

基于内容的月球影像检索并行框架设计*

陈慧中^{1,2}, 陈永光^{1,3}, 景宁¹, 陈葶¹, 刘义¹

(1. 国防科技大学 电子科学与工程学院, 湖南长沙 410073;

2. 西南电子电信技术研究所上海分所, 上海 200434;

3. 军械工程学院, 河北石家庄 050003)

摘要: 基于内容的月球影像检索面向月球探测计划获取的大量遥感数据, 根据影像视觉内容, 提供一种方便而高效的检索方式。为提高其检索效率, 在对基于内容的月球影像检索过程进行分析的基础上, 运用 Petri 网完成过程建模与并行化分析。提出一种基于内容的月球影像检索并行框架, 并据此部署实验系统, 将嫦娥月球影像等实际数据集投入其中进行检索实验。结果表明, 基于内容的月球影像检索并行框架能够有效提升查询检索效率。

关键词: 月球影像; 基于内容的图像检索; 并行框架 Petri 网

中图分类号: TP791 **文献标志码:** A **文章编号:** 1001-2486(2013)01-0115-08

A parallel architecture of content based retrieval for lunar images

CHEN Huizhong^{1,2}, CHEN Yongguang^{1,3}, JING Ning¹, CHEN Lu¹, LIU Yi¹

(1. College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China;

2. Shanghai Branch, Southwest Electronic and Telecommunication Research Institute, Shanghai 200434, China;

3. Ordnance Engineering College, Shijiazhuang 050003, China)

Abstract: Content-based lunar image retrieval provides a convenient and efficient way for accessing relevant lunar exploration images by their visual contents. To increase the efficiency, the process of content-based lunar image retrieval was analyzed and modeled using Petri nets, and a parallel mechanism was designed based on the model. A parallel architecture was then proposed for the content based retrieval of lunar exploration images. According to the architecture, an experimental system was implemented. Experiments upon real datasets including Chang'e lunar exploration images confirm that the proposed parallel architecture can effectively improve the constructive and retrieval efficiency.

Key words: lunar images; content-based image retrieval; parallel architecture; Petri net

20 世纪末以来, 全球范围内掀起了新一轮的月球探测热潮, 迄今已实施多个新的月球探测计划, 比如: 美国的 CLEMENTINE、LRO, 欧洲的 SMART-1, 日本的 SELENE, 印度的 Chandrayaan-1 以及我国的嫦娥一号、嫦娥二号等。这些探月计划获得了大量的月球遥感影像数据并且通过互联网向全世界公开发布。然而, 不同来源的影像数据组织方式存在巨大的差异。传统的基于数据卷组织的检索方式依赖地名点、月面经纬度以及传感器信息等元数据, 十分繁琐且效率低下, 并对用户的领域知识有非常高的要求。因此, 业内人士开始关注基于内容的月球影像检索方法, 例如, Meyer 和 Deans^[1]提出了一种基于纹理特征的行星探测数据内容检索方法, 在 Apollo 月球影像

数据和 MER MI、MOC NA 火星探测数据上进行了实验验证。基于内容的月球影像检索借鉴了基于内容的图像检索 (CBIR, Content Based Image Retrieval) 技术, 检索时, 能自动地分析影像的内容特征, 从大量的非同源数据中匹配检索出具有相似性视觉特征的月球影像, 从而满足更为高效而广泛的数据获取需求。然而, 月球影像数据量大、内容特征多样、维度较高, 其特征的提取与匹配均十分耗时。为提升性能, 有必要针对基于内容的月球影像检索特点对整个过程进行建模和并行优化设计, 这是该检索方法能够投入实际应用的重要保证。目前国内外这方面的研究偏少, 相关研究主要着重于特征与匹配, 并未涉及整体的框架设计, 也未对整体检索效率进行深入思考。

* 收稿日期: 2012-05-06

基金项目: 国家 863 计划项目 (2008AA12A211, 2011AA12A306); 国家自然科学基金资助项目 (60902036, 61070035)

作者简介: 陈慧中 (1982—), 女, 上海人, 博士研究生, E-mail: chen_huizhong@yahoo.cn;

景宁 (通信作者), 男, 教授, 博士, 博士生导师, E-mail: ningjing@nudt.edu.cn

在相关的基于内容图像检索方面,有不少学者对其并行做出研究,典型的包括:Barlass^[2]设计了一种混合集群框架用以加速图像检索任务,采用基于 MPI 的普通计算节点集群与基于 CUDA 的 GPU 计算集群共同实现多个检索任务的并行; Zhang 等^[3]基于 MapReduce 并行编程模型提出了一种分布式图像检索系统; Vlachou 等^[4]提出一种能够支持分布式索引结构的基于 P2P 网络的分布式图像检索框架; Schweitzer^[5]提出一种使用样例代表图像加速检索过程的分布式算法; Yadav 等^[6]基于 GPU 并行技术基于使用色彩信息对压缩域的医学影像进行检索; Montagnata 等^[7]使用大规模网格技术并行检索医学影像图片。Premchaiswadi 等^[8]提出一种使用连接查询和相关反馈的在线图像检索框架,使用多线程技术加以并行等。相关的研究大多关注已有并行技术、模型等在图像检索领域的应用,对于视觉内容特征较特殊的月球影像检索针对性不强。

与以往研究不同,本文从提高基于内容的月球影像检索整体效率出发,对整个检索过程进行分析与描述,利用 Petri 网^[9]构建了过程模型并进行并行分析,据此提出了一种基于内容的月球影

像检索并行框架。使用普通台式机对框架进行了部署后进行的多个并行效率实验结果表明,该框架能够以较为低廉的成本为基于内容的月球影像检索带来明显的性能提升。

1 基于内容的月球影像检索过程模型

1.1 过程描述

基于内容对月球影像进行检索时,首先需建立影像内容检索库,即对月球影像数据库中的影像数据分别提取其全局内容特征和局部内容特征,然后将特征分别存储于月球影像全局特征库和月球影像局部特征库,在此基础上建立全局特征索引与局部特征索引;当对查询样例影像进行检索时,首先提取样例影像的全局特征,通过全局特征索引在全局特征库中进行近邻查找,找到度量空间上距离最近的 k_c 个全局近邻组成候选集合 **candidates**,然后提取样例影像的局部特征,在候选集合范围内通过局部特征索引进行查询匹配,根据匹配结果得到前 k 个匹配度最高的候选者组成特征查询结果集合 **result**,最后从月球影像数据库中找出特征查询结果对应的月球影像 **I_result**作为检索结果返回。

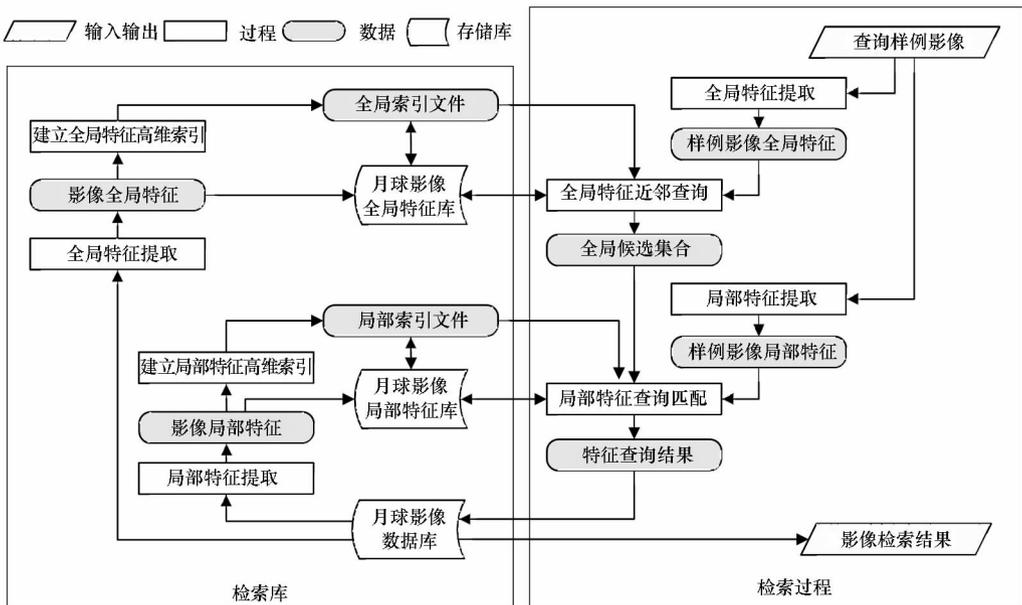


图 1 基于内容的月球影像检索总体过程示意

Fig. 1 Illustration of content-based lunar image retrieval

基于内容的月球影像检索总体过程如图 1 所示,其中,存储库与关键运算过程具体描述如下:

存储库 1 (月球影像库) 存储月球影像数据,记为 I_BASE ; 存储 $I = \{I_i | i = 1, 2, \dots, N\}$, 其中 I_i 定义为一幅月球影像,库中影像总的数目

为 N 。

存储库 2 (全局特征库) GF_BASE 存储月球影像的全局特征向量;对于每一幅月球影像 I_i ,有且仅有一个全局特征向量 $\vec{gf}_i = \{gf_{i1}, gf_{i2}, \dots, gf_{iDg}\}$, 其中 Dg 为全局特征向量的维数;则共存储 N 个

全局特征向量,记为 $\mathbf{gfv} = \{\vec{gfv}_i | i = 1, 2, \dots, N\}$ 。

存储库 3(局部特征库) LF_BASE 存储月球影像的局部特征向量;对于每一幅月球影像 I_i ,有 M_i 个局部特征向量 $\vec{lfv}_{ij} = \{lf_{ij1}, lf_{ij2}, \dots, lf_{ijDl}\}$,即 $\mathbf{lf}_i = \{\vec{lfv}_{ij} | j = 1, 2, \dots, M_i\}$,其中 Dl 为局部特征向量的维数;则共存储 N 个局部特征向量集合,记为 $\mathbf{lfv} = \{\mathbf{lf}_i | i = 1, 2, \dots, N\}$ 。

过程 1(全局特征提取) 对输入的月球影像提取一个 Dg 维全局特征向量;可定义为函数 $gf_extract(\cdot)$,其输入为月球影像库中影像 I_i 或者查询样例影像 Q ,输出为 \vec{gfv}_i 或者查询样例影像的全局特征向量 \vec{qgfv} 。

过程 2(局部特征提取) 对输入的月球影像提取多个 Dl 维局部特征向量;可定义为函数 $lf_extract(\cdot)$,其输入为月球影像库中影像 I_i 或者查询样例影像 Q ,输出为 \mathbf{lf}_i 或者查询样例影像的局部特征向量集合 $\mathbf{qlf} = \{\vec{qlfv}_j | j = 1, 2, \dots, M_q\}$ 。

过程 3(全局索引建立) 对月球影像的全局特征建立用于近邻查询的高维索引结构,月球影像库中所有的影像只建立一个全局特征索引;可定义为函数 $index_gf(\cdot)$,其输入为 \mathbf{gfv} ,输出为全局特征索引文件 gf_index 。

过程 4(局部索引建立) 对月球影像的局部特征建立用于查询匹配的高维索引结构,月球影像库中每一幅影像均建立一个局部特征索引;可定义为函数 $index_lf(\cdot)$,其输入为月球影像 I_i 的局部特征向量集合 \mathbf{lf}_i ,输出为 I_i 的局部特征索引文件 lf_index_i 。

过程 5(全局特征近邻查询) 通过全局特征索引查找查询样例影像的全局特征在全局特征库中的近邻;可定义为 $gf_knn(\cdot)$,其输入为 \vec{qgfv} 和 gf_index ,输出为与 \vec{qgfv} 距离最近的 k_c 个全局近邻组成候选集合 **candidates**。

过程 6(局部特征查询匹配) 通过局部特征索引逐一匹配查询样例影像与候选集合影像的局部特征,根据匹配率求得前 k 个匹配度最高的候选者作为特征查询结果;可定义为 $lf_match(\cdot)$,其输入为 \mathbf{lfv} 和 **candidates**,输出为特征查询结果 **result**。

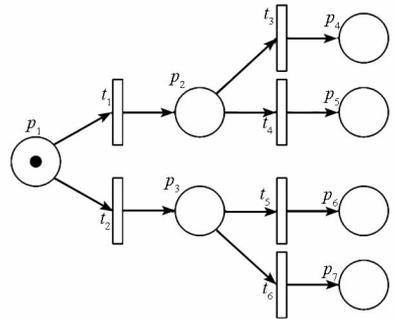
1.2 过程建模

为了更好地进行流程分析与并行化设计,下面利用 Petri 网对基于内容的月球影像检索过程进行建模。Petri 网是由库所、变迁、有向弧和托肯组成的一种有向网。可以用直观的图形化和严格的形式化语义描述网格中顺序、并发、选择等逻辑

关系和特性,是一种广泛用于并发、异步系统建模和分析的重要工具^[10]。

使用五元组 $PN = (P, T, F, M_i, M_o)$ 定义基于内容的月球影像检索 Petri 网。其中, $P = \{p_1, p_2, \dots, p_n\}$ 为库所集合,用于描述系统可能的局部状态, $n > 0$ 为库所个数; $T = \{t_1, t_2, \dots, t_m\}$ 为变迁集合,用于描述修改系统状态的事件, $m > 0$ 为库所个数, $P \cup T \neq \emptyset$ 且 $P \cap T = \emptyset$; F 为库所与变迁之间的有向弧集合,用以引述事件能够发生的局部状态以及状态的转换, $F \subseteq (P \times T) \cup (T \times P)$; M_i 为输入库所的集合,描述过程模型中托肯的初始位置, $M_i \subseteq P$; M_o 为输出库所的集合,描述过程模型中托肯的终止位置, $M_o \subseteq P$ 。

根据上述定义,可以为月球影像内容检索库的生成过程构建如图 2 所示的 Petri 网模型,共包含 7 个库所和 6 个变迁,其中,代表起始状态的输入库所 $M_i = \{p_1\}$,代表过程终止状态的输出库所 $M_o = \{p_4, p_5, p_6, p_7\}$ 。模型增加考虑了两个重要的 I/O 操作过程:变迁 t_4 ,代表全局特征入库的过程,记为 $store_gf(\cdot)$;变迁 t_5 代表局部特征入库的过程,记为 $store_lf(\cdot)$ 。托肯指示了从 M_i 到 M_o 的过程中所处的不同状态,图中的托肯(黑点)标示了起始库所。



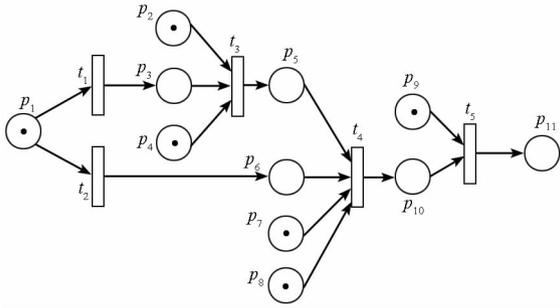
- $p_1: \mathbf{I}; p_2: \mathbf{gfv}; p_3: \mathbf{lfv};$
- $p_4: gf_index; p_5: GF_BASE;$
- $p_6: \{lf_index_i | i = 1, 2, \dots, N\};$
- $p_7: LF_BASE;$
- $t_1: gf_extract(\cdot); t_2: lf_extract(\cdot);$
- $t_3: index_gf(\cdot); t_4: store_gf(\cdot);$
- $t_5: index_lf(\cdot); t_6: store_lf(\cdot).$

图 2 月球影像内容检索库生成过程 Petri 网模型

Fig. 2 Model of feature extraction by Petri net

类似地,为基于内容的月球影像查询检索过程构建 Petri 网模型(如图 3 示意),共包含 10 个库所和 5 个变迁,其中,除上节过程描述中的重要运算过程之外,增加考虑从影像库中取出结果影像的过程,即变迁 t_5 ,记为 $get_result(\cdot)$ 。模型

的输入 $M_I = \{p_1, p_2, p_4, p_7, p_8, p_9\}$, 输出 $M_O = \{p_{11}\}$, 图中的托肯 (黑点) 标示了起始库所。



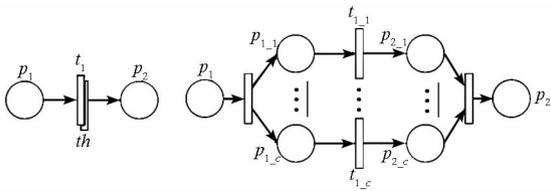
$p_1: Q; p_2: \mathbf{gfv}; p_3: \overrightarrow{qgf\bar{v}};$
 $p_4: \mathbf{gf_index}; p_5: \mathbf{candidates};$
 $p_6: \mathbf{qlf} \quad p_7: \mathbf{lfv}$
 $p_8: \{\mathbf{lf_index}_i | i = 1, 2, \dots, N\};$
 $p_9: \mathbf{I}; p_{10}: \mathbf{result}; p_{11}: \mathbf{I_result};$
 $t_1: \mathbf{gf_extract}(\cdot); t_2: \mathbf{lf_extract}(\cdot);$
 $t_3: \mathbf{gf_knn}(\cdot); t_4: \mathbf{lf_match}(\cdot);$
 $t_5: \mathbf{get_result}(\cdot).$

图 3 基于内容的月球影像查询检索过程 Petri 网模型
 Fig. 3 Model of lunar image retrieval by Petri net

2 基于内容的月球影像检索并行框架

2.1 过程分析与并行设计

为提高索引库构建与查询检索的效率, 本节在 Petri 网模型基础上对过程进行分析并提出并行化设计方案。采取的并行模式主要包含两种: 其一, 适应于单机多核处理器架构的共享存储多线程并行模式; 其二, 基于消息传递的分布式多计算节点并行模式。为更好地描述这两种并行的过程, 提出如图 4 所示两种变迁结构以表述两种不同的并行方式, 其中 th 为多线程并行的线程个数, c 为多节点并行的节点个数。



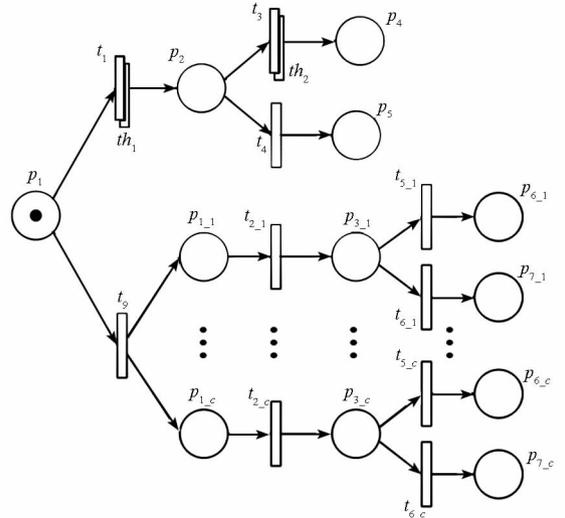
(a) 共享内存并行变迁结构 (b) 消息传递并行变迁结构

图 4 两种并行模式的变迁结构

Fig. 4 Two types of parallel transfer

根据如图 2 所示的 Petri 网模型, 月球影像内容检索库的生成过程中, 主要的计算密集型变迁包括变迁 t_1 全局特征提取、 t_2 局部特征提取、 t_3 建立全局特征索引和 t_5 建立局部特征索引, I/O 密集型变迁包含 t_4 全局特征入库与 t_6 局部特征入库。变迁 t_1 与 t_2 可以同时发生, 变迁完成之后,

t_3, t_4, t_5 和 t_6 也可以同时发生。从变迁本身考虑, t_1 与 t_3 针对所有影像的全局特征集, 可以采取单机多核并行; t_2 与 t_5 针对的是多个影像产生的多个局部特征数据集, 可以以单幅影像为单位进行多节点并行; 采取分布式的特征存储策略对 I/O 密集型变迁 t_4 与 t_6 实现并行写盘操作以提高 I/O 效率。考虑了多种并行的月球影像内容检索库的生成过程 Petri 网模型如图 5 所示, 其中, 增加一个变迁 t_7 用于分发原始月球影像给多个节点以计算局部特征, 记为 $distribute(\cdot)$ 。



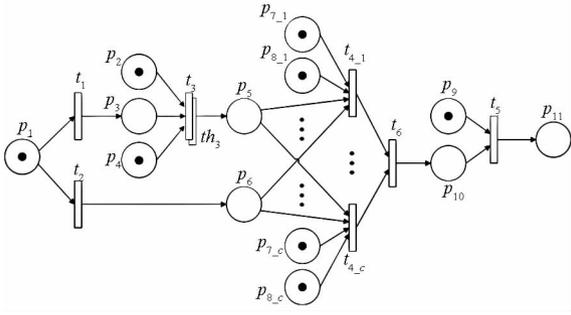
$P_1: \mathbf{I}; p_2: \mathbf{gfv};$
 $p_4: \mathbf{gf_index}; p_5: \mathbf{GF_BASE};$
 $\bigcup_{i=1}^c p_{1,i}: \mathbf{I}; \bigcup_{i=1}^c p_{3,i}: \mathbf{lfv};$
 $\bigcup_{i=1}^c p_{6,i}: \{\mathbf{lf_index}_i | i = 1, 2, \dots, N\};$
 $\bigcup_{i=1}^c p_{7,i}: \mathbf{LF_BASE};$
 $t_1: \mathbf{gf_extract}(\cdot); t_{2,1} \dots t_{2,c}: \mathbf{lf_extract}(\cdot);$
 $t_3: \mathbf{index_gf}(\cdot); t_4: \mathbf{store_gf}(\cdot);$
 $t_{5,1} \dots t_{5,c}: \mathbf{index_lf}(\cdot);$
 $t_{6,1} \dots t_{6,c}: \mathbf{store_lf}(\cdot);$
 $t_7: \mathbf{distribute}(\cdot).$

图 5 并行检索库生成过程 Petri 网模型

Fig. 5 Model of parallel feature extraction by Petri net

基于内容的月球影像查询检索过程中, 主要计算密集型变迁包括变迁 t_1 全局特征提取、 t_2 局部特征提取、 t_3 全局特征近邻查询和 t_4 局部特征查询匹配, I/O 密集型变迁为 t_5 检索结果获取。 t_1 与 t_2 针对检索影像, 两者可以同时发生; t_3 针对所有影像的全局特征集, 可以采取单机多核并行; t_4 对多个局部特征数据集求取匹配度, 可以以单幅影像局部特征数据为单位进行多节点并行; 由于检索结果数据量较小, t_5 根据原始月球遥感影

像的存储方式直接获取即可。考虑了多种并行的月球影像查询检索过程 Petri 网模型如图 6 所示,其中,增加一个变迁 t_6 用于收集各节点的局部特征匹配结果并排序得到特征检索结果,记为 $\text{sort}(\cdot)$ 。



$p_1: Q; p_2: \text{gfv}; p_3: \overrightarrow{qgv};$
 $p_4: \text{gf_index}; p_5: \text{candidates}; p_6: \text{qlf};$
 $p_9: \mathbf{I}; p_{10}: \text{result};$
 $p_{11}: \mathbf{I_result}; t_1: \text{gf_extract}(\cdot); t_2: \text{lf_extract}(\cdot);$
 $t_3: \text{gf_knn}(\cdot); t_{6,1} \dots t_{6,c}: \text{lf_match}(\cdot);$
 $t_5: \text{get_result}(\cdot); t_6: \text{sort}(\cdot);$
 $\bigcup_{i=1}^c p_{7,i}: \text{lfv}; \bigcup_{i=1}^c p_{8,i}: \{ \text{lf_index}_i | i=1,2,\dots,N \}.$

图 6 月球影像并行查询检索过程 Petri 网模型

Fig. 6 Model of parallel lunar image retrieval by Petri net

2.2 并行框架建立

根据过程建模与并行化设计,本文提出一种基于内容的月球影像检索并行框架。如图 7 所示,共分为三个主要部分:月球影像库,主控节点与一组并行计算节点。月球影像库包含了一个或多个存储原始月球遥感影像的数据源,存储的月球影像数据通过一个数据访问层可以被主控节点与多个计算节点访问到。主控节点主要负责过程控制、人机交互与全局特征相关处理。多个并行计算节点则负责完成计算量较大的与局部特征相关的运算。

在实际应用中,可在多台普通的个人电脑上部署此框架,并选用其中计算性能最优的一台作为主控节点,以达到以较为低廉的成本获得较优效率的目的。

2.3 理论分析

本小节对索引库构建与检索过程效率进行理论上的分析。这里,采用表示 $T(\cdot)$ 过程执行时间,则总的串行检索库构建时间为:

$$T_{\text{construct}} = \sum_{i=1}^N \left(\begin{matrix} T(\text{gf_extract}(I_i)) + T(\text{store_gf}(I_i)) \\ + T(\text{lf_extract}(I_i)) + T(\text{store_lf}(I_i)) \end{matrix} \right)$$

根据过程模型(图 5)和并行框架(图 7)设

计,并行检索库构建时间为:

$$T_{\text{construct}} = T_{th_1} + T_{th_2} + \text{Max}_{s=1}^{th_1} \left(\sum_{i=1}^{Ns} T(\text{gf_extract}(I_i)) \right) \\ + \text{Max}_{s=1}^{th_2} \left(\sum_{i=1}^{Ns} T(\text{store_gf}(I_i)) \right), T(\text{index_gf}(I)) \\ + \text{Max}_{s=1}^c \left(\sum_{i=1}^{Ns} T(\text{lf_extract}(I_i)) \right) \\ + \text{Max} \left(\sum_{i=1}^{Ns} T(\text{store_lf}(I_i)), \sum_{i=1}^{Ns} T(\text{index_lf}(I_i)) \right)$$

其中, Ns 为分配到线程或节点 s 上处理的影像数目, T_{th_1} 为提取全局特征时 th_1 个线程的管理时间损耗, T_{th_2} 为存储全局特征时 th_2 个线程的管理时间损耗。从上式可以得出,由于并行的特征库时间损耗由最慢的一个处理单元决定,因此分配到各线程及各节点 s 的负载越均衡,越能取得较好的时间效率;此外,线程的管理时间损耗 T_{th_1} 、 T_{th_2} 随线程数的 th_1 和 th_2 增加而增加,在达到最佳平衡时,能够取得较优的特征库构建效率。

对于单个查询影像 Q ,一次总的串行查询时间为:

$$T_{\text{query}} = T(\text{gf_extract}(Q)) + T(\text{lf_extract}(Q)) \\ + T(\text{gf_knn}(Q)) + T(\text{lf_match}(Q, \mathbf{I}))$$

根据过程模型(图 6)和并行框架(图 7)设计,一次并行查询检索总时间为:

$$T_{\text{query}} = T_{\text{comm}} + \text{Max}(T(\text{gf_extract}(Q)) + T(\text{gf_knn}(Q)), \\ T(\text{lf_extract}(Q))) + \text{Max}_{s=1}^c (T(\text{lf_match}(Q, \mathbf{I}_s)))$$

其中, T_{comm} 为各计算节点与主控节点间分发查询任务及收集查询结果时的数据通信时间, \mathbf{I}_s 为计算节点 s 需要进行局部特征查询匹配的候选影像集合。从上式可以得出,分配到各节点的局部特征查询匹配负载越均衡,越能取得较好的时间效率;此外,随着节点数目 c 的增加, \mathbf{I}_s 的大小就可能越小,从而查询匹配时间有所减少,但同时,由于数据通信量的增加, T_{comm} 将有所增加。因此,在达到计算时间与通信时间最佳平衡时,能够取得较优的查询匹配效率。

根据上述理论分析,整体检索库构建与查询的时间消耗是由各关键过程时间组合而成,由此,关键过程的并行性能提升能够带来整体效率提升。同时,对于单节点多线程并行,在线程数和线程管理消耗达到平衡时取得较优的时间效率;对于多节点并行,在节点数目和数据通信量达到平衡时取得较优的时间效率。

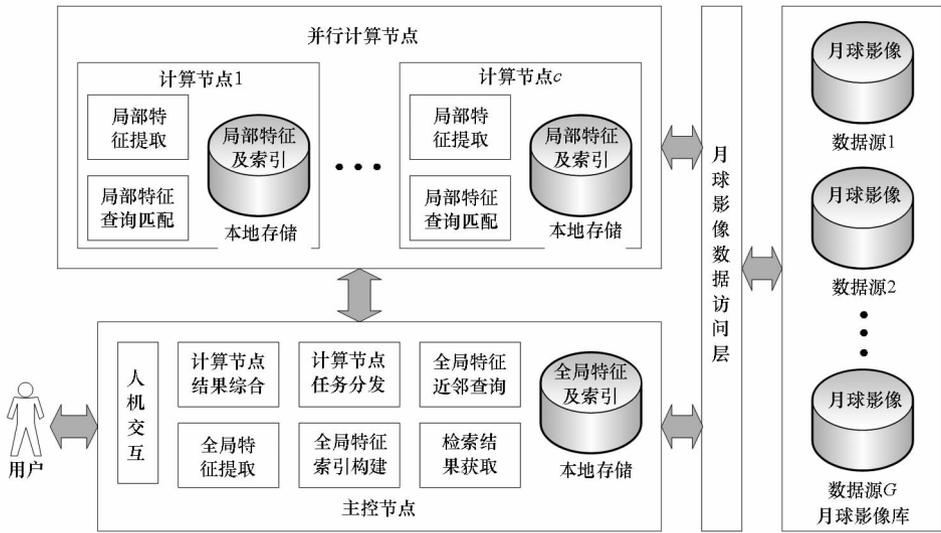


图 7 基于内容的月球影像检索系统并行框架

Fig. 7 Parallel framework of content-based lunar image retrieval system

3 实验与分析

实验部分主要对提出的框架进行部署,形成一个基于内容的月球影像检索并行实验系统;通过对关键环节的并行效率分析,验证框架对检索库构建以及查询检索的效率提升。

3.1 框架部署

根据所提并行框架,使用普通台式计算机部署实现了一个基于内容的月球影像检索并行实验系统,共由一个主控节点与四个计算节点构成,节点配置各不相同,详细信息如表 1 所示。

表 1 实验系统节点配置信息

Tab. 1 Information of experimental parallel nodes

节点名称	CPU	内存	核数	操作系统
主控节点	Intel Core i3 2.93 GHz	2.0G	4	Microsoft Windows XP Professional
计算节点 1	Intel Core2 Duo E6500 2.93GHz	2.0G	2	Linux; OpenSuse 11.2
计算节点 2	Intel Core2 Duo E6500 2.93GHz	2.0G	2	Linux; OpenSuse 11.2
计算节点 3	Intel Core2 Duo E5200 2.5GHz	2.0G	2	Linux; OpenSuse 11.2
计算节点 4	Intel Celeron E430 1.8GHz	1.0G	1	Linux; OpenSuse 11.2

3.2 并行效率实验

并行效率实验通过对几个核心过程的并行效果测试,评对本文所提并行框架。共包含四个部分:在单机多核的主控节点上进行的采用共享内存多线程编程模式实现的并行全局特征提取入库 $gf_extract(\cdot)$ 、 $store_gf(\cdot)$ 与全局特征近邻查询 $gf_knn(\cdot)$;在多计算节点上进行的采用消息映射编程模式实现的局部特征提取入库 $lf_extract(\cdot)$ 、 $store_lf(\cdot)$ 与局部特征查询匹配 $lf_match(\cdot)$ 。

实验系统使用的月球影像库包含两个数据源:嫦娥二号月球影像数据与 CLEMENTINE 月球影像数据;其中,嫦娥数据共有 3227 幅月球影像, CLEMENTINE 数据源包含 8241 幅月球影像。实验程序使用 Visual studio 2008 和 Eclipse PTP 3.01 + GCC 开发,采用 OpenCV 2.0 作为图像处理库。实验中,月球影像全局特征采用 HU 不变矩^[11] 和 Tamura 纹理特征^[12],维数 $Dg = 10$;局部特征采用 SURF 局部特征点^[13],维数 $Dl = 128$;返回结果大小 $k = 10$;候选集大小 $k_c = 100$ 。

图 8 给出了全局特征提取并行实验结果,使用并行加速比与并行效率评价。可以看到,在 主控节点的四核处理器环境下,随着线程数目的变化,并行加速比从 1.256 到 2.27 不等,并行效率从 0.472 到 0.625 不等;当线程数目增加时,加速比与效率随着增加,线程数 $th_1 = th_2 = 7$ 时取得最大值,之后趋于稳定并比最高值略有下降。这是由于线程数小于处理器核数时,没有充分利用每个处理器的计算资源,加速比较小,超过时,由于存储特征的 I/O 操作也是并行的,加速比继续增

加,线程数达到一定的数目($th_1 = th_2 = 7$)时,性能不再提升而达到最大值,之后由于维护多个线程本身的开销增大,并行加速比与效率均略有下降。

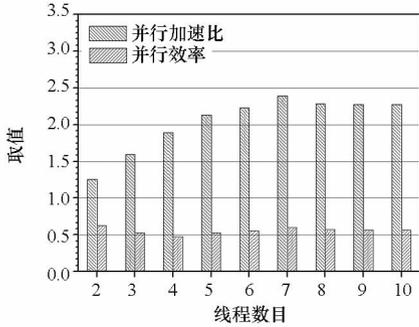


图 8 全局特征提取并行实验结果

Fig. 8 Experimental results of parallel global feature extraction

图 9 给出了全局特征近邻查询的实验结果,为十幅不同查询影像的平均值。在不同线程数目下,并行加速比为 1.57 到 2.6,并行效率为 0.43 到 0.79;与全局特征提取不同,全局特征近邻查询时最佳的并行加速比在线程数目和处理器核数相同($th_3 = 4$)时取得,这是由于在近邻查询过程中,I/O 的比重相对较小,运行时间主要用于计算上,而计算能力则与处理器的核数更加紧密相关。

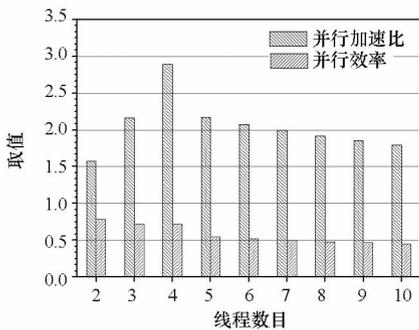


图 9 全局特征近邻查询并行实验结果图

Fig. 9 Experimental results of parallel k-NN search of global features

表 2 局部特征提取并行实验结果

Tab.2 Experimental results of parallel local feature extraction

	计算节点 1/2	计算节点 3	计算节点 4	多节点并行
平均查询匹配时间 (s)	0.909	1.157	2.339	0.320
加速比	2.841	3.616	7.309	/

表 2 给出了在多个计算节点上进行的局部特征提取并行实验结果,为 30 幅月球影像的特征提取平均时间。由于计算节点 1/2 与 3、4 的配置各不相同,表中第一行给出了在各个节点上串行运

行的查询匹配时间与多节点并行执行的时间,第二行给出了多节点并行相对于不同的单个节点串行计算的并行加速比(这里,由于各节点处理核的计算能力各不相同,故不对并行效率进行对比)。可以看到,多节点并行的特征提取时间比单节点串行提取有大幅度的减少,对于配置最好的计算节点 1/2,加速比可达 2.8 左右,对于配置最差的计算节点 4,并行加速比更高,超过了 7。可以说,使用多节点并行对于月球影像的局部特征提取起到了显著的性能提升作用。

表 3 局部特征查询匹配并行实验结果

Tab.3 Experimental results of parallel local feature matching

	计算节点 1/2	计算节点 3	计算节点 4	多节点并行
平均查询匹配时间 (s)	0.953	1.083	1.698	0.407
加速比	2.342	2.661	4.171	/

表 3 给出了在多个计算节点上进行的局部特征查询匹配的并行实验结果,为 10 幅月球影像的查询匹配平均时间。可以看到,采用多节点并行处理的查询匹配时间与单节点串行相比有明显的减少,并行加速比根据各单节点的配置不同,从 2.342 到 4.171 不等。可见,多计算节点并行处理的局部特征查询匹配也取得了良好的并行效果。这里,相比局部特征提取,查询匹配的多节点并行加速比整体要低一些。这是由于在查询匹配过程中,各节点完成匹配后有一个收集结果综合排序的过程(图 6 中的变迁 t_6),这个是没有多节点并行的;此外,由于候选集不一定均匀分布在每个计算节点上,因此查询匹配时存在一定的计算负载不均衡。而局部特征提取时,主控平均地分发任务给多个计算节点,其计算负载是相对均衡的;各节点接受任务后各自完成提取并存储于本地,之后无再聚合的操作,计算的独立性也是较好的;因此可以取得更加高的并行加速比。

综合上述实验结果可以得到,基于本文框架部署的原型系统在各个关键计算环节上都得到了明显的性能提升,由此,本文提出的基于内容的月球影像检索并行框架能够明显提升月球影像检索库的构建效率和相似影像查询检索效率。

4 结论

基于内容的月球影像检索能够为普通用户提供更为直观而有效的月球探测数据检索手段。以提高其检索性能为目的,本文利用 Petri 网进行月

球影像检索过程分析与建模以保证其合理性,据此提出的基于内容的月球影像检索并行框架能够以较低廉的成本有效地提升基于内容的月球影像检索整体效率。后续研究工作将更多关注多节点负载均衡问题,考虑通过实验构建多节点负载模型,在此基础上设计一种合理的数据划分机制,进一步提升查询检索的效率。

致谢:感谢中国科学院国家天文台提供的嫦娥二号月球遥感影像数据作为实验数据来源之一。

参考文献 (References)

- [1] Meyer C, Deans M. Content based retrieval of images for planetary exploration [C]//Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007: 1377 - 1382.
- [2] Barlas G. An analytical approach to optimizing parallel image registration retrieval [J]. IEEE Transactions on Parallel and Distributed Systems, 2010, 21(8): 1074 - 1088.
- [3] Zhang J, Liu X, Luo J, et al. DIRS: distributed image retrieval system based on mapreduce [C]//Proceedings of IEEE International Conference on Pervasive Computing and Application, 2010: 93 - 98.
- [4] Vlachou A, Doulkeridis C, Mavroeidis D, et al. Designing a peer-to-peer architecture for distributed image retrieval [C]//Proceedings of International Conference on Adaptive Multimedia Retrieval: Retrieval, User and Semantics, 2008: 182 - 195.
- [5] Schweitzer H. A distributed algorithm for content based indexing of images by projections on ritz primary images [J]. Data Mining and Knowledge Discovery, 1997, (1): 375 - 390.
- [6] Yadav K, Srivastava A, Ansari M A. GPU-parallel implementation of color based medical image retrieval in compressed domain [J]. IJCA Special Issue on "Novel Aspects of Digital Imaging Applications", 2011: 8 - 14.
- [7] Montagnata J, Duquel H, Pierson J M, et al. Medical image content-based queries using the grid [C]//Proceedings of the first European HealthGrid conference, Lyon, France, 2003: 142 - 151.
- [8] Premchaiswadi W, Tungkatsathan A. On-line content-based image retrieval system using joint querying and relevance feedback scheme [J]. WSEAS Transactions on Computers, 2010, 5(9): 465 - 474.
- [9] 袁崇义. Petri 网原理与应用 [M]. 北京: 电子工业出版社, 2005.
YUAN Chongyi. Principles of Petri nets [M]. Beijing: Electronic Industrial Press, 2005. (in Chinese)
- [10] 李文, 高鹏, 陈英武, 等. 遥感数据处理任务调度的 Petri 网模型及求解算法 [J]. 国防科技大学学报, 2011, 33(1): 138 - 142.
LI Wen, GAO Peng, CHEN Yingwu, et al. A petri net model and algorithm for remotely sensed data processing task scheduling problem [J]. Journal of National University of Defense Technology, 2011, 33(1): 138 - 142. (in Chinese)
- [11] Hu M K. Visual pattern recognition by moment invariants [J]. IRE Transactions on Information Theory, 1962, (8): 179 - 187.
- [12] Tamura H, Shunji M, Yamawaki T. Textural features corresponding to visual perception [J]. IEEE Transactions on Systems, Man, And Cybernetics, 1978, 8(6): 460 - 473.
- [13] Bay S H, Tuytelaars T, Gool L V. Surf: speeded up robust features [C]//Proceeding of European Conference on Computer Vision, 2006: 404 - 417.