

Dados Preliminares: Um Experimento de Paralelização para a Segmentação de Imagens Landsat-TM

Silene de Freitas Fernandes*[#]
silene@univap.br

Antônio Miguel V. Monteiro[#]
miguel@dpi.inpe.br

* Universidade do Vale do Paraíba
Instituto de Pesquisa e Desenvolvimento
Av. Shinshima Hifumi 2911- Urbanova
São José dos Campos - SP
Tel (012) 349-100

[#] Instituto Nacional de Pesquisas Espaciais
Departamento de Processamento de Imagens
Av. Dos Astronautas - Jd. da Granja
São José dos Campos - SP
Tel (012) 345-6000

junho, 1997

Resumo

Este trabalho, pretende avaliar, propor e testar algumas soluções computacionais buscando a redução do tempo de processamento empreendido na fase de segmentação, durante a classificação digital de imagens orbitais, proposta para o problema de desflorestamento da Amazônia Legal e apresentadas em BATISTA, 1996 e SHIMABUKURO, 1997. Para isso explorar-se-á o paralelismo inerente ao algoritmo de Crescimento de Regiões, usado no suporte a estas metodologias e implementado no Sistema de Processamento de Informações Georeferenciadas - SPRING, escrevendo versões paralelas para o mesmo, utilizando as bibliotecas de troca de mensagem Parallel Virtual Machine - PVM e Message Passing Interface - MPI, buscando gerar uma solução independente de hardware específico, mais flexível e de fácil porte. Com isso pretendemos demonstrar a viabilidade do uso em programas das soluções digitais propostas, e investigar o uso de Computação de Alto Desempenho baseado no modelo de troca de mensagens para aplicações de processamento de imagens geradas por plataformas orbitais.

Abstract

The aim of this research project are to evaluate, propose and test a computational solutions, that could cut down on the processing time spent over the segmentation phase, part of the digital classification methodology proposed in BATISTA, 1996 and SHIMABUKURO, 1997 for the deforestation problem over the legal amazonian area, for the deforestation problem over the legal Amazonian Area, trying to make viable, from a computational perspective, the operational use of the digital solutions proposed. The basic idea is to recover the inherent parallelism of the Region Growing algorithm used for the segmentation and implemented on the SPRING - Sistema de Processamento de Informações Georeferenciadas, by writing parallel versions for this algorithm, using the message passing model based on the libraries PVM - Parallel Virtual Machine - and MPI - Message Passing Interface -, libraries, offering this way a solution that although can not be the best in performance is hardware independent, more flexible and easily portable. An investigation on the use of High Performance Computing based on the message passing model for applications that make intensive use of orbital image processing is another outcome of this research project.

1. INTRODUÇÃO

1.1 Solução Automática: Um Desafio Computacional

Os imageadores orbitais e os imageadores embarcados têm sido cada dia mais explorados na busca de aquisição das informações necessárias ao planejamento estratégico e a formulação de políticas de desenvolvimento racionais. Recuperar destas imagens a informação de interesse, requer capacidades de processar, analisar e interpretar esta massa de dados digitais. O tratamento automático, pode reduzir custos, e a modificação algorítmica de alguns procedimentos, pode reduzir o tempo de processamento para estas etapas viabilizando a periodicidade necessária a solução dos problemas.

Metodologias que fazem uso de imageamento remoto, usam frequentemente a fotointerpretação, onde o especialista interpreta a cena com base num produto fotográfico. O procedimento de classificação automática objetiva "automatizar" a etapa de fotointerpretação, fazendo com que a interpretação de toda cena se baseie nas propriedades de cada pixel e/ou de seus vizinhos com a mínima intervenção possível para o operador. O desafio computacional está em gerar estas soluções automáticas, usando imagens de satélites e/ou imagens de sensores embarcados, com a eficiência necessária para o suporte à implantação de programas "operacionais", com metodologias que façam uso efetivo de imageamento remoto e processamento digital.

1.2 Um Problema Real e Propostas de Solução Digital

Recentes levantamentos sobre a evolução do desflorestamento bruto na Amazônia estão sendo feitos pelo Instituto Nacional de Pesquisas Espaciais - INPE, em conjunto com a Fundação de Ciência, Aplicação e Tecnologia Espaciais - FUNCATE, através do Projeto de Levantamento das Áreas Desflorestadas na Amazônia Legal - PRODES [7]. O PRODES tem como objetivo identificar e mapear os incrementos de desflorestamentos de origem antrópica e gerar um banco de dados sobre as áreas desflorestadas. Todavia, devido aos problemas de georeferenciamento da interpretação visual sobre imagens de diferentes anos, esse objetivo tem sido comprometido.

Portanto, é pretendida uma nova metodologia, que faça uso efetivo do dado digital, na análise e geração dos mapas, potencializando a obtenção dos resultados finais com mais agilidade e precisão. As metodologias descritas em [2] e [3] têm por objetivo "automatizar" o processo de interpretação, minimizando as etapas de análise e interpretação visual, apresentando uma alternativa de menor custo de processamento e de melhor precisão, implicando um ganho de qualidade para o PRODES. Ambas metodologias apresentam as mesmas fases no processo de classificação da imagem, mas elas se diferenciam na forma de entrada dos dados. Enquanto [2] utiliza como entrada imagens Landsat-TM, utilizando 3 bandas com a resolução de 30x30m. A metodologia exposta em [3] utiliza 3 bandas da imagem Landsat-TM, para gerar 1 banda, que é a banda de mistura, reamostrada de 30 para 60m. Em [3], o pré-processamento das 3 bandas visando gerar uma imagem fração sombra, além de proporcionar um grande contraste entre as áreas florestas e áreas desflorestadas, também reduz consideravelmente o número de "pixels" necessários à análise da imagem.

As fases de segmentação e classificação empregam os algoritmos implementados no SPRING [10]. A fase de segmentação se baseia no algoritmo de crescimento de regiões, o qual agrega "pixels" com propriedades similares. A Tabela 1.1 mostra que o tempo de segmentação utilizando a imagem sombra foi menor.

Imagem	Número de Pixels	Nº de Bandas	Equipamento	Tempo de Segmentação (h)
Landsat-TM	2700x1700	3	IBM/RISC 600-530, 96 Mb de RAM	24
Landsat-TM	6100x6100	3	SUN SPARCstation - 20 , 96 Mb RAM	35
Sombra	3072x3326	1	SUN SPARCstation - 20 , 96 Mb RAM	6

Tabela 1.1 - Comparação do tempo de Segmentação nas Metodologias descritas em [2] e [3].

A Tabela 1.1 ilustra que o tempo de processamento para a segmentação da imagem sombra foi de 6h, enquanto que a segmentação utilizando as bandas 3, 4 e 5 do Landsat-TM foi de 24h para parte de uma cena descrita em [2] e 35h para a cena inteira utilizada em [3].

Nosso trabalho visa uma melhora significativa no tempo de processamento envolvido na fase de segmentação do processo de classificação utilizado nas metodologias propostas em [2] e [3].

1.3 Objetivos

Baseado nas soluções propostas em [2] e [3] vimos que a fase de segmentação é componente necessária e merece uma atenção especial, pois é uma fase que consome muito tempo de processamento. Nosso objetivo principal é reduzir este tempo, explorando o paralelismo inerente ao algoritmo de Crescimento de Regiões, visando potencializar o uso das metodologias descritas em [2] e [3]. O ponto inicial é o estudo detalhado do algoritmo de segmentação [1], [2], [4],[8] já implantado no SPRING [4], [9], [10], com o objetivo de avaliar os laços críticos, analisar sua complexidade, explorar o paralelismo avaliado [8] e escrever versões paralelas através da biblioteca de troca de mensagem Parallel Virtual Machine Enhanced - PVMc [6], [12] e uma proposta para o Parallel Environment - PE [13].

Toda implementação paralela e testes de desempenho serão feitos num IBM - SP-modelo Class 305 ("Single Staged Switching, 2-80 processor nodes"), com 4 nós e uma "High Performace Switch", instalado na Universidade do Vale do Paraíba - UNIVAP.

2. O PROBLEMA DE CLASSIFICAÇÃO E SEGMENTAÇÃO DE IMAGENS

2.1 Os Problemas da Classificação Automática

Dentre as diversas abordagens para a criação de um procedimento automatizado de classificação, encontra-se uma que divide o problema em duas partes. A primeira parte engloba os procedimentos que buscam identificar regiões homogêneas nas imagens. A segunda parte engloba as estratégias que buscam rotular, ou seja, estabelecer o significado de cada uma das regiões já identificadas. Na fase da identificação das regiões homogêneas, existe uma operação cuja finalidade é particionar a imagem, gerando partições que tenham uma correlação muito forte com os objetos ou áreas reais presentes na cena representada pelas imagens. Este procedimento para o particionamento da

imagem é conhecido como Segmentação da Imagem. O procedimento de buscar um particionamento da imagem em regiões, as quais devem ser compostas por "pixels" que apresentam conformidade com algum critério de similaridade, gera a operação conhecida como Segmentação por Regiões.

2.2 O Algoritmo de Segmentação por Crescimento de Regiões

Crescimento de Regiões é uma classe de técnicas usadas nos algoritmos de segmentação de imagens em que tipicamente, regiões são construídas por um processo de aglomeração que une "pixels" a regiões quando são adjacentes e tem propriedades similares. Cada "pixel" da imagem recebe um rótulo através do processo de Crescimento de Regiões. "Pixels" terão o mesmo rótulo se e somente se eles pertencerem a mesma região [1]. De acordo com [5], as regiões têm que ter as seguintes propriedades:

1) a região tem que estar conectada, isto é, para qualquer par de "pixels" pertencendo a uma região, X_i e X_j , deve haver uma sequência $\{X_p, X_q, X_{k+1}, \dots, X_j\}$ tal que X_k e X_{k+1} sejam vizinhos para uma vizinhança estabelecida e todos os "pixels" da sequência pertençam a região;

2) $I = \cup R_k$, I é toda imagem e R_k é a k^{th} região;

3) $R_i \cap R_j = \emptyset$ (não tem sobreposição)

4) Se $H(R)$ é uma função booleana para medir a homogeneidade da região, então

4.1) $H(R_i) = \text{verdadeiro}$

4.2) $H(R_i \cup R_j) = \text{falso}$

Vários critérios de similaridade podem ser usados para agregar "pixels" ou regiões, no algoritmo implementado no SPRING [4] e utilizados em [2] e [3], duas regiões R_i e R_j poderiam ser agregadas se:

1) $R_j = vp(R_i)$

2) $R_j = vp(R_i)$;

3) $dist(R_i, R_j) < \text{limiar de similaridade}$

onde: vp é o Vizinho mais Próximo,

$dist(R_i, R_j)$ é a distância Euclidiana entre a média espectral de R_i e R_j e o

$vp(R) = R_k / dist(R, R_k) = \min \{dist(R, R')\}$

onde R' pertence a vizinhança de R

O algoritmo de segmentação tem os seguintes parâmetros

<similaridade>: menor limiar a partir do qual duas regiões são consideradas similares e portanto, devem ser agrupadas. Na implementação do SPRING, o critério de similaridade é a distância Euclidiana entre as médias espectrais das regiões envolvidas no teste;

<área>: a menor área a ser considerada como uma região, definida em número de "pixels".

3. COMPUTAÇÃO DE ALTO DESEMPENHO PARA PROCESSAMENTO DE IMAGENS

Processamento digital de imagens possui tarefas inerentemente paralelas, devido ao tamanho e a regularidade das estruturas dos dados envolvidos, os quais tipicamente consistem de grandes matrizes bidimensional de elementos ("pixels"), e por causa das características paralelas dos algoritmos usados, tipicamente consistindo de execuções repetidas (laços) do mesmo conjunto de operações em todos os "pixels" e/ou usam da combinação de cada "pixel" com um pequeno conjunto de "pixels" localizados a uma distância pequena. Portanto, os laços podem ser o ponto de partida para se tentar reduzir o tempo da fase de segmentação da classificação digital de uma imagem, pois podemos optar pelo particionamento da cena e a consequente distribuição das partes em diferentes processadores, os quais analisariam suas partes concorrentemente.

4. ETAPAS DO TRABALHO

estudo da arquitetura SP-2	
estudo da bibliotecas PVMe e PE	
aquisição e estudo do programa fonte	
compilação e execução do fonte serial no SP	
tomada de tempo para área piloto	
análise dos dados e tarefas	
análise dos trechos paralelizáveis	
análise dos laços de código paralelizáveis	
análise da granularidade	
estudo de representações	
implementação do código usando PVMe	
tomadas de tempo variando o número de processadores com PVMe	
avaliação comparativa da nova metodologia com as propostas em [2] e [3]	
verificação da viabilidade operacional	
proposta de implementação usando PE	

<input type="checkbox"/>	tarefa completada
<input type="checkbox"/>	tarefa em andamento

5. REFERÊNCIAS BIBLIOGRÁFICAS

1. BADER, D.A. et all. *Parallel Algorithms for Image Enhancement and Segmentation by Region Growing with an Experimental Study*. University of Maryland, May, 1995
2. BATISTA, G. T. et all. *A New Approach for Deforestation Assessment*. International Geoscience and Remotely-Sensing Symposium, IGARSS'96, 1996.
3. SHIMABUKURO, Y. E. et all. *Segmentação e Classificação da Imagem Sombra do Modelo de Mistura para Mapear Desflorestamento na Amazônia*. Instituto Nacional de Pesquisas Espaciais, INPE-6147-PUD/029, Maio, 1997.
4. BINS, L. S. et all. *Um Método de Classificação Não Supervisionado por Regiões*. Simpósio Brasileiro de Computação Gráfica -SIBGRAPI V, Anais, p.65-68, 1992.
5. BROWN, C. M., BALLARD, D. H. *Region Growing*. Prentice Hall Inc, 2ª ed., 1982.
6. GEIST A. et all. *PVM 3 User's Guide and Reference Manual*. U.S. Department of Energy, the National Science Foundation, and the State of Tennessee, May, 1994.
7. ALVES, D. S. et all. *The Amazonia Information System*. Washington, D.C., ISPRS Archives, Commission VI, Anais, p.259-266, 1992.
8. LILJA, D. J. *Exploiting the Parallelism Available in Loops*. Computer, vol. 27, nº2, p.13-26, February, 1994.
9. MONTEIRO, A. M. V. *PARIMA: Paralelismo e Imagens. Paralelização de Alguns Algoritmos Básicos para a Classificação Automática de Imagens Digitais Geradas por Satélites de Recursos Naturais, Radars Orbitais e Radars Embarcados*. Proposta de Projeto para o CNPQ. Instituto Nacional de Pesquisas Espaciais, 1997.
10. INPE-DPI *SPRING, Manual do Usuário*. São José dos Campos, 1996. <http://www.inpe.br/spring>
11. MPI: *A Message Passing Interface Standard*. Message Passing Forum, June, 1996.

12. *PVMc: Optimized PVM on the SP System*. http://www.rs6000.ibm.com/software/sp_products/pvmc.html
13. IBM Parallel Environment for AIX: Hitchhiker's Guide. IBM Corporation, November, 1996.