航, 陈芬, 郁梅, 蒋刚毅 0857

结合空间信息的模糊聚类侧扫声纳图像分割

李阳, 庞永杰, 盛明伟 0865

基于颈动脉粥样硬化斑块的多序列MR图像去噪与配准

李军伟, 汪晓妍, 张剑华, 何俊丽, 管秋, 陈胜勇 0871

小波变换与纹理合成相结合的图像修复

张东, 唐向宏, 张少鹏, 黄俊泽 0882

光滑拟合模型的IIPBDR图像修复算法

鱼江, 翟东海, 聂洪玉, 李同亮 0895

图像分析和识别

噪声方差未知条件下的视频目标跟踪

陈金广, 任冰青, 马丽丽, 温静 0906

基于暗原色先验的图像快速去雾

曾浩, 尚媛园, 丁辉, 周修庄, 付小雁 0914

人体运动分析的实例学习方法

李毅, 刘兴川, 孙亭 0922

图像理解和计算机视觉

多参数加权的无缝纹理映射算法

刘彬, 陈向宁, 薛俊诗 0929

交互式云南绝版套色木刻数字化合成

王雪松, 李杰, 侯剑侠, 徐丹, 普园媛 0937

计算机图形学

基于像素的多边形等距区域子分算法

严志刚, 寿华好 0945

个性化定制的虚拟健身系统设计与实现

赵汉理, 孟庆如, 韩丽贞, 潘志庚 0953

遥感图像处理

LANDSAT-8卫星影像长条带数据处理

李亚林, 李山山, 冯钟葵, 张洪群, 陈勃 0963

MapReduce模式下高光谱图像端元提取算法加速

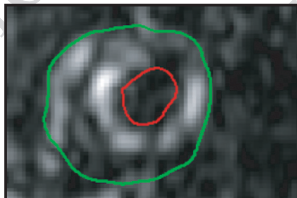
曲海成, 籍瑞庆, 刘万军, 梁雪剑 0973

基于GA-BP神经网络模型的登革热时空扩散模拟

李卫红, 陈业滨, 闻磊 0981

CONTENTS

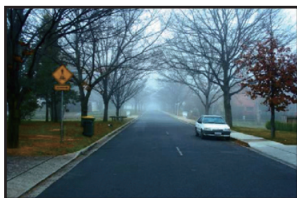
JOURNAL OF IMAGE AND GRAPHICS



Multiple sequence MR image de-noising and registration method based on carotid atherosclerosis plaque (P0871)



Improved IIPBDR algorithm based on smooth-fitting model (P0895)



Fast image haze removal base on dark channel prior (P0914)



Design and implementation of the personalized virtual fitness system(P0953)

Image Processing and Coding

- Human perception features based rate distortion optimization for HEVC
Fei Mayan, Peng Zongju, Li Chihang, Chen Fen, Yu Mei, Jiang Gangyi 0857
- Side-Scan sonar image segmentation via fuzzy clustering with spatial constrains
Li Yang, Pang Yongjie, Sheng Mingwei 0865
- Multiple sequence MR image de-noising and registration method based on carotid atherosclerosis plaque
Li Junwei, Wang Xiaoyan, Zhang Jianhua, He Junli, Guan Qiu, Chen Shengyong 0871
- Image inpainting based on combination of wavelet transform and texture synthesis
Zhang Dong, Tang Xianghong, Zhang Shaopeng, Huang Junze 0882
- Improved IIPBDR algorithm based on smooth-fitting model
Yu Jiang, Zhai Donghai, Nie Hongyu, Li Tongliang 0895

Image Analysis and Recognition

- Video target tracking algorithm with the noise variance unknown
Chen Jinguang, Ren Bingqing, Ma Lili, Wen Jing 0906
- Fast image haze removal base on dark channel prior
Zeng Hao, Shang Yuanyuan, Ding Hui, Zhou Xiuzhuang, Fu Xiaoyan 0914
- Example-based approach for human motion analysis from videos
Li Yi, Liu Xingchuan, Sun Ting 0922

Image Understanding and Computer Vision

- Seamless texture mapping algorithm based on multi parameter weighted
Liu Bin, Chen Xiangning, Xue Junshi 0929
- Interactive digital synthesis of Yunnan out-of-print woodcuts
Wang Xuesong, Li Jie , Hou Jianxia, Xu Dan, Pu Yuanyuan 0937

Computer Graphics

- Pixel based domain subdivision algorithm for polygon offsetting
YAN Zhigang, SHOU Huahao 0945
- Design and implementation of the personalized virtual fitness system
Zhao Hanli, Meng Qingru, Han Lizhen, Pan Zhigeng 0953

Remote Sensing Image Processing

- Study on processing method of LANDSAT-8 long strip satellite imagery
Li Yalin, Li Shanshan, Feng Zhongkui, Zhang Hongqun, Chen Bo 0963
- Acceleration of hyperspectral image endmember extraction based on MapReduce pattern
Qu Haicheng, Ji Ruiqing, Liu Wanjun, Liang Xuejian 0973
- Simulation of spatio-temporal diffusion of dengue fever based on the GA-BP neural network model
Li Weihong, Chen Yebin, Wen Lei 0981

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2015)07-0973-08

论文引用格式: Qu H C, Ji R Q, Liu W J, Liang X J. Acceleration of hyperspectral image endmember extraction based on MapReduce pattern[J]. Journal of Image and Graphics, 2015, 20(7): 0973-0980. [曲海成, 籍瑞庆, 刘万军, 梁雪剑. MapReduce 模式下高光谱图像端元提取算法加速[J]. 中国图象图形学报, 2015, 20(7): 0973-0980.] [DOI: 10.11834/jig.20150714]

MapReduce 模式下高光谱图像端元提取算法加速

曲海成^{1,2}, 籍瑞庆¹, 刘万军¹, 梁雪剑¹

1. 辽宁工程技术大学软件学院, 葫芦岛 125105; 2. 哈尔滨工业大学电子与信息工程学院, 哈尔滨 150001

摘要: **目的** 随着成像光谱仪的发展, 高光谱遥感图像的空间分辨率和光谱分辨率越来越大, 这给高光谱遥感图像解译处理及应用带来挑战。本文提出一种基于 MapReduce 模式的分布式混合并行处理模型来加速高光谱解混处理。**方法** 为降低算法计算复杂度, 对原串行算法进行并行化设计, 并采用行列式分块计算法对原算法进行化简计算; 最后在分布式集群环境下, 采用 Jama 和 JCuda 组件来加速算法执行过程中的矩阵运算操作。**结果** 针对 224 波段, 400 × 400 像素空间分辨率的高光谱图像, 采用分布式混合计算模型进行解混处理比原始的处理方法在速度上有近十倍的提高, 且算法计算量越大, 加速效果越明显。**结论** 本文提出了一种基于 MapReduce 模式的分布式混合并行处理方法来加速最大单纯形体体积端元提取算法, 加速效果明显; 采用分块法求解行列式可以大大降低算法复杂度。该方法对计算任务可并行划分、主机与节点间数据交换量少且计算复杂类算法加速效果明显。

关键词: 光谱解混; 分布式计算; 端元提取; 分块行列式

Acceleration of hyperspectral image endmember extraction based on MapReduce pattern

Qu Haicheng^{1,2}, Ji Ruiqing¹, Liu Wanjun¹, Liang Xuejian¹

1. College of Software, Liaoning Technical University, Huludao 125105, China;

2. School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China

Abstract: **Objective** With the development of imaging spectrometer, the spectral resolution and space resolution have been enhanced dramatically which makes a challenge to hyperspectral unmixing processing. So a new distributed hybrid parallel model has been proposed to accelerate hyperspectral unmixing processing. **Method** In order to reduce the computational complexity of endmember extraction algorithm, the original serial method has been redesigned for parallel computation and a fast implementation of improved method has been proposed based on partitioned determinant operations. At the same time, the Jama and JCuda components have been used to accelerate the computation in distributed cluster environment. **Result** The proposed distributed hybrid parallel method plays a large role in accelerating hyperspectral unmixing based on maximum simplex volume algorithm. The improved MapReduce model method is near ten times more rapid than the original method for the hyperspectral image which size is 400 × 400 × 224. And the more computational load, the more speed up. **Conclusion** In this paper, the proposed distributed hybrid parallel method can increase the hyperspectral unmixing processing speed dramatically. At the same time, the partitioned determinant solving method can reduce the complexity of MS-

收稿日期: 2015-03-13; 修回日期: 2015-05-09

基金项目: 国家自然科学基金项目(61172144); 国家高技术研究发展计划(863)基金项目(2012AA12A405)

第一作者简介: 曲海成(1981—), 男, 讲师, 哈尔滨工业大学信息与通信工程专业在读博士研究生, 主要研究方向遥感高性能计算技术。

E-mail: quhaicheng@lntu.edu.cn

Supported by: National Natural Science Foundation of China (61172144); National High Technology Research and Development Program of China (2012AA12A405)

VA algorithm. The experimental results indicate that the proposed method can achieve great speedup to the algorithms which have characters of parallel executive tasks、lower data transmission between main node and sub nodes and massive calculations.

Key words: spectral unmixing; distributed computation; endmember extraction; partitioned determinant

0 引言

高光谱遥感图像以其较高的光谱分辨率和空间分辨率以及“图谱合一”的特性在光谱解译领域发挥重要作用^[1]。光谱解混是高光谱图像解译中重要步骤,也是高光谱图像处理中关键环节,光谱解混好坏直接影响高光谱图像解译的效率和质量。然而随着传感器技术的发展,光谱分辨率越来越高,光谱波段数越来越多,获取到的数据立方体也越来越大^[2]。一景高光谱遥感影像存储大小可达几个 G,这给高光谱图像深度解译处理带来挑战。传统的处理手段已经远远满足不了实际需求,迫切需要新的快速处理算法以及特定处理硬件支持解混处理^[2]。国内外学者从算法改进以及硬件加速处理方面做了很多开拓性工作。解混算法改进方面,围绕端元数目确定、端元提取和丰度计算这 3 部分开展研究工作,主要是从算法本身入手,通过公式化简、修改迭代停机准则、减少搜索空间以及等价替换等手段来降低算法的时间或空间复杂度;硬件加速处理方面,主要研究集中在 DSP(digital signal processing)、FPGA(field programmable gate array)、GPU(graphics processing unit)或分布式集群或异构网等并行计算技术。由于端元提取算法是线性光谱解混的先行步骤,且端元提取精度在一定程度上决定了混合像元分解的精度,因此,国内外学者主要精力集中在端元提取算法的精度和实现效率上。近年来建立在凸面几何学理论基础之上的凸面单体的顶点作为端元的算法是目前应用最多、最有效的端元提取算法。N-FINDR(自动端元检测解混算法)、SGA(simplex growing algorithm)和 VCA(vertex component analysis algorithm)是这类算法的典型代表。然而实际应用过程中,这类算法存在初始端元不明确、搜索空间范围大、数据降维造成信息损失和单形体体积计算时间过长等问题。针对上述问题,Chang 等人^[3]利用虚拟维数确定端元数目,采用迭代误差分析算法的端元提取结果作为 N-FINDR 的初始端元,加快了算

法收敛速度且使端元提取结果更加稳定^[3];Xiong 等人^[4]采用感兴趣区域(ROI)、随机纯像元指数(RPPI)和矩阵分块技术来解决原始 N-FINDR 算法搜索空间大、计算结果不稳定及过量计算等问题;赵春晖等人^[5]提出改进 N-FINDR 算法停机准则和数据特征预处理方法,并使用支持向量机对提取到的端元进行二次提取,来提升数据可分性和端元提取精度。丁海勇等人^[6]利用卡方分布原理构造端元候选集合, N-FINDR 算法中作为搜索区域,从而提高了算法的实现效率。N-FINDR 及其改进算法都是基于降维后的影像数据进行的操作,数据经过降维必定损失细节信息,Chang 等人^[7]在提出 SGA 算法的基础上,又进一步提出了基于数据不降维的改进算法,并分析了二者的差异;Nascimento 等人^[8]提出 VCA 算法,以线性光谱混合模型的几何描述为基础,通过反复寻找正交向量并计算图像矩阵在正交向量上的投影距离来逐一提取端元,这一算法也是数据不降维的经典端元提取快速算法。在硬件加速方面,目前 PPI(pixel purity index algorithm)、N-FINDR、VCA 和 SGA 等算法都有了 FPGA 版本;而它们的 GPU 版本已经出现,目前仍在不断地改进中^[9]。在 N-FINDR 算法基础上发展起来的基于最大单形体体积的端元提取算法(MSVA)以其不降维,端元顺序提取,解混效果好等优点被广泛关注^[10]。然而,算法实现过程中需要大量计算高维空间单形体体积,且提取端元数目越多,计算量越大。随着高光谱影像存储体积增大,单机处理已显得无能为力。近几年发展起来的基于 MapReduce 模式的分布式并行计算技术为解决海量遥感影像并行处理提供新思路^[11];基于 GPU 的高性能计算技术,为提高单机处理效率提供可能^[12]。为此提出将二者结合起来,构建基于 MapReduce 和 GPU 的混合并行处理架构来加速高光谱解混算法实现。

1 MapReduce 混合并行处理模型

1.1 MapReduce 分布式编程模型

MapReduce^[11]是一种简化的分布式编程模式。

其名字源自于该模型中的两项核心操作:映射(Map)和化简(Reduce)。MapReduce 核心思想是把能自动分割执行的问题拆解成 Map 和 Reduce 执行的方式,在数据被分割后通过 Map 过程将数据映射成不同的区块,分配给计算机集群处理,达到分布式运算的效果;然后通过 Reduce 过程将各子节点计算结果汇总,从而在主控节点输出最终的处理结果。在处理之前,将数据集分布至各个节点;处理时,每个节点就近读取本地存储的数据,进行映射处理(Map),将处理后的数据进行合并(Combine)、排序(Shuffle and Sort)后在发送(至 Reduce 节点),从而避免了节点间大量数据传输,提高了处理效率。另外数据在 MapReduce 模式的各个阶段都是以键/值(key/value)作为输入和输出,整个 MapReduce 模式的分布式计算流程如图 1 所示。

1.2 GPU 加速算法实现

CUDA(compute unified device architecture)是显

卡厂商 NVIDIA 推出的运算平台。CUDA™ 是由 NVIDIA 推出的通用并行计算架构,该架构下可以调用 GPU 来解决复杂的计算问题。由于 MapReduce 程序运行在 Java 环境下,而 CUDA 环境下 GPU 程序采用类 C 语言实现,为实现不同语言环境下程序协同工作,采用 JCuda(Java bindings for CUDA)组件来做中间件。JCuda 是利用 Java 调用 CUDA 程序的相关组件库。

程序中通过使用 JCuda 中的 JCublas 库,实现对矩阵进行求逆运算,在运算过程中,JCublas 调用 GPU 做并行处理,从而高效地实现矩阵求逆的问题。另外,程序实现过程中采用 Jama(一个 Java 处理矩阵运算的相关数学库,其中包含很多矩阵运算的相关方法和函数)来加速串行算法实现过程中的矩阵运算,从而简化算法实现难度,提高执行效率。

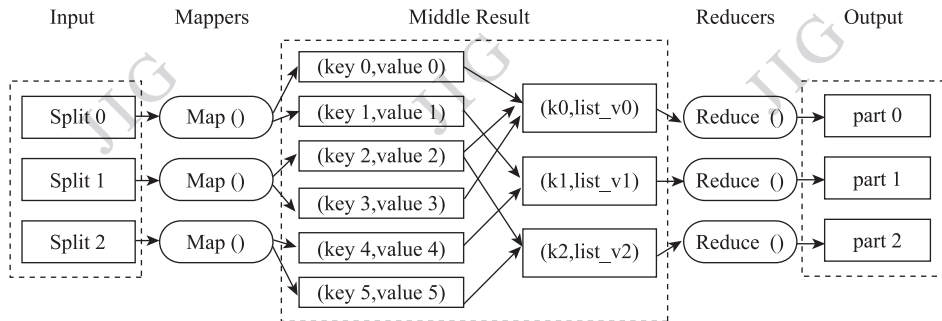


图 1 MapReduce 的工作流程图

Fig. 1 Flow chart of MapReduce processing

2 端元提取算法及其并行化实现

2.1 最大单形体体积法端元提取

最大单形体体积端元提取法^[10]是在经典的 N-FINDR 算法基础上发展起来的,该算法引入高维空间中的单形体体积公式,该公式与高光谱数据的维数无关,因此克服了 N-FINDR 算法受高光谱数据维数限制的固有缺陷,避免了由于数据降维操作而产生的信息损失,实现端元顺序提取。算法具体描述如下:

假设 m 维单形体内嵌于 m 维平行多面体之中,则从几何上看它们的体积之间存在对应关系

$$V_m = \frac{1}{m!} \bar{V}_m \quad (1)$$

式中, $\bar{V}_m$ 为平行多面体体积, V_m 为高维空间单形体体积。

在高维空间下平行多面体体积公式为

$$\bar{V}_m = V(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m) = \sqrt{\det(\mathbf{A}_m^T \mathbf{A}_m)} = \sqrt{G(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m)} \quad (2)$$

式中, $\mathbf{v}_i (i=1, 2, \dots, m)$ 为 m 维空间的列向量, $\mathbf{A}_m = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m]$ 为 m 行 m 列矩阵, $G(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m)$ 为克莱姆行列式。

$$G(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m) = \begin{vmatrix} \mathbf{v}_1^T \mathbf{v}_1 & \mathbf{v}_1^T \mathbf{v}_2 & \cdots & \mathbf{v}_1^T \mathbf{v}_m \\ \mathbf{v}_2^T \mathbf{v}_1 & \mathbf{v}_2^T \mathbf{v}_2 & \cdots & \mathbf{v}_2^T \mathbf{v}_m \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{v}_m^T \mathbf{v}_1 & \mathbf{v}_m^T \mathbf{v}_2 & \cdots & \mathbf{v}_m^T \mathbf{v}_m \end{vmatrix} \quad (3)$$

需要注意的是,上述单形体包含原点。

对高光谱图像而言,每个像元以散点的形式存在其特征空间中,图像中的每个像元映射为特征空间中的一个点,假设 $p_1, p_2, \dots, p_n$ 是高光谱图像中的 n 个独立像元(光谱向量),则令 $A_{n-1} = (p_n - p_1, p_n - p_2, \dots, p_n - p_{n-1})$,则在特征空间里以这 n 个点为顶点的 $n-1$ 维单形体体积为

$$V_n = V(p_n - p_1, p_n - p_2, \dots, p_n - p_{n-1}) = \frac{1}{(n-1)!} \sqrt{|A_{n-1}^T A_{n-1}|} \quad (4)$$

式中, $|\cdot|$ 表示行列式运算符。

2.2 改进最大单形体体积端元提取算法

由式(4)可以看出,算法实现过程中,每提取一个端元需要大量计算行列式值,随着端元提取数目的增加,计算量呈指数级增加。为此,考虑如下化简策略:

令

$$G_m(v_1, v_2, \dots, v_m) = \begin{vmatrix} v_1 v_1^T & v_1 v_2^T & \dots & v_1 v_m^T \\ v_2 v_1^T & v_2 v_2^T & \dots & v_2 v_m^T \\ \vdots & \vdots & \ddots & \vdots \\ v_m v_1^T & v_m v_2^T & \dots & v_m v_m^T \end{vmatrix} = \begin{bmatrix} A & C \\ D & B \end{bmatrix} \quad (5)$$

又令

$$G_{m-1}(v_1, v_2, \dots, v_{m-1}) = \begin{vmatrix} v_1 v_1^T & v_1 v_2^T & \dots & v_1 v_{m-1}^T \\ v_2 v_1^T & v_2 v_2^T & \dots & v_2 v_{m-1}^T \\ \vdots & \vdots & \ddots & \vdots \\ v_{m-1} v_1^T & v_{m-1} v_2^T & \dots & v_{m-1} v_{m-1}^T \end{vmatrix} = |A| \quad (6)$$

则

$$C = \begin{bmatrix} v_1 v_m^T \\ v_2 v_m^T \\ \vdots \\ v_{m-1} v_m^T \end{bmatrix} \quad D = \begin{bmatrix} v_m v_1^T \\ v_m v_2^T \\ \vdots \\ v_m v_{m-1}^T \end{bmatrix} \quad B = [v_m v_m^T]$$

$$G(V_1, V_2, \dots, V_m) = \begin{bmatrix} A & C \\ D & B \end{bmatrix} \text{ 成立。又当 } A \text{ 是 } n$$

阶可逆矩阵, B 是 m 阶方阵,有

$$\begin{vmatrix} A & C \\ D & B \end{vmatrix} = |A| \cdot |B - DA^{-1}C| \quad (7)$$

下面对式(7)证明。

因为 $A = G(V_1, V_2, \dots, V_{m-1})$ 表示 $m-1$ 维空间平行多面体体积的平方,若 $|A| = 0$,则表示空间所有点汇聚一点或一个平面,则这些点不满足凸集要求,此时对像元解混没有意义,因此有 $|A| > 0$,因此矩阵 A 可逆。又因为 B 为一阶方阵,对其进行初等变换

$$\begin{bmatrix} A & C \\ D & B \end{bmatrix} \xrightarrow{r_2 + (-DA^{-1})r_1} \begin{bmatrix} A & C \\ 0 & B - DA^{-1}C \end{bmatrix} \quad (8)$$

即将矩阵 A 和 C 分别乘以 $-DA^{-1}$,然后分别加到矩阵 D 和 B 上,可以得到变换矩阵

$$\begin{bmatrix} A & C \\ 0 & B - DA^{-1}C \end{bmatrix} \quad (9)$$

从而得到 $\begin{bmatrix} A & C \\ D & B \end{bmatrix} = |A| \cdot |B - DA^{-1}C|$,问题得证。

2.3 算法串行化实现

首先读入高光谱数据后,对原始数据进行预处理,去除水吸收和低信噪比波段数据,然后执行算法核心内容。算法核心主要对各个像元进行循环筛选,并计算单形体体积,以此来判断该像元点是否为待求端元。首先从高光谱图像的特征空间(光谱向量空间)中寻找一个距离原点最远像元点,作为第1个端元;然后遍历特征空间寻找第2个像元点,它与第1个端元的直线距离最远,并把它作为第2个端元;接着遍历特征空间,寻找第3个像元点,其与前两个端元构成的三角形面积最大,作为第3个端元;依次类推,直到找到图像中所有的端元。在每一次端元求解过程中,都要对所有光谱向量进行一次遍历,以找到能使当前的像元点与已找到的像元点构成的高维平行多面体的体积最大,该点即为端元。当从大量像元数据中提取多个端元时,主要计算时间消耗在求解高维平行多面体体积上,且随着提取端元数目的增加,计算量呈指数级变化。另外循环遍历寻找端元过程中,每个像元向量都参与计算,计算体积时,不同像元间互不影响。因此,考虑将算法改写成并行实现,对大量像元数据进行分块求解,以减少在循环遍历上的时间消耗。

2.4 MapReduce 分布式计算模型描述

针对算法串行实现耗时过长问题,利用 Hadoop

平台实现了对算法的并行化处理过程, 具体实现过程如下:

输入: 像元矩阵数据 N 。

输出: 端元矩阵 M 。

1) 在 Map 阶段, 读取已确定的端元, 并求出该数据分片中的待定端元, 并以像元行号作为其 key 值, 其该高维平行多面体的体积以及该像元的值作为 value 形成 $\langle \text{key}, \text{value} \rangle$ 的键值对输出。

2) 对每个数据片分别计算平行多面体体积。

3) 在 Reduce 阶段, 每个 Reduce 输出高维平行多面体的体积最大的像元行号及像元的值, 最终以体积升序排列并输出。

4) 在主程序中, 提取输出文件中最后一个像元的行号及像元的值, 即为端元。

5) 循环执行步骤 1) — 4), 直到提取到的端元数量符合要求, 即程序结束。

这种模式的不足是, 在 Hadoop 机制下的 MapReduce 模式, 每发布一次 MapReduce 任务都很消耗时间, 而每次求一个端元都需要发布一次任务, 所以, 没能很好地发挥出 MapReduce 高效的性能; 同时, 采用了缓存文件来对已求出的端元进行共享, 在数据读取过程中存在问题。

2.5 改进的 MapReduce 模型程序描述

对程序进行优化改进, 提取局部图像各自 N 个端元, 然后对输出的局部端元进行处理, 消除冗余, 达到减少多次发布计算任务带来的时间消耗的目的, 并且避免了使用缓存文件进行端元共享。其具体实现过程如下:

输入: 像元矩阵数据 N 。

输出: 端元矩阵 M 。

1) 在 Map 阶段, 读取该数据分片中的像元数据, 在内部进行多次循环, 以求出该分片中的多个端元, 并以像元行号作为其 key 值, 其该高维平行多面体的体积以及该像元的值作为 value 形成 $\langle \text{key}, \text{value} \rangle$ 的键值对输出。

2) Hadoop 会根据 key 值进行默认排序操作。

3) 在 Reduce 阶段, 每个 Reduce 输出像元行号及像元的值, 最终以体积升序排列并输出。

将每次发布 MapReduce 任务的大循环, 改为每一个 Map 函数中的内部循环; 将每次求出一个端元, 多次求解, 改为一次求解多个局部的端元。

3 实验与加速性能分析

3.1 实验环境

实验过程中在 Linux 环境下搭建 Hadoop 集群, 共 4 台机器, 系统使用的操作系统是 Ubuntu14.04, Hadoop 使用的是 Hadoop-2.20 版本, 该集群各节点配置情况见表 1。

表 1 Hadoop 集群各节点配置情况

Table 1 Configuration of each node in Hadoop cluster

节点名称	CPU 型号	内存	硬盘
master	Intel(R) Core(TM) i5-3470CPU@ 3.20 GHz	8 G	1 TB
slave1	Intel(R) Core(TM) i3-2120CPU@ 3.30 GHz	2 G	500 GB
slave2	Intel(R) Pentium(R) CPU G630@ 2.70 GHz	2 G	500 GB
slave3	Intel(R) Core(TM) i5-3470CPU@ 3.20 GHz	2 G	500 GB

其中, 一台主机作为 NameNode 节点, 同时作为 Resource Manager 服务中的 master 节点, 其他 3 台机器作为 DataNode 节点, 且作为 NodeManager 服务中的 slave 节点, 图 2 展示了 4 台机器之间的网络拓扑关系。

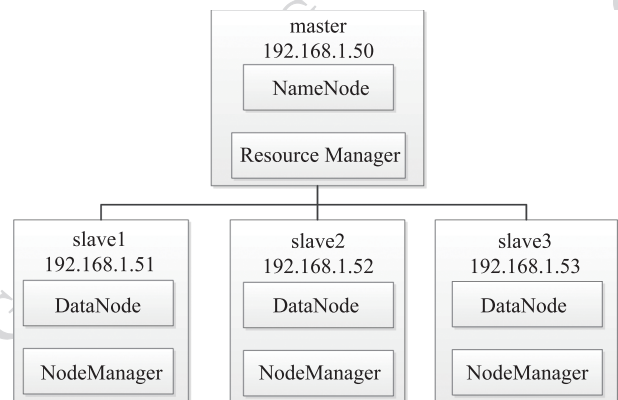


图 2 4 台机器之间的拓扑关系图

Fig. 2 Topological relation between master node and slave nodes

3.2 单机处理实验

3.2.1 端元提取速度比较

实验内容为改进 MapReduce 模型程序和原始 MapReduce 模型程序在单机下处理相同数据时, 提取出 10 个端元所需要的时间。两者均在 Hadoop 伪

分布模式下运行。处理的数据集为 1995 年 7 月获取的 Nevada 州的 Cuprite 采矿区的 AVIRIS Cuprite 数据,截图数据大小为 400×400 像素, 224 个波段,波长范围 $0.4 \sim 2.5 \mu\text{m}$,空间分辨率和光谱分辨率分别为 20 m 和 10 nm。实验过程中,1~2、104~113、148~167 及 221~224 波段由于存在水蒸气吸收和信噪比较低而被去除,因此计算端元提取精度时只用 188 个波段。首先针对该高光谱数据提取 10 个端元的时间如表 2 所示。

表 2 10 个端元的提取速度
Table 2 Running time for extracting ten endmembers

序号	原始 MapReduce 模型	改进 MapReduce 模型
1	345.228	39.458
2	345.687	39.197
3	347.043	39.754
平均	345.986	39.328

从表 2 中可以看出,在单机伪分布的环境下,改进的 MapReduce 模型程序所消耗的时间更少。这

是因为在多次循环发布并执行 MapReduce 任务需要消耗更多数据传输时间,而在改进的程序中,将多次循环迭代改为一次执行内部的多次循环,从而消除了多次循环发布并执行 MapReduce 任务的时间开销,从而提升了端元提取速度。

3.2.2 端元提取精度比较

为验证最大单形体体积端元提取算法精度,分别采用 N-FINDR 算法、VCA 算法、MSVA 算法、MVSA (minimum volume simplex analysis) 算法^[13] 以及 SISAL (simplex identification via split augmented lagrangian) 算法^[14] 从 Cuprite 数据中提取 5 个纯像元,分别代表白云母、方解石、明矾石、高岭石、铍长石 5 种矿物的光谱,然后与 USUG (united states geological survey) 光谱库中 5 种物质光谱进行光谱角匹配,结果如表 3 所示,光谱角数值越小,表示两条光谱相似程度越大。

从表 3 可以看出,MSVA 算法与快速 MSVA 算法提取的端元是一致的;MSVA 算法端元提取效果优于 N-FINDR 和 VCA 算法;由于 MVSA 和 SISAL 为假设不存在纯像元端元提取算法,计算结果相对稍差。

表 3 不同端元提取方法提取端元与光谱库中相应光谱进行光谱角匹配情况
Table 3 Spectral angle distance between USGS mineral spectra and their corresponding endmembers extracted by several endmember identification algorithms

算法	明矾石	铍长石	方解石	高岭石	白云母	平均
N-FINDR	4.81	4.29	7.60	9.92	5.05	6.33
VCA	6.91	5.38	9.53	9.65	6.47	7.59
MSVA	4.55	4.80	5.99	6.89	5.44	5.53
MVSA	12.72	8.41	5.69	15.04	5.36	9.44
SISAL	9.78	5.13	12.78	13.53	8.00	9.84
Fast MSVA	4.55	4.80	5.99	6.89	5.44	5.53

3.3 集群环境下测试

3.3.1 负载对程序运行时间的影响

实验内容为是在分片大小为 128 MB 下,不同的数据集依次在不同节点上运行提取各分片局部 10 个端元的程序。数据集大小分别为 366.4 MB、690.8 MB、1740.8 MB 和 3307.5 MB;节点情况分别为单机伪分布 (master)、2 个节点 (slave1、slave2)、3 个节点 (slave1、slave2、slave3) 和 4 个节点 (master、

slave1、slave2、slave3)。不同数据大小的运行时间如图 3 所示。

从图 3 可以看出,当数据量 (计算负载) 增加时,在单机伪分布的情况下,程序运行时间呈现线性递增;在集群测试中节点个数一定的情况下,程序运行会缓慢递增。在 Hadoop 中 MapReduce 模式下,对端元提取做并行运算时,数据量增大会发挥出其运算优势。

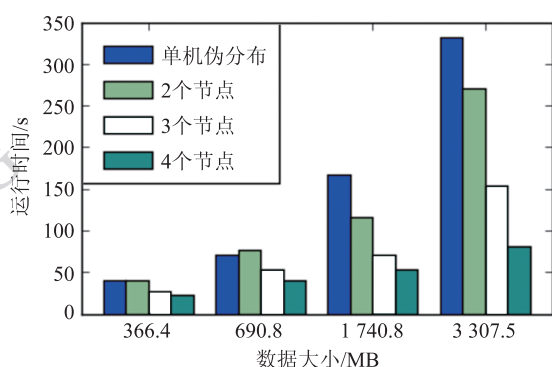


图3 不同数据大小的运行时间图

Fig. 3 Running time for different loads

当数据量小时,在 MapReduce 模式中,对数据进行分片,执行 MapReduce 任务,以及监控各个节点的执行情况所占用的开销远大于程序本身执行的消耗,所以会呈现出不适于并行运算的情况。

3.3.2 节点个数对程序运行时间的影响

在分片大小 $B = 128$ MB 下,数据大小 D 分别 366.4 MB、690.8 MB 和 1740.8 MB 的数据集依次在单机伪分布、2 个节点、3 个节点和 4 个节点运行提取各分片局部 10 个端元的运行时间如下图 4 所示。

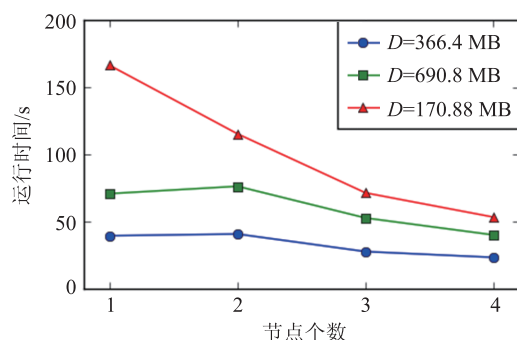


图4 不同节点下的运行时间图

Fig. 4 Running time for different nodes

从图 4 可以看出,当节点个数增加时,在数据量较小的情况下,程序运行时间呈现出缓慢减小;在数据量较大的情况下,程序运行时间会有比较明显的减少。在 Hadoop 中 MapReduce 模式下,对端元提取算法做并行运算时,节点个数增加会发挥出其运算优势。

4 结 论

本文旨在提高光谱遥感图像解混算法处理效率。提出了基于 MapReduce 分布式模型的异构混

合并行算法加速处理模式。并采用该模式来加速基于最大单形体体积的端元提取算法。在实验过程中发现算法中存在矩阵求逆和循环迭代两个部分最耗费时间。试图通过 GPU 加速矩阵求逆运算。但通过实验对比发现,350 阶以内的矩阵求逆运算,使用 Jama 库运算比采用 JCuda 调用 GPU 进行并行加速运算效果更好,当矩阵阶数远大于 350 阶时,采用 JCuda 调用 GPU 加速矩阵求逆运算效果明显。为进一步提高并行加速效果,对原始算法进行改进,减少多次发布 MapReduce 任务的时间消耗。对比单机伪分布、2 个节点、3 个节点和 4 个节点采用改进的 MapReduce 模型程序进行实验可知,Hadoop 多节点对数据进行处理时发挥了很好的并行加速效果,当数据量随之增加的情况下,并行加速的效果并没有减弱,反而是在大数据量的情况下,并行加速的效果表现的更为明显。由此可见,基于 MapReduce 模式的分布式混合并行处理模型可以应用到高光谱图像解混计算领域,尤其对计算任务可分、主机与节点间数据交换量少、计算复杂类算法加速效果明显。因此,该方案可以进一步拓展到其他高光谱图像处理算法领域。

参考文献 (References)

- [1] Goetz A F H. Three decades of hyper-spectral remote sensing of the earth: a personal view [J]. Remote sensing of Environment, 2009, 113: s5-s16.
- [2] Bioucas-Dias J, Plaza A, Camps-Valls G, et al. Hyperspectral remote sensing data analysis and future challenges [J]. IEEE Geoscience and Remote Sensing Magazine, 2013, 1(2): 6-36.
- [3] Chang C I, Xiong W, Liu W, et al. Linear spectral mixture analysis based approaches to estimation of virtual dimensionality in hyperspectral imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2010, 48(11): 3960-3979.
- [4] Xiong W, Chang C I, Wu C C, et al. Fast algorithms to implement N-FINDR for hyperspectral endmember extraction [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2011, 4(3): 545-564.
- [5] Zhao C H, Qi B, Wang Y L. An improved N-FINDR hyperspectral endmember extraction algorithm [J]. Journal of Electronics & Information Technology, 2012, 34(2): 499-503. [赵春晖, 齐滨, 王玉磊. 一种改进的 N-FINDR 高光谱端元提取算法 [J]. 电子与信息学报, 2012, 34(2): 499-503.] [DOI: 10.3724/SP.J.1146.2011.00575]
- [6] Ding H Y, Shi W Z. Fast N-FINDR algorithm for endmember ex-

- traction based on chi-square distribution[J]. *Journal of Remote Sensing*, 2013, 17(1): 122-137. [丁海勇, 史文中. 利用卡方分布改进 N-FINDR 端元提取算法[J]. *遥感学报*, 2013, 17(1): 122-137.] [DOI:10. 11834/jrs. 20131241]
- [7] Wu C C, Lo C S, Chang C. Improved process for use of a simplex growing algorithm for endmember extraction[J]. *Geoscience and Remote Sensing Letters*, IEEE, 2009, 6(3): 523-527.
- [8] Nascimento J M P, Dias J M B. Vertex component analysis: a fast algorithm to unmix hyperspectral data [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2005, 43(4): 898-910.
- [9] Plaza A, Du Q, Chang Y L, et al. High performance computing for hyperspectral remote sensing [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2011, 4(3): 528-544.
- [10] Geng X, Zhao Y, Wang F, et al. A new volume formula for a simplex and its application to endmember extraction for hyperspectral image analysis [J]. *International Journal of Remote Sensing*, 2010, 31(4): 1027-1035.
- [11] Li C H, Zhang X F, Jin H, et al. MapReduce: a new programming model for distributed parallel computing[J]. *Computer Engineering and Science*, 2011, 33(3): 129-135. [李成华, 张新访, 金海, 等. MapReduce: 新型的分布式并行计算编程模型[J]. *计算机工程与科学*, 2011, 33(3): 129-135.] [DOI:10. 3969/j. issn. 1007-130X. 2011. 03. 023]
- [12] Lu F S, Song J Q, Yin F K, et al. Survey of CPU/GPU synergetic parallel computing [J]. *Computer Science*, 2011, 38(3): 5-9. [卢凤顺, 宋君强, 银福康, 等. CPU/GPU 协同并行计算研究综述[J]. *计算机科学*, 2011, 38(3): 5-9.] [DOI:10. 3969/j. issn. 1002-137X. 2011. 03. 002]
- [13] Li J, Bioucas-Dias J M. Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data [C]//*Geoscience and Remote Sensing Symposium*. Boston, USA: IEEE Press, 2008: 250-253.
- [14] Bioucas-Dias J M. A variable splitting augmented Lagrangian approach to linear spectral unmixing [C]//*Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*. Grenoble, France: IEEE Press, 2009: 1-4.