

# Agentes de Mineração de Imagens de Satélite

Ciro D. G. Moura<sup>1</sup>, Nicksson C. A. Freitas<sup>2</sup>, Marcelino P. S. Silva<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação, Universidade do Estado do Rio Grande do Norte (UERN) / Universidade Federal Rural do Semi-Árido (UFERSA) – Mossoró, RN – Brasil

<sup>2</sup>LES – Laboratório de Engenharia de Software, Departamento de Informática, Universidade do Estado do Rio Grande do Norte (UERN) – Mossoró, RN – Brasil

{ciro.dgm, nickssonarrais, prof.marcelino}@gmail.com

**Abstract.** *Image Mining is a field with huge potential and relevant challenges. For satellite images, this technique and resources as well algorithms can bring appropriate responses to important problems. However, due to this approach limitations, multiagent systems present features that, properly applied, can bring advances on pattern analysis in satellite images. In this context, the goal of this research is to present a methodology and a software for satellite image mining through multiagent systems. The prototype is being developed in Java and uses resources from TerraView and GeoDMA.*

**Resumo.** *Mineração de Imagens é uma área com grande potencial e relevantes desafios. No caso de imagens de satélite, esta técnica e seu conjunto de recursos e algoritmos podem viabilizar soluções e respostas adequadas a importantes problemas. Entretanto, devido às limitações desta abordagem, sistemas multiagentes possuem características que, adequadamente aplicadas, podem trazer avanços significativos na análise de padrões em imagens de satélite. Neste contexto, o objetivo deste trabalho é apresentar uma metodologia e respectiva implementação computacional para a mineração de imagens de satélite através de sistemas multiagentes. O protótipo está sendo desenvolvido em Java e utiliza funcionalidades do TerraView e o do GeoDMA.*

## 1. Introdução

A mineração de dados é uma das etapas no processo de Descoberta de Conhecimento em Banco de Dados (DCBD). Essa etapa é responsável pela aplicação de algoritmos específicos para extrair padrões de dados [Fayyad, 1997].

Na mineração de imagens, as imagens de um acervo (banco de imagens) são recuperadas segundo critérios inerentes à aplicação. A seguir, uma fase de pré-processamento aumenta a qualidade dos dados, os quais são então submetidos a uma série de transformações e de extração de características que geram importantes informações a respeito das imagens. A partir destas informações, a mineração pode ser realizada através de técnicas específicas, com o intuito de descobrir padrões significativos. Esses padrões resultantes são então avaliados e interpretados para a

obtenção do conhecimento final, que pode ser aplicado no entendimento de problemas, na tomada de decisões ou outras aplicações estratégicas [Silva, 2006].

Sistemas de agentes ou multiagentes e o processo de DCBD são duas áreas de pesquisa cada vez mais inter-relacionadas, que proporcionam benefícios para ambas as partes. Para Silva e Ralha (2011), essas duas áreas, a princípio, seguem com metas e objetivos distintos, porém há vários aspectos de ambas as áreas que coincidem, tais como: interação usuário-sistema, papéis humanos, modelagem dinâmica, fatores de domínio, fatores organizacionais e sociais, entre outros.

Nesse contexto, o objetivo deste trabalho é apresentar uma metodologia e respectiva ferramenta para a mineração de imagens através de sistemas multiagentes. O protótipo foi desenvolvido em Java e trabalha em conjunto com o software TerraView e o seu plugin, o GeoDMA.

## **2. Trabalhos Relacionados**

A interação entre mineração de dados e agentes é aplicada em diferentes campos de pesquisa. Apresentaremos em seguida alguns trabalhos relacionados ao tema.

Silva (2011) apresenta um protótipo do AAgent Mining Integration (AGMI) que foi testado com dados reais de licitações extraídos do Sistema ComprasNet. Vários experimentos foram realizados explorando aspectos de distribuição do processamento e autonomia dos agentes.

Pultar, Raubal, e Goodchild, (2008) trazem o projeto e a implementação de um protótipo que lê os dados de páginas na web e segue links para adquirir conhecimento. O agente cria um banco de dados com textos de páginas da web, minera informações de localização e então converte para um formato de dados geoespaciais.

Xavier Júnior (2012) propôs um Sistema Multiagente que permite que os usuários possam guardar as informações visualizadas em seus históricos de acesso, e baseado nesse histórico, o sistema cria agrupamentos de informações relevantes através do uso de métodos de agrupamento de dados relacionais. Com as informações armazenadas no perfil dos usuários, os agentes podem interagir e recomendar informações geográficas relevantes aos usuários.

Todos esses trabalhos exemplificam o uso de agentes de mineração de dados, mas nenhuma das metodologias é aplicada a mineração de imagens de satélite. Tendo como base conceitos abordados nesses trabalhos, este trabalho apresenta uma metodologia diferente.

## **3. Técnicas e Ferramentas**

A metodologia proposta e descrita na próxima seção utiliza técnicas e ferramentas de mineração de dados geográficos e de sistemas multiagentes.

### **3.1 GeoDMA**

A ferramenta utilizada para minerar imagens é o GeoDMA (Geographical Data Mining Analyst). Ela realiza todas as fases de processamento necessárias para manipular dados de sensoriamento remoto, incluindo os processos de segmentação, extração e seleção de

atributos, treinamento, classificação e análise exploratória dos dados. Segundo Korting et al. (2009), o GeoDMA funciona como um plugin para o sistema TerraView. Isto significa que toda a estrutura que manipula e visualiza bancos de dados geográficos é proporcionada pelo TerraView. Então o GeoDMA é executado em conjunto com o TerraView, e produz os resultados que são exibidos na sua tela principal.

Dessa forma, uma imagem é inserida no Terraview e, utilizando o GeoDMA, é possível segmentar a imagem e extrair atributos espaciais e espectrais de cada um desses segmentos. Após a etapa de extração, na etapa de treinamento seleciona-se alguns dos segmentos e seus respectivos atributos, rotulando cada um desses segmentos com uma determinada classe. Com as amostras selecionadas nesta etapa, o próximo passo é a classificação, cujo algoritmo escolhido para este trabalho foi a árvore de decisão C4.5 [Quinlan, 1993], uma vez que o GeoDMA pode carregar uma árvore já existente ou salvar uma nova para que ela possa ser usada em futuras classificações.

### **3.2 Sistemas Multiagentes**

Um agente é capaz de perceber seu ambiente por meio de sensores e de agir sobre este ambiente por intermédio de atuadores. Os quatro tipos básicos de agentes são o agente reativo simples, o agente reativo com estado, o agente baseado em objetivo e o agente baseado em utilidade [Russel e Norving, 2010]. Um Sistema Multiagente (SMA) consiste de um número de agentes que interagem uns com os outros de forma cooperativa ou competitiva. No caso mais geral, os agentes em um SMA estarão representando ou agindo em nome de usuários ou proprietários com diferentes objetivos e motivações [Wooldridge, 2009].

Segundo Sycara (1998), as características de SMA são tais que: cada agente possui uma visão limitada; não há controle global do sistema; trabalham com dados descentralizados; a computação é assíncrona, de maneira a permitir a comunicação entre entidades heterogêneas. Para Weiss (2000) e Wooldridge (2009) uma das razões para o desenvolvimento da área é que os conceitos subjacentes à SMA não são restritos a um único domínio de aplicação, sendo interessantes no desenvolvimento e análise de modelos e teorias de interatividade nas sociedades humanas. Isto ocorre devido ao SMA ser considerado uma metáfora natural para o entendimento e construção de uma ampla faixa do que podem ser rudemente denominados de sistemas artificiais sociais.

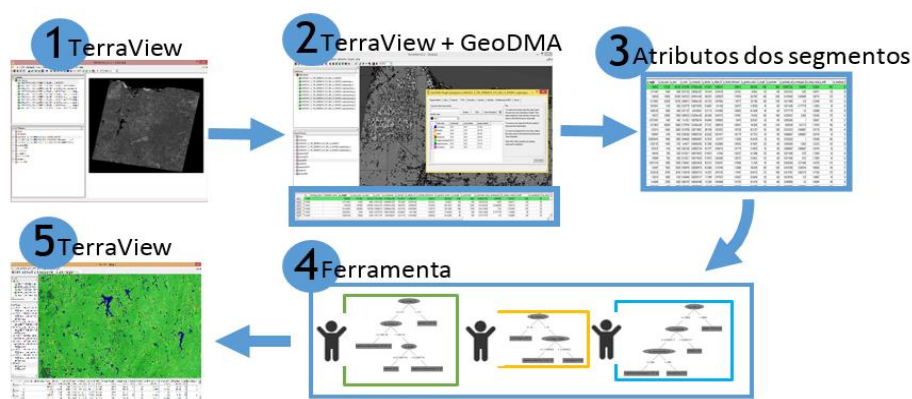
Aplicando os conceitos de agentes e SMA neste trabalho, cada agente realiza uma busca por padrões nas imagens de acordo com seu perfil e, dessa forma, o SMA minera a imagem como um todo.

## **4. Metodologia**

O GeoDMA realiza todo o processo para classificar uma única imagem. Visando superar esta limitação, esta proposta com sistemas multiagente viabiliza um processo articulado e cooperativo. Uma vez que o GeoDMA segmenta a imagem e extrai os atributos de cada um dos segmentos, armazenando-os no banco de dados, o protótipo realizará a mineração através dos agentes segundo a árvore de decisão gerada.

A Figura 1 apresenta o diagrama do processo. (1) O TerraView gerencia o banco de dados e manipula as imagens; (2) O GeoDMA segmenta as imagens e extrai os atributos espaciais e espectrais de cada um dos segmentos; (3) Esses atributos são

armazenados no banco de dados; (4) A ferramenta então é conectada ao mesmo banco de dados utilizado pelo TerraView, e as tabelas que possuem os atributos obtidos são selecionados para criar o ambiente dos agentes; neste ponto, o usuário cria ou escolhe os agentes que serão utilizados visando atingir os seus resultados esperados. Cada agente possui sua árvore de decisão para que os segmentos possam ser minerados pelo mesmo. (5) Finalmente, com os dados classificados, o usuário visualiza o resultado da mineração no TerraView.



**Figura 1. Diagrama do processo**

Os agentes são do tipo reativos simples, ou seja, são agentes que selecionam ações com base na sua percepção atual [Russel e Norving, 2010]. A partir da sua árvore, cada agente possuirá um perfil que lhe permitirá minerar padrões específicos em novas imagens. Então, digamos que o agente A1 tenha o perfil de minerar corpos d'água. Na interface do protótipo o usuário terá recursos para selecioná-lo e atribuir ao mesmo a tarefa de minerar em novas imagens objetos que sejam mananciais, lagoas, açudes ou outros objetos pertinentes ao seu perfil.

Dessa forma, o agente A1 é capaz de identificar todos os segmentos que satisfaçam sua árvore de decisão, superando a limitação do GeoDMA, pois os agentes não minerarão apenas os segmentos de uma imagem, mas todos os segmentos das imagens que forem alocadas em seu ambiente.

## 5. Estudo de caso

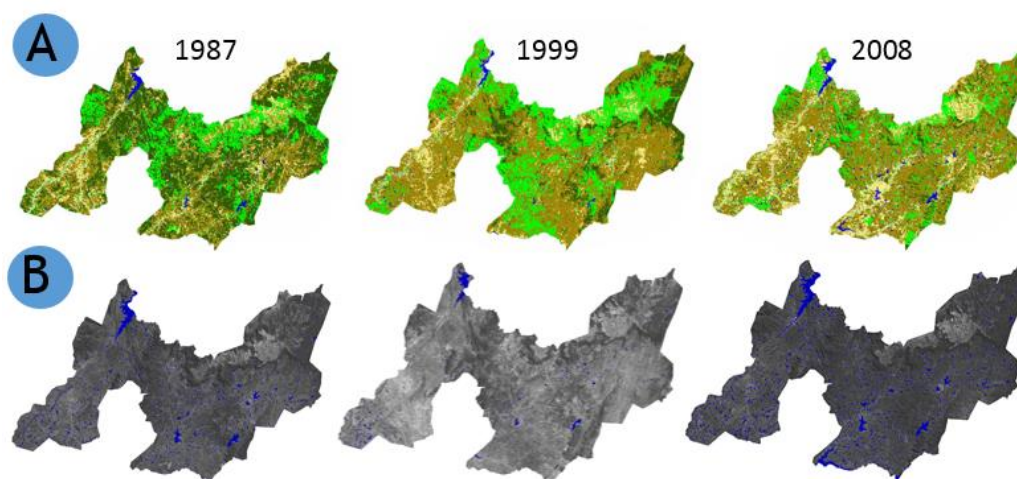
Pereira (2010) realizou um estudo que tinha como objetivo mapear e identificar as áreas sujeitas ao processo de desertificação no estado do Rio Grande do Norte. Para isso, usou-se uma série histórica de imagens do satélite Landsat 5 TM dos anos de 1987, 1999 e 2008, e a cena 215/64. Foi realizado um recorte abrangendo 13 municípios: Jucurutu, Timbaúba dos Batistas, São Fernando, Florânea, Cruzeta, São José do Seridó, Acari, São Vicente, Tenente Laurentino Cruz, Lagoa Nova, currais novos, Bodó e Cerro Corá. Esses municípios foram escolhidos pois pertencem a uma região de grande susceptibilidade ao fenômeno da desertificação.

O software utilizado foi SPRING. Para segmentar as imagens, foi utilizada a técnica de crescimento de regiões, a qual agrega pixels com propriedades similares. Os valores de similaridade e área utilizados foram 10 e 15, respectivamente. Na etapa de classificação foi utilizado o classificador do tipo supervisionado "Bathacharia" e as seis

classes apresentadas nas imagens foram: caatinga preservada, caatinga degradada, solo exposto, lavoura, mata ciliar e corpos d'água.

Como forma de demonstrar a metodologia e o protótipo desenvolvido, realizamos uma comparação com o estudo apresentado por Pereira (2010), mas buscando padrões que indiquem a existência de corpos d'água. Para isso, utilizou-se a mesma série de imagens, obtidas no Banco de Imagens da DGI/INPE (Divisão de Geração de Imagens do Instituto Nacional de Pesquisas Espaciais. DGI, 2011). Apenas a banda 5 foi selecionada, pois corpos d'água são diferenciados de áreas com vegetação, áreas de cultivo, bem como solo exposto.

Utilizando o GeoDMA, as imagens foram segmentadas com a mesma técnica de crescimento de regiões e os mesmos valores para similaridade e área, a diferença fica na parte do treinamento, que é realizado apenas na imagem de 1987. Após a etapa de treinamento, uma árvore de decisão é gerada pelo GeoDMA para classificar a imagem de 1987. Essa árvore gerada serviu de perfil para um agente que foi o responsável pela classificação das outras duas imagens. A Figura 2 apresenta a comparação visual entre as imagens do trabalho de Pereira (2010) (A) e os resultados obtidos neste trabalho (B).



**Figura 2. Resultado da classificação na área de corte: (A) Imagens adaptadas do trabalho da Pereira (2010) e (B) e imagens mineradas pelo agente.**

Mesmo tendo utilizado softwares diferentes e diferentes amostras na etapa de treinamento, os resultados foram satisfatórios. Observa-se na Tabela 1 que a porcentagem de área ocupada pela classe água é bem próxima.

**Tabela 1. Porcentagem da área ocupada pela classe água.**

	1987	1999	2008
Pereira (2010)	1.94%	1.41%	3.12%
Resultados obtidos	2.09%	1.99%	3.36%

Da mesma forma como foi possível classificar utilizando um agente, outros agentes poderiam ter sido criados com diferentes perfis (lavoura, solo exposto etc.) e usados para minerar as imagens.

## 6. Conclusão

Este trabalho propõe uma metodologia e ferramenta que utiliza as características de agentes para auxiliar no processo de extração de conhecimento de imagens de satélite. Esses agentes serão capazes de minerar dados geográficos a partir das características espaciais e espectrais dos objetos presentes em imagens de sensoriamento remoto. O SMA, através dos diferentes perfis dos seus agentes, terá a funcionalidade, após devidamente treinados, de minerar objetos de diferentes padrões (solo exposto, vegetação densa, corpos d'água, dentre outros) em um repositório de imagens.

A ferramenta ainda está em desenvolvimento e, os resultados parciais estão sendo obtidos e avaliados através de protótipo. Novos estudos de caso validarão a versão funcional da ferramenta e, conseqüentemente, a metodologia proposta.

## Referências

- DGI. DGI - Divisão de Geração de Imagens (INPE), 2011. Catálogo de Imagens. Disponível em: <<http://www.dgi.inpe.br/CDSR/>>.
- Fayyad, U. Data Mining and Knowledge Discovery in Databases: Implications for Scientific Database. Proc. of the 9th International Conference on Scientific and Statistical Database Management, pp. 2-11, 1997.
- Korting, T. S.; et al. GeoDMA - Um sistema para mineração de dados de sensoriamento remoto. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14, 2009, Natal. Anais do XIV Simpósio Brasileiro de Sensoriamento Remoto. Natal: INPE, 2009. p. 7813-7820.
- Pereira, G. O. Identificação de Áreas Suscetíveis à Desertificação no Rio Grande do Norte. Relatório final. (PIBIC/CNPq/INPE). INPE, 2010.
- Pultar, E., Raubal, M. e Goodchild, M. F.. 2008. GEDMWA: Geospatial Exploratory Data Mining Web Agent. In Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '08). ACM, New York, NY, USA, Article 74, 4 pages.
- Quinlan, J. R. C4.5: Programs for Machine Learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993. ISBN 1-55860-238-0.
- Russel, S. Norving, P. Artificial Intelligence: A Modern Approach. 3ª. Ed. Prentice Hall. 2010. ISBN-10: 0-13-604259-7.
- Silva, C. V. S. Agentes de Mineração e sua Aplicação no Domínio de Auditoria Governamental. 2011. 122f.. Tese (Mestrado em Informática) - UnB, Brasília, 2011.
- Silva, C. V. S. e Ralha, C. G. Detecção de Cartéis em Licitações Públicas com Agentes de Mineração de Dados. RESI - Revista Eletrônica de Sistema de Infomação. v. 10, n. 1, artigo 8. 2011. ISSN: 1677-3071.
- Silva, M. P. S. Mineração de dados em bancos de imagens. 2006. 123f.. Tese (Doutorado em Computação Aplicada) - INPE, São José dos Campos, 2006.
- Sycara, K. P.. Multiagent systems. AI Magazine, 19(2):79–92, 1998.
- Weiss, G. Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence. The MIT Press, 2000.
- Wooldridge, M. An Introduction to MultiAgent Systems. John Wiley & Sons, LTD, Chichester, England, 2nd edition, 2009.
- Xavier Júnior, J. C. NatalGIS: Um Sistema Multiagente de Recomendação de Informações Geográficas baseado em Agrupamento de Dados Relacionais. 2012. 189f.. Tese (Doutorado em Ciências) - UFRN, Natal, 2012.