

# 基于 CUDA 的并行 K-means 聚类图像分割算法优化\*

霍迎秋<sup>1</sup> 秦仁波<sup>2</sup> 邢彩燕<sup>2</sup> 陈曦<sup>2</sup> 方勇<sup>1</sup>

(1. 西北农林科技大学机械与电子工程学院, 陕西杨凌 712100;

2. 西北农林科技大学信息工程学院, 陕西杨凌 712100)

**摘要:** 为提高 K-means 聚类算法的运算速度, 基于 CUDA 架构提出一种分块、并行的 K-means 算法, 并采用“合并访问”、“多级规约求和”、“负载均衡”和“指令优化”等策略优化并行算法。实验结果表明, 并行 K-means 算法的分割效果与串行 K-means 算法相同, 但运行速度得到了极大的提高, 加速比最高达到 560, 很好地解决了农业工程实际中由于分割算法带来的瓶颈问题, 能够极大地提高农业劳动生产率。

**关键词:** 图像分割 聚类分割算法 统一计算架构 图形处理器 并行优化

**中图分类号:** TP311 **文献标识码:** A **文章编号:** 1000-1298(2014)11-0047-07

## 引言

图像分割算法是精准农业研究的重要方面, 其广泛应用于智能机械除草<sup>[1-6]</sup>、精准喷施<sup>[7-8]</sup>、机器人采摘<sup>[9-12]</sup>和病害识别<sup>[13-14]</sup>等精准农业研究的各个方向。

阈值分割算法, 包括单阈值法<sup>[15-17]</sup>和自动阈值法<sup>[1,18-19]</sup>, 以其实现简单、运行速度快的特点得到广泛应用, 但它的缺点是不能对复杂背景的图像进行有效分割, 且受光照影响较大。

为了提高复杂背景下目标图像的分割效果, 需引入更多的人工智能, 提高分割算法的自适应性和普适性。典型的算法有: 基于 K-means 聚类算法<sup>[20]</sup>的杂草图像分割, 基于模糊聚类的成熟水果图像分割<sup>[21]</sup>, 基于颜色特征和面积阈值的棉花害螨图像分割<sup>[22]</sup>, 基于数据融合思想的图论分割法分割玉米叶部病害<sup>[23]</sup>等。人工智能的应用提高了算法的自适应性, 但是也带来了新的问题, 由于智能算法涉及到大量运算, 耗时较长, 实时性不高。而另一方面, 为了提高劳动生产率, 精准农业的实时性要求越来越高。如何在引入人工智能提高算法分割质量的同时, 提高算法的运行速度, 进而提高农业劳动生产率, 成为越来越多研究者关注的重点。

随着图形处理器(Graphic processing unit, GPU)的通用计算能力不断增强, 基于 GPU 设计并行算法成为近年来研究的热点<sup>[24-26]</sup>。本文以经典的

K-means 聚类分割算法为研究对象, 自然背景下的杂草和玉米图像分割为研究内容, 提出一种基于 CUDA 架构的分块、并行的 K-means 聚类分割算法, 并采用“合并访问”、“多级规约求和”、“负载均衡”和“指令优化”等策略优化并行算法, 设计 3 个实验进行对比分析。

## 1 基于串行 K-means 算法图像分割

### 1.1 K-means 聚类算法简介

K-means 算法是一种无监督、自适应的聚类算法, 在图像分割中广泛应用<sup>[27-28]</sup>。算法的基本思想是, 通过迭代不断调整聚类中心, 最终将  $n$  个像素构成的图像集合  $X = \{x_i | x_i \in \mathbf{R}^d, i = 1, 2, \dots, n\}$  划分为  $k$  个不同的像素子集或子类  $C_1, C_2, \dots, C_k$ , 且子集合满足下面 3 个条件:

- (1)  $C_i \neq \emptyset, i = 1, 2, \dots, k$ 。
- (2)  $C_i \cap C_j = \emptyset, i, j = 1, 2, \dots, k; i \neq j$ 。
- (3)  $\bigcup_{i=1}^k C_i = X$ 。

以此来实现对图像数据的分割。

K-means 算法的主要步骤:

输入: 聚类个数  $k$  和包含  $n$  个像素的图像数据集。输出: 满足方差最小标准的  $k$  个子类。

(1) 在  $n$  个数据集中, 随机选定  $k$  个初始聚类中心:  $\mu_1, \mu_2, \dots, \mu_k$ 。

(2) 计算每个像素  $x_i (i = 1, 2, \dots, n)$  到各聚类

收稿日期: 2014-05-07 修回日期: 2014-06-26

\* 国家自然科学基金资助项目(61271280)和国家级大学生科技创新重点资助项目(201310712068)

作者简介: 霍迎秋, 博士生, 实验师, 主要从事压缩感知、并行计算研究, E-mail: fallying@nwsuaf.edu.cn

通讯作者: 方勇, 教授, 博士生导师, 主要从事分布式编码、压缩感知研究, E-mail: fangyong@nwsuaf.edu.cn

中心  $\mu_j (j = 1, 2, \dots, k)$  的欧氏距离  $D(x_i, \mu_j) = \sqrt{\sum_{m=1}^d (x_{im} - \mu_{jm})^2}$ , 按最近邻域原则将其划分到聚类中心  $\mu_j$  表示的子类  $C_j$  中, 得到  $k$  个子类:  $C_1, C_2, \dots, C_k$ 。

(3) 设  $d_i = \sum_{x \in C_i} D(x, \mu_i)$  为第  $i$  个子类中所有像素到其聚类中心  $\mu_i$  的欧氏距离和, 每个像素到其聚类中心的欧氏距离和  $s = \sum_{i=1}^k d_i = \sum_{i=1}^k \sum_{x \in C_i} D(x, \mu_i)$ , 其中  $i = 1, 2, \dots, k$ 。与上一个循环中计算的欧氏距离和  $s$  对比, 如果  $s$  发生变换, 则进行下一步运算; 如果没有改变, 则算法结束。

(4) 产生新的聚类中心  $\mu'_i = \frac{1}{n_i} \sum_{x \in C_i} x (i = 1, 2, \dots, k)$ , 其中  $n_i$  为子类  $C_i$  中像素数; 转步骤(2)。

## 1.2 基于串行 K-means 算法的图像分割

选取典型的 3 类杂草图像和玉米幼苗图像作为测试图像, 每类 5 张。图像宽度 1 600 像素, 高度 1 200 像素。

将待分割的杂草图像视为  $n$  个像素组成的图像集合  $X$ , 每个像素包含  $L, a, b$  3 个分量, 通过 K-means 算法进行聚类分割。考虑自然背景下杂草图像的复杂性, 通过多次实验测试发现, 选择 4 个聚类中心进行聚类分割时, 分割效果最好, 因此设  $k = 4$ 。聚类完成后选择绿色像素区域作为分割的区域, 然后进行腐蚀和膨胀去除噪声和斑点, 获得杂草分割图像。基于 CPU 设计串行 K-means 算法对杂草图像进行分割测试。运行串行算法的主机配置参数: Intel Core(TM)2 Duo CPU E7400 2.84 GHz, 2 GB 内存。

每类杂草任选 1 幅图像和 1 幅玉米图像, 进行 K-means 聚类分割。分割效果如图 1 所示。

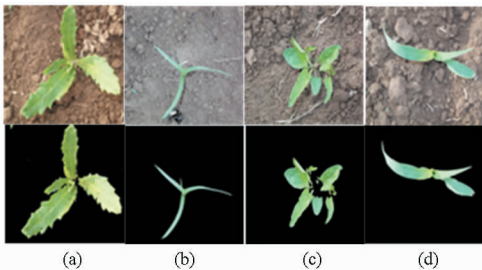


图 1 串行分割图组

Fig. 1 Images segmented by the sequential

K-means algorithm

(a) 刺儿菜 (b) 白颖藎草 (c) 打碗花 (d) 玉米

由图 1 可知, 基于串行 K-means 的杂草分割算法能够有效分割杂草和玉米图像。

在 3 类杂草和玉米图像中, 每类选取 5 张图像,

计算分割算法的运行时间。统计结果如表 1 所示。

表 1 串行 K-means 算法运行时间

Tab. 1 Elapsed time of the sequential K-means algorithm

| 植物<br>种类 | 运行时间/ms |        |        |        |        | 平均耗<br>时/ms |
|----------|---------|--------|--------|--------|--------|-------------|
|          | 图像 1    | 图像 2   | 图像 3   | 图像 4   | 图像 5   |             |
| 刺儿菜      | 38 092  | 35 934 | 30 924 | 36 405 | 38 954 | 36 061.6    |
| 白颖藎草     | 44 054  | 39 256 | 42 981 | 38 891 | 40 546 | 41 145.6    |
| 打碗花      | 22 126  | 25 639 | 27 861 | 26 845 | 23 471 | 25 188.4    |
| 玉米       | 33 342  | 40 969 | 46 542 | 38 067 | 42 563 | 40 296.6    |

对表 1 的实验结果分析发现, 刺儿菜、白颖藎草、打碗花、玉米平均分割耗时分别为 36 061.6、41 145.6、25 188.4、40 296.6 ms。由此可得, 串行分割算法耗时较长。以此类推可知, 待分割图像的背景越复杂, 分割算法运行的时间越长。因此, 提高复杂背景下目标图像的分割速度, 进而提高农业劳动生产率显得尤为重要。

## 2 基于 CUDA 并行设计 K-means 算法

### 2.1 CUDA 并行设计架构

CUDA (Compute unified device architecture) 是一种支持在 GPU 上进行通用、并行计算的编程体系架构, 由 Nvidia 公司设计开发<sup>[18-20]</sup>。GPU 由若干个 SM 组成, 每个 SM 由若干个流处理器 SP、存储器、控制逻辑和少量的其他计算单元组成。每个 SP 都是独立的运算单元。CUDA 架构能够将计算量均衡、灵活的分配到执行单元 SM 上, 具有很高的并行计算性能和良好的可移植性。线程是最小的执行单元。大量的线程被组织成线程块, 同一个线程块中的线程可以通过共享存储器以很低的代价进行通信。CUDA 架构通过将线程块映射到 SM 上实现应用程序的并行计算。SM 通过线程调度器实现任务的分配, 每个 SM 能够运行一个或多个线程块。根据线程块数目和每个线程块中线程数目的不同, 线程调度器能够自动的将一个或多个块分配到 SM 上运行。本研究采用 GTX770 显卡, 主要参数: 2 GB 的显存, 8 个 SM, 1 536 个 SP, 每个 SM 有 48KB 共享内存和 65 536 个寄存器。

### 2.2 K-means 算法的并行性分析

K-means 算法主要包含 3 部分: 计算每个像素点的最小距离 (聚类划分)、计算所有像素点的最小距离和以及产生新的聚类中心。通过分析发现: ① 每个像素点的最小距离的计算是相互独立的, 可完全并行, 并且这一部分是聚类分割算法中计算量最大的一部分, 如果对其进行并行设计, 会极大地提高算法的运行速度。② 计算所有像素点的最小距离和, 可通过并行设计方法“规约求和”

对其进行并行设计,以提高“求和”计算速度。  
 ③产生新的聚类中心包含两部分计算量:统计每个类的像素数和计算每个类的像素值之和,都可以通过设计合适的数据结构,采用原子操作或“规约求和”方法实现。

### 2.3 基于 CUDA 的并行 K-means 算法设计

基于 CUDA 进行并行算法设计,必须充分考虑 GPU 硬件的特点,开启尽可能多的并行执行的线程,将并行计算最大化。将图像数据分割为若干个数据子块,每个线程块处理一个数据子块。如图 2 所示。本文设计 3 个 CUDA 核函数来实现 K-means 算法并行化。

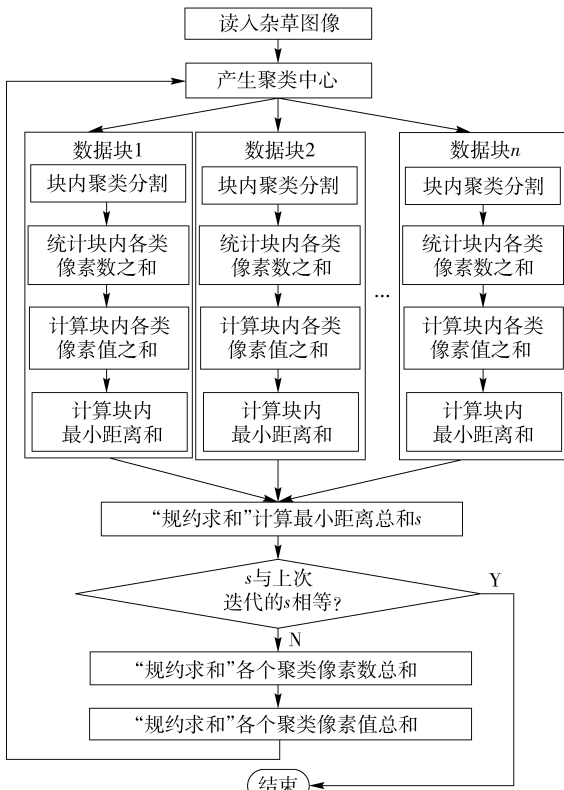


图 2 K-means 算法并行设计

Fig. 2 Design of parallel K-means algorithm

#### 2.3.1 设计核函数 Kernel1 << <gird, block >>>

主要实现块内像素的聚类分割、块内最小距离求和、块内每个类内像素值求和及每个类像素数统计。详细设计如下:

(1) 核函数参数  $gird$  和  $block$  的维度分别为  $(hw + b - 1)/b$  和  $s \times 1$ ,  $s$  为 512,  $h$  为图像的高度,  $w$  为图像的宽度,  $b$  为  $block$  个数。每个线程块处理一个数据子块,块内每个线程处理一个像素。开启线程总数大于等于图像像素数。

(2) 在块内共享内存中,开辟 5 个大小为  $s \times 1$  的线性空间  $s_1, s_2, s_3, s_4, s_5$ 。其中  $s_5$  存放每个像素点的最小欧氏距离;  $s_1, s_2, s_3, s_4$  代表 4 个类。每个

线程计算一个像素到  $k$  个聚类中心的欧氏距离,根据“最小欧氏距离”得出像素的隶属度(或类别)写入全局内存;利用原子操作统计每个聚类的像素数;向相应类别的线性空间写入像素值,如图 3 所示。同步块内线程,利用“规约求和”(本文涉及到的“规约求和”方案均如图 4 所示)计算块内最小距离和以及块内各类像素值之和。每个线程块依据块索引将“规约求和”结果写入全局内存。

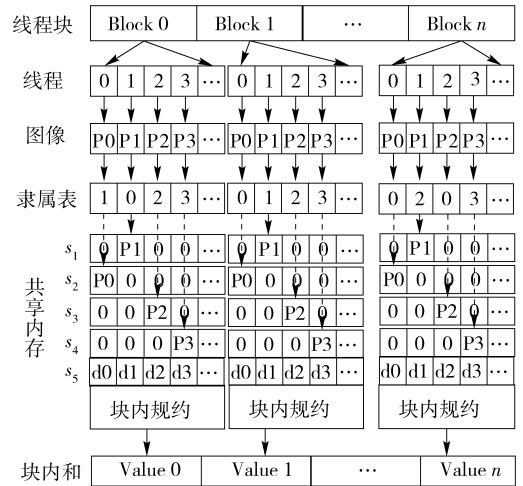


图 3 聚类划分

Fig. 3 Clustering segmentation

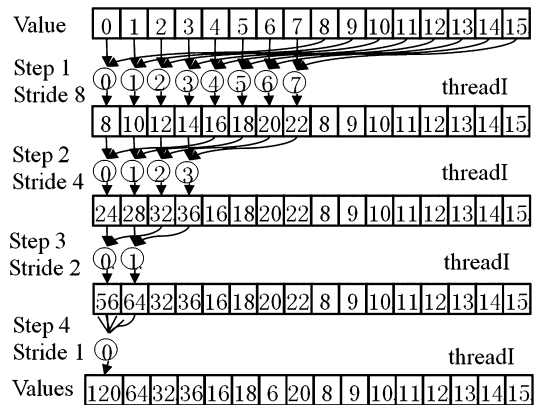


图 4 规约求和示意图

Fig. 4 Parallel reduction

“规约求和”基本思想,假设对 16 个数据进行求和,如图 4 所示,开启 16/2 个线程分别进行数据两两相加,如 ID 为 0 的线程累加第 1、9 个数据,结果写入第 1 个数据的存储空间, ID 为 1 的线程累加第 2、10 个数据,结果写入第 2 个数据的存储空间,同理可推其他线程计算的数据;每个循环参与计算的线程数减半,如第 1 个循环 8 个线程参与计算,第 2 循环 4 个线程参与计算,直至最后 1 个循环 ID 为 0 的线程计算最终相加结果。“规约求和”将求和运算的计算复杂度从  $n$  降为  $\lg n$ ,极大地加快了求和运算的速度。

### 2.3.2 设计核函数 Kernel2 <<<gird,block>>>

在块内共享内存中分配一个大小为  $s$  的线性空间,通过在全局内存上进行二次“规约求和”实现最小距离总和,其中  $gird$  和  $block$  的维度分别为 1 和  $s$ , $s$  取值为 512。将上一个核函数规约求和写入全局内存的数据的索引(id)表示为: $0, \dots, s, \dots, ks, \dots$ 。每个线程读取多个数据,在线程内寄存器变量中进行加和,将结果写入块内共享内存。进行块内“规约求和”得出最小距离总和。

### 2.3.3 设计核函数 Kernel3 <<<1,block>>>

线程块大小为 512。在共享内存中分配 4 个大小为  $s$  的线性空间  $s_0, s_1, s_2, s_3$ ,代表 4 个聚类。通过二次“规约求和”实现各类像素值总和计算。设计思路与 Kernel2 相同:每个线程从全局内存中读取多个数据,在线程内寄存器变量进行加和,最后将结果写入块内共享内存中,再通过“规约求和”实现各类的像素值之和,可得 Value 0、Value 1、Value 2、Value 3,分别表示每个聚类的像素值总和;并利用公式“新聚类中心点的像素值等于类内像素值之和除以类内像素数”,更新聚类中心。详细设计如图 5 所示。



图 5 更新聚类中心

Fig.5 Update the clustering center

每类杂草各选 3 幅图片作为测试样本,分别运用并行、串行 K-means 算法对图像进行聚类分割,计算两种算法所用时间的比值(加速比),实验结果如表 2 所示。并行分割算法分割结果,如图 6 所示。通过图 6 与图 1 对比分析发现,基于 CUDA 分块、并行的 K-means 算法能够对复杂背景下杂草和玉米进行有效分割,并行算法的分割效果与串行 K-means 算法相同。

通过对表 2 分析发现,并行 K-means 聚类分割算法显著提高了分割算法的运行速度,最高加速比为 84.17,最低加速比为 49.52。

## 3 负载均衡优化

鉴于 2.3 节设计的方案 1 中,核函数 Kernel1 计算任务量大,使用的共享内存及寄存器过多,导致设

表 2 串、并行 K-means 算法运行时间对比

Tab.2 Comparison of the sequential algorithm with the parallel algorithm

| 植物种类 | 序号 | 串行算法运行时间/ms | 并行算法运行时间/ms | 加速比   |
|------|----|-------------|-------------|-------|
| 刺儿菜  | 1  | 38 527      | 556.64      | 69.21 |
|      | 2  | 42 744      | 830.47      | 51.46 |
|      | 3  | 41 287      | 490.50      | 84.17 |
| 白颖藎草 | 1  | 44 054      | 662.36      | 66.51 |
|      | 2  | 39 256      | 542.12      | 72.41 |
|      | 3  | 42 981      | 569.65      | 75.45 |
| 打碗花  | 1  | 22 126      | 446.75      | 49.52 |
|      | 2  | 25 639      | 468.49      | 54.73 |
|      | 3  | 27 861      | 465.28      | 59.88 |
| 玉米   | 1  | 55 167      | 774.94      | 71.19 |
|      | 2  | 40 969      | 542.79      | 75.47 |
|      | 3  | 48 542      | 653.31      | 74.30 |

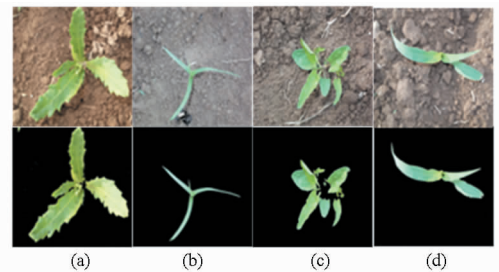


图 6 并行 K-means 算法分割图组

Fig.6 Images segmented by the parallel K-means algorithm

(a) 刺儿菜 (b) 白颖藎草 (c) 打碗花 (d) 玉米

备资源利用不均衡,使得并行 K-means 算法计算效率不高。针对资源分配不均衡的问题,重新划分核函数的计算任务及组织方式对并行算法进行优化;同时,采用共享内存的“规约求和”方法代替原子操作对块内各个聚类像素个数进行统计。优化之后的并行算法整体设计流程图,如图 7 所示。

详细核函数及任务划分设计如下:

(1) 核函数 Kernel1 <<<gird,block>>>,实现聚类分割、块内最小距离求和。在块内共享内存分配 2 个大小为  $size \times 1$  的线性空间  $s_1, s_2$ 。每个线程块将要处理的图像数据子块从全局内存复制到块内共享内存  $s_1$  中,块内每个线程再从  $s_1$  中读取图像像素数据,并计算最短欧氏距离,将最小距离写入  $s_2$  中,最后做块内“规约求和”计算块内最小距离和。

(2) 设计核函数 Kernel2 <<<gird,block>>>,在块内共享内存中开辟 4 个大小为  $s \times 1$  线性空间  $s_0, s_1, s_2, s_3$ ,其中  $s$  值为 512。根据像素的隶属信息表,向相应类别的线性空间写入“1”,同步块内线程,利用“规约求和”统计块内每类像素数;将每个聚类的块内求和结果写入全局内存。详细设计如图 8 所示。再通过全局内存上的“二次规约求和”,

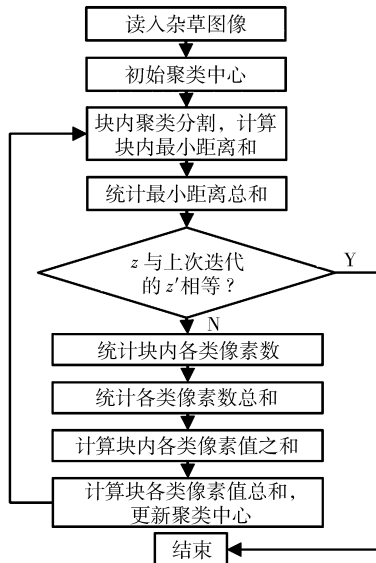


图 7 并行算法设计流程图

Fig. 7 Design of the parallel algorithm

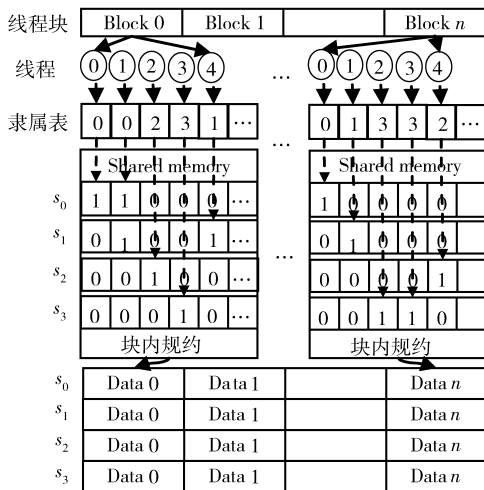


图 8 块内各类像素数统计

Fig. 8 Counting pixels of each cluster within a block

求得各个聚类的像素总和(同方案 1)。

(3) 设计核函数  $\text{Kernel3} \ll \langle \text{grid}, \text{block} \rangle \gg$ , 在块内共享内存中开辟 4 个大小为  $\text{size} \times 1$  线性空间  $s_0, s_1, s_2, s_3$ , 其中  $\text{size}$  值为 512。根据像素隶属度将数据从全局内存复制到块内共享内存  $s_0, s_1, s_2, s_3$ ; 进行块内“规约求和”计算块内每个类像素值之和, 将结果写入全局内存, 设计思路同图 8; 通过全局内存上的“二次规约求和”, 求得各个聚类的像素值之和(同方案 1)。

实验采用刺儿菜杂草图片, 分别通过串行方案、并行方案 1 和并行方案 2(本节方案)进行聚类分割 100 次。串行算法耗时 38 404 ms, 并行方案 1 耗时 535.92 ms, 并行方案 2 耗时 259.35 ms。表明在任务合理划分以及资源均衡利用情况下, 并行聚类分割算法运行效率有了极大提高, 加速比是并行方案 1 的 2 倍多。

### 4 指令级优化

并行设计方案 2 通过负载均衡的优化, 极大的提高了并行算法的运行速度。但是, 通过 Visual Profiler 对并行算法进行性能分析发现: 算法存储器带宽利用率不高, 线程计算延迟较大。分析结果如表 3 所示。由于线程并不会被存储访问阻塞, 它只会因为数据依赖性而阻塞, 为了隐藏延迟, 本文提出指令级优化方案: 每个线程对多个像素点进行聚类划分, 使单个线程运行更多的指令。

表 3 单线程计算单像素点存储带宽  
Tab.3 Memory bandwidth of a single thread computing a single pixel

| 存储器           | 数据传输带宽/(GB·s <sup>-1</sup> ) |
|---------------|------------------------------|
| Local Loads   | 0                            |
| Local Stores  | 2.51                         |
| Shared Loads  | 235.77                       |
| Shared Stores | 195.98                       |
| Global Loads  | 34.16                        |
| Global Stores | 1.17                         |
| L1/SM         | 469.6                        |
| L2/Cache      | 30.54                        |

对并行方案 2 进行优化得出方案 3: 在聚类划分计算最短欧氏距离和的过程中, 每个线程计算若干个像素到聚类中心的距离, 并通过线程内寄存器变量将这几个像素的最小欧氏距离进行加和, 将结果写入块内共享内存, 最后在共享内存上进行“规约求和”计算块内最小距离和, 其他设计不变。

详细设计如图 9 所示, 每个线程计算的  $N$  个像素点的索引(id)为:  $b_{id}sN + t_{id} + is; i = 0, 1, \dots, N$ 。其中  $b_{id}$  为线程块号,  $t_{id}$  为块内线程号,  $s$  为线程块的大小 512,  $N$  为 1 到 100 的整数。

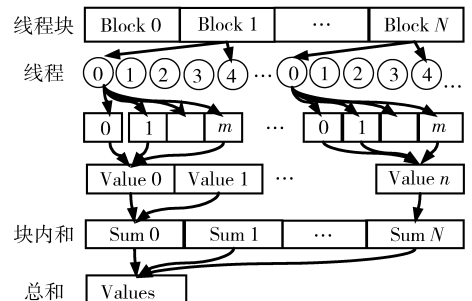


图 9 单线程计算多个像素

Fig. 9 Single thread computing multiple pixels

寄存器的访存速度比共享内存的访存速度快很多, 充分利用寄存器资源, 能够获得更好的效果。同时, 本方案(方案 3)也减轻了块内像素点最小距离“规约求和”的计算量, 客观上加速了算法的运行。

任选 1 幅待分割图像作为测试样本,运用串行分割算法对其进行聚类分割,算法运行 100 个循环,取其平均值作为串行算法分割时间。设计并行分割算法,每个线程计算  $t(t=1,2,\dots,L)$  个像素的最小距离。为消除其他无关因素的影响,对于每个  $t$  值,算法运行 100 个循环,取其平均值作为并行算法运行时间;然后计算出并行分割算法在每个  $t$  值时的加速比。如图 10 所示。设置线程块大小为 128、256 和 512,测试不同设置时算法的加速比。

对图 10 进行分析可知,当线程块大小为 256,每个线程计算 40 个像素时,加速比最高,为 560 倍。此时方案 3 设计的并行算法运行速度远远高于方案 2(加速比为 148 倍)。通过 Visual Profiler 分析一级缓存、二级缓存、共享内存和全局内存等的带宽利用率,详细数据如表 4 所示。通过对表 3 和表 4 对比分析发现,方案 3 较之单线程计算单像素的方案 2,一级缓存、二级缓存和全局内存的带宽利用率都得到了很大的提高。因此,实验结果证明指令级优化

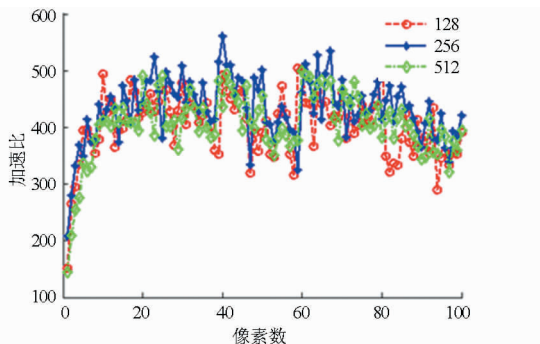


图 10 单线程多像素点性能分析

Fig. 10 Performance of a single thread computing multiple pixels

表 4 单线程计算 40 个像素存储带宽

Tab. 4 Memory bandwidth of a single thread computing 40 pixels

| 存储器           | 数据传输带宽/(GB·s <sup>-1</sup> ) |
|---------------|------------------------------|
| Local Loads   | 0                            |
| Local Stores  | 0.001 7                      |
| Shared Loads  | 188.97                       |
| Shared Stores | 191.35                       |
| Global Loads  | 162.44                       |
| Global Stores | 0.448 31                     |
| L1/SM         | 543.20                       |
| L2/Cache      | 162.96                       |

方案可行,且远优于方案 2,能够极大地提高算法的运行速度。

## 5 结束语

为提高 K-means 聚类分割算法的运行速度,分析了算法的工作原理及并行性,提出了分块、并行的 K-means 算法,并采用“合并访问”、“多级规约求和”和“负载均衡”等优化策略优化并行算法,提高算法的运行速度;通过 Visual Profiler 对并行算法进行详细分析,针对并行算法的缺点和不足,提出基于多像素的指令级优化方案优化算法;设计并行算法与串行算法,进行对比分析。实验结果表明,基于 CUDA 的并行 K-means 算法能够极大提高算法运行速度,相对于串行算法加速比最高达 560 倍,可广泛应用于智能机械除草、机器人采摘、变量喷施、病害识别和农业大数据处理等涉及高密度计算和高实时性要求的应用中,替代传统串行 K-means 算法,加速整个算法的运行,进而提高农业劳动生产率。

## 参 考 文 献

- 胡炼,罗锡文,曾山,等.基于机器视觉的株间机械除草装置的作物识别与定位方法[J].农业工程学报,2013,29(10):12-18.  
Hu Lian, Luo Xiwen, Zeng Shan, et al. Plant recognition and localization for intra-row mechanical weeding device based on machine vision[J]. Transactions of the CSAE, 2013, 29(10):12-18. (in Chinese)
- 毛文华,张银桥,王辉,等.杂草信息实时获取技术与设备研究进展[J].农业机械学报,2013,44(1):190-195.  
Mao Wenhua, Zhang Yinqiao, Wang Hui, et al. Advance technique sand equipments for real-time weed detection [J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(1):190-195. (in Chinese)
- Jin Jian, Tang Lie. Corn plant sensing using real-time stereo vision[J]. Journal of Field Robotics, 2009, 26(6-7):591-608.
- Pérez-Ruiz M, Slaughter D C, Gliever C J, et al. Automatic GPS-based intra-row weed knife control system for transplanted row crops[J]. Computers and Electronics in Agriculture, 2012, 80(1):41-49.
- Nieuwenhuizen A T, Hofstee J W, Henten E J. Performance evaluation of an automated detection and control system for volunteer potatoes in sugar beet fields[J]. Biosystems Engineering, 2010, 107(1):46-53.
- Berge T W, Goldberg S, Kaspersen K, et al. Testing image-based site-specific weed control in cereals[C]//Poster at the 15th European Weed Research Society Symposium, 2010.
- 葛玉峰,周宏平,郑加强,等.基于机器视觉的室内农药自动精确喷雾系统[J].农业机械学报,2005,36(3):86-89.  
Ge Yufeng, Zhou Hongping, Zheng Jiaqiang, et al. Indoor pesticide smart spraying system based on machine vision [J]. Transactions of the Chinese Society for Agricultural Machinery, 2005, 36(3):86-89. (in Chinese)
- 尹东富,陈树人,裴文超,等.基于处方图的室内变量喷药除草系统设计[J].农业工程学报,2011,27(4):131-135.

- Yin Dongfu, Chen Shuren, Pei Wenchao, et al. Design of map-based indoor variable weed spraying system[J]. Transactions of the CSAE, 2011, 27(4): 131 - 135. (in Chinese)
- 9 吕继东, 赵德安, 姬伟. 苹果采摘机器人目标果实快速跟踪识别方法[J]. 农业机械学报, 2014, 45(1): 65 - 72.  
Lü Jidong, Zhao Dean, Ji Wei. Fast tracing recognition method of target fruit for apple harvesting robot[J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(1): 65 - 72. (in Chinese)
- 10 Peng Hongxing, Zou Xiangjun, Xiong Juntao, et al. Research of target recognition based on active shape model[J]. Key Engineering Materials, 2012, 522: 720 - 725.
- 11 Bulanon D M, Kataoka T. Fruit detection system and an end effector for robotic harvesting of Fuji apples[J]. CIGR Journal, 2010, 12(1): 203 - 210.
- 12 Naoshi Kondo, Koki Yata, Michihisa Iida, et al. Development of an end-effector for a tomato cluster harvesting robot[J]. Engineering in Agriculture, Environment and Food, 2010, 3(1): 20 - 24.
- 13 Camargo A, Smith J S. Image pattern classification for the identification of disease causing agents in plants[J]. Computers and Electronics in Agriculture, 2009, 66(1): 121 - 125.
- 14 霍迎秋, 唐晶磊, 尹秀珍, 等. 基于压缩感知理论的苹果病害识别方法[J]. 农业机械学报, 2013, 44(10): 227 - 232.  
Huo Yingqiu, Tang Jinglei, Yin Xiuzhen, et al. Apple disease recognition based on compressive sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(10): 227 - 232. (in Chinese)
- 15 侯学贵, 陈勇, 郭伟斌. 除草机器人田间机器视觉导航[J]. 农业机械学报, 2008, 39(3): 106 - 108.  
Hou Xuegui, Chen Yong, Guo Weibin. Machine vision-based navigation for a weeding robot[J]. Transactions of the Chinese Society for Agricultural Machinery, 2008, 39(3): 106 - 108. (in Chinese)
- 16 Tang Jinglei, He Dongjian, Jing Xu, et al. Maize seedling/weed multiclass detection in visible/near infrared image based on SVM[J]. Journal of Infrared Millimeter Waves, 2011, 30(2): 97 - 103.
- 17 吴兰兰, 刘剑英, 文友先, 等. 基于支持向量机的玉米田间杂草识别方法[J]. 农业机械学报, 2009, 40(1): 162 - 166.  
Wu Lanlan, Liu Jianying, Wen Youxian, et al. Weed identification method based on SVM in the corn field[J]. Transactions of the Chinese Society for Agricultural Machinery, 2009, 40(1): 162 - 166. (in Chinese)
- 18 黄士凯, 祁力钧, 张建华, 等. 基于行宽的玉米行间杂草识别算法[J]. 中国农业大学学报, 2013, 18(1): 165 - 171.  
Huang Shikai, Qi Lijun, Zhang Jianhua, et al. An identification algorithm of weeds among multi-row corn based on the mapping of the corn rows width[J]. Journal of China Agricultural University, 2013, 18(1): 165 - 171. (in Chinese)
- 19 毛文华, 王一鸣, 张小超, 等. 基于机器视觉的苗期杂草实时分割算法[J]. 农业机械学报, 2005, 36(1): 83 - 86.  
Mao Wenhua, Wang Yiming, Zhang Xiaochao, et al. Real-time algorithm based on machine vision to segment weeds at seedling [J]. Transactions of the Chinese Society for Agricultural Machinery, 2005, 36(1): 83 - 86. (in Chinese)
- 20 赵川源, 何东健, 乔永亮. 基于多光谱图像和数据挖掘的多特征杂草识别方法[J]. 农业工程学报, 2013, 29(2): 192 - 198.  
Zhao Chuanyuan, He Dongjian, Qiao Yongliang. Identification method of multi-feature weed based on multi-spectral images and data mining[J]. Transactions of the CSAE, 2013, 29(2): 192 - 198. (in Chinese)
- 21 陈科尹, 邹湘军, 熊俊涛, 等. 基于视觉显著性改进的水果图像模糊聚类分割算法[J]. 农业工程学报, 2013, 29(6): 157 - 164.  
Chen Keyin, Zou Xiangjun, Xiong Juntao, et al. Improved fruit fuzzy clustering image segmentation algorithm based on visual saliency[J]. Transactions of the CSAE, 2013, 29(6): 157 - 164. (in Chinese)
- 22 刁智华, 王欢, 宋寅卯, 等. 复杂背景下棉花病叶害虫图像分割方法[J]. 农业工程学报, 2013, 29(5): 147 - 152.  
Diao Zhihua, Wang Huan, Song Yinmao, et al. Segmentation method for cotton mite disease image under complex background [J]. Transactions of the CSAE, 2013, 29(5): 147 - 152. (in Chinese)
- 23 虎晓红, 李炳军, 刘芳. 多颜色空间中玉米叶部病害图像图论分割方法[J]. 农业机械学报, 2013, 44(2): 177 - 181.  
Hu Xiaohong, Li Bingjun, Liu Fang. Image segmentation based on graph theory in multi-color space for maize leaf disease[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(2): 177 - 181. (in Chinese)
- 24 Joo Yul Park, Ki Seok Chung. Parallel LDPC decoding using CUDA and OpenMP [J]. EURASIP Journal on Wireless Communications and Networking, 2011: 172.
- 25 Dariusz Mrozek, Miłosz Brożek, Bożena Małysiak-Mrozek. Parallel implementation of 3D protein structure similarity searches using a GPU and the CUDA[J]. Journal of Molecular Modeling, 2014, 20(2): 2067.
- 26 董莹莹, 赵星, 王纪华. 基于 GPU 的虚拟植物生长的双尺度自动机模型实现方法[J]. 农业工程学报, 2011, 27(5): 207 - 211.  
Dong Yingying, Zhao Xing, Wang Jihua. Rendering plants with dual-scale automaton model based on GPU[J]. Transactions of the CSAE, 2011, 27(5): 207 - 211. (in Chinese)
- 27 Chen Jiazhong, Cao Hua, Ju Zengwei, et al. Non-attention region first initialisation of k-means clustering for saliency detection [J]. Electronics Letters, 2013, 49(22): 1384 - 1386.
- 28 Krista Rizman Zalik. An efficient k-means clustering algorithm[J]. Pattern Recognition Letters, 2008, 29(9): 1385 - 1391.

- tea vibrating fine manipulation equipment[J]. Transactions of the CSAE, 2013, 29(10):269–277. (in Chinese)
- 14 杨贤强, 沈生荣. 人工光照对离体茶鲜叶氨基酸及香气成分的影响[J]. 福建茶叶, 1991(1):11–21.
- 15 范仕胜, 晋秀, 杨清, 等. 人工光照萎凋对茶叶主要品质成分与酶活性的影响[J]. 湖北农业科学, 2012, 51(6):1152–1155.
- Fan Shisheng, Jin Xiu, Yang Qing, et al. Effects on main quality components and enzyme activity of tea leaves by artificial light withering[J]. Hubei Agricultural Sciences, 2012, 51(6):1152–1155. (in Chinese)
- 16 任广鑫. 基于近红外分析技术的红茶成分分析与产地识别的研究[D]. 合肥:安徽农业大学, 2012.

## Development of 6CQCL – 50 Multi-functional Preprocessing Machine for Fresh Tea

He Xuejun<sup>1</sup> Wang Jin<sup>1</sup> Tang Xiaolin<sup>2</sup> Lu Guodong<sup>1</sup> Wang Bingwen<sup>1</sup> Ren Guangxin<sup>2</sup>

(1. State Key Laboratory of Fluid Power Transmission and Control, Zhejiang University, Hangzhou 310027, China

2. Hangzhou Tea Research Institute, China CO-OP, Hangzhou 310016, China)

**Abstract:** In order to enrich the kind of teas being produced, and improve the utilization of fresh tea materials, a multi-functional preprocessing machine for fresh tea was developed which could realize spreading of green tea, withering of black tea, and green-making of oolong tea. The general structures as well as the working principle of some key parts like mechanical transmission system, leaf turning device, ventilation and heating system were introduced. With multi-branch pipe for air intake, the uniformity of wind field was ensured; pump with high pressure and low flow rate, together with small spray nozzle ensured the uniformity of moisture; temperature was accurately controlled according to different layers. Experimental results show that teas produced with this machine had higher physical and chemical indexes and scores compared to the control groups. The results show that the design was reasonable, and the machine was reliable.

**Key words:** Fresh tea Preprocessing machine Spreading Withering Green-making

(上接第 53 页)

## CUDA-based Parallel K-means Clustering Algorithm

Huo Yingqiu<sup>1</sup> Qin Renbo<sup>2</sup> Xing Caiyan<sup>2</sup> Chen Xi<sup>2</sup> Fang Yong<sup>1</sup>

(1. College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China

2. College of Information Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China)

**Abstract:** K-means clustering algorithm is an excellent algorithm which has been widely used in the image processing and data mining. However, the algorithm arouses a high computational complexity. This paper made a parallel analysis of K-means algorithm in detail, and proposed a partitioning and parallel K-means algorithm based on CUDA (Compute unified device architecture). In addition, some optimization strategies, e. g., coalesced memory access, parallel reduction, load balance and instruction optimization, were discussed to obtain the higher performance. Experimental results show that the parallel K-means algorithm achieves 560x speedup over the sequential C codes, while maintains the same effect. Hence it solves the bottleneck of the algorithm perfectly, which is an attractive alternative to the sequential K-means algorithm for image segmentation and clustering analysis.

**Key words:** Image segmentation K-means clustering algorithm CUDA GPU Parallel optimization