

# ABORDAGEM MULTIAGENTE PARA MINERAÇÃO DE IMAGENS DE SATÉLITE

Ciro Daniel Gurgel de Moura  
Marcelino Pereira dos Santos Silva  
Nicksson Ckayo Arrais de Freitas

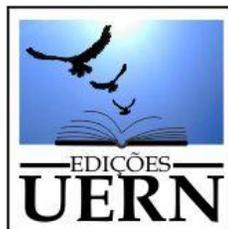


Ciro Daniel Gurgel de Moura  
UERN

Marcelino Pereira dos Santos Silva  
UERN

Nicksson Ckayo Arrais de Freitas  
UERN

# ABORDAGEM MULTIAGENTE PARA MINERAÇÃO DE IMAGENS DE SATÉLITE



**Reitor**

Prof. Pedro Fernandes Ribeiro Neto

**Vice-Reitor**

Prof. Aldo Gondim Fernandes

**Pró-Reitor de Pesquisa e Pós-Graduação**

Prof. João Maria Soares

**Comissão Editorial do Programa Edições UERN:**

Prof. João Maria Soares

Profa. Anairam de Medeiros e Silva (**Editora Chefe**)

Prof. Eduardo José Guerra Seabra

Prof. Humberto Jefferson de Medeiros

Prof. Sérgio Alexandre de Moraes Braga Júnior

Profa. Lúcia Helena Medeiros da Cunha Tavares

Prof. Bergson da Cunha Rodrigues

**Assessoria Técnica:**

Daniel Abrantes Sales

## **Catálogo da Publicação na Fonte. Universidade do Estado do Rio Grande do Norte.**

Moura, Ciro Daniel Gurgel de

Abordagem multiagente para mineração de imagens de satélite / Ciro Daniel Gurgel de Moura, Marcelino Pereira dos Santos Silva, Nicksson Ckayo Arrais de Freitas - Mossoró - RN: EDUERN, 2017.

92p.

ISBN: 978-85-7621-172-3

1. Sensoriamento Remoto - Tecnologia. 2. Banco de dados - Sensoriamento Remoto. 3. Mineração de dados - Bancos de imagens. I. Silva, Marcelino Pereira dos Santos. II. Freitas, Nicksson Ckayo Arrais de. III. Universidade do Estado do Rio Grande do Norte. IV. Título.

UERN/BC

CDD 621.3678

Bibliotecária: Aline Karoline da Silva Araújo CRB 15 / 783

*Ao meu Deus, meus pais, minha família e amigos.*

*Ciro Moura*

*A Deus, por sua graça e misericórdia, e à minha  
família, pelo amor e compreensão.*

*Marcelino Pereira*

*À minha família e aos meus amigos, pelos incentivos e  
pelo apoio constante.*

*Nicksson Freitas*

# Sumário

CAPÍTULO 1	INTRODUÇÃO.....	5
CAPÍTULO 2	REFERENCIAL TEÓRICO .....	7
2.1	SENSORIAMENTO REMOTO .....	7
2.2	MINERAÇÃO DE IMAGENS DE SENSORIAMENTO REMOTO.....	17
2.3	AGENTES DE MINERAÇÃO DE DADOS .....	29
2.4	TRABALHOS RELACIONADOS.....	41
CAPÍTULO 3	METODOLOGIA.....	45
3.1	SELEÇÃO DE DADOS DO REPOSITÓRIO DE IMAGENS DE SATÉLITES .....	46
3.2	SEGMENTAÇÃO E EXTRAÇÃO DE CARACTERÍSTICAS.....	52
3.3	AGENTGEO.....	53
3.4	AVALIAÇÃO E INTERPRETAÇÃO .....	61
CAPÍTULO 4	ESTUDOS DE CASO E RESULTADOS.....	63
4.1	ESTUDO DE CASO 1: NORDESTE .....	64
4.2	ESTUDO DE CASO 2: ESTADO DO RIO GRANDE DO NORTE .....	70
CAPÍTULO 5	CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS .....	82
	REFERÊNCIAS.....	84

## CAPÍTULO 1 INTRODUÇÃO

Uma das melhores formas de se obter informações sobre o nosso planeta é através dos dados obtidos pelos inúmeros satélites de sensoriamento remoto. E com a disponibilidade destes dados é possível realizar estudos em várias áreas, tais como agricultura, hídrica, urbana, florestal, meteorológica, limnológica, entre outras.

Com a enorme quantidade de informações que os satélites de sensoriamento remoto adquirem com o passar do tempo, os bancos de dados espaciais crescem continuamente, gerando gigantescos volumes de informações. Esses bancos de dados, também chamados de banco de dados de sensoriamento remoto, são uma coleção de imagens da paisagem, que fornecem uma oportunidade única de compreender como, quando e onde mudanças ocorreram em nosso planeta.

A análise desses bancos de dados espaciais, que são repositórios de imagens de sensoriamento remoto, demanda técnicas adequadas de exploração de dados para revelar conhecimento útil. Atualmente, tem-se uma capacidade limitada para extrair informações a partir destas bases de dados.

A mineração de dados em bancos de imagens vem ao encontro destas questões através de técnicas, algoritmos e sistemas capazes de extrair conhecimento (informações estratégicas), relacionamentos espaciais e padrões interessantes não explícitos em grandes acervos de imagens.

Os investimentos, a disponibilidade e a demanda sobre estes acervos de imagens, aliados à impossibilidade humana na análise manual e individual destes dados, motivam a pesquisa deste tema em direção ao fomento de novas ferramentas e técnicas para análise automática (ou semiautomática) e inteligente de bancos e acervos de imagens de sensoriamento remoto.

Um agente é tudo o que pode ser considerado capaz de perceber seu ambiente por meio de sensores e de agir sobre esse ambiente, por intermédio de atuadores. Suas características são: autonomia, interação, reação e iniciativa. Um sistema multiagente consiste de um número de agentes que interagem uns com os outros. Dessa forma, os agentes em um sistema multiagente estarão representando ou agindo em nome dos usuários com os mais diversos objetivos e motivações.

O volume de dados geográficos e o dinamismo da geração, transmissão e análise destes dados, aliados às potencialidades dos agentes e sistemas multiagente, desafia a integração dessas tecnologias para que haja um avanço nesta frente de pesquisa.

Dessa maneira, este livro tem como objetivo apresentar uma metodologia e respectiva implementação computacional, que forneçam meios para realizar a mineração de imagens de sensoriamento remoto através de sistemas multiagente, visando o avanço quantitativo e qualitativo do processo de descoberta de conhecimento nos repositórios de sensoriamento remoto.

## CAPÍTULO 2 REFERENCIAL TEÓRICO

Neste capítulo são apresentados os assuntos que serviram de base para o desenvolvimento deste livro. Inicialmente, na Seção 2.1 são abordados conceitos sobre Sensoriamento Remoto. Na Seção 2.2 é apresentada uma forma de se extrair conhecimento dos bancos de dados de Sensoriamento Remoto, abordando a Mineração de Imagens. Por fim, na Seção 2.3 descrevemos alguns conceitos sobre Agentes, Sistemas Multiagente (SMA) e Agentes de Mineração de Dados.

### 2.1 SENSORIAMENTO REMOTO

O primeiro satélite de sensoriamento remoto operacional (LANDSAT-1) foi lançado em 1972, desde então, tem havido uma grande experiência em nível mundial na coleta, processamento e análise dos dados de sensoriamento remoto (SILVA; CÂMARA; ESCADA, 2008).

Historicamente, reconhece-se que o termo Sensoriamento Remoto foi criado para designar o desenvolvimento dessa nova tecnologia de instrumentos capazes de obterem imagens da superfície terrestre a distâncias remotas. Por isso, a definição mais conhecida ou clássica de sensoriamento remoto é: Sensoriamento remoto é uma técnica de obtenção de imagens dos objetos da superfície terrestre sem que haja um contato físico de qualquer espécie entre o sensor e o objeto (MENESES; ALMEIDA, 2012).

Em CCRS (2007) há outra definição para sensoriamento remoto: Sensoriamento Remoto é uma área das ciências aplicadas que busca, a partir de um conjunto de atividades (detecção, aquisição e análise), obter informações dos objetos que compõem a superfície terrestre, isto sem a necessidade de contato direto com os mesmos. Essas atividades envolvem a interpretação e extração de informações da energia

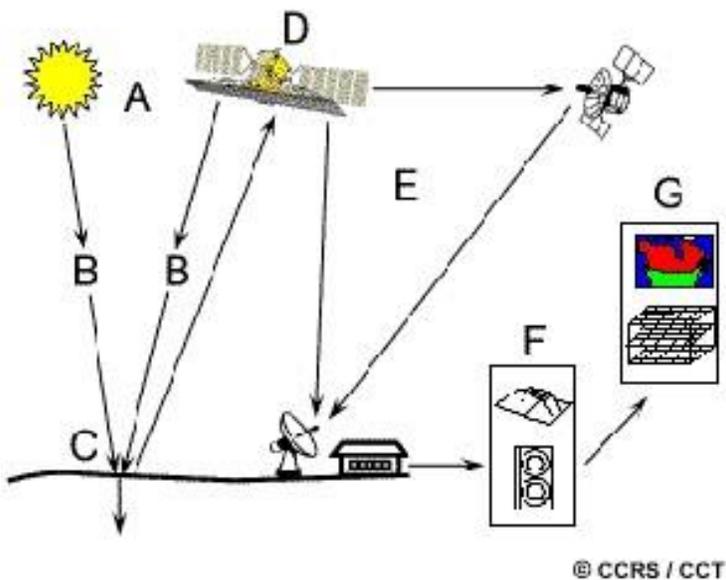
eletromagnética que é utilizada na obtenção dos dados por Sensoriamento Remoto e emitida ou refletida pelos objetos terrestres e, conseqüentemente, registrada por sensores remotos.

Pode-se perceber a importância do uso desses sensores remotos para adquirir e monitorar informações da superfície terrestre, pois, como Diógenes (2011) cita em seu trabalho: os sensores são capazes de capturar informações na paisagem não perceptíveis aos humanos, uma vez que gravam a energia emitida ou refletida pelos corpos em diferentes comprimentos de onda, dentro do espectro eletromagnético. Em sensoriamento remoto alguns elementos devem ser destacados, como sintetizado na Figura 1 (CCRS, 2007). A seguir serão listadas as principais etapas para aquisição e uso de imagens:

- (A) Fonte de energia ou iluminação** - um dos primeiros requisitos para obtenção de dados em sensoriamento remoto é a disponibilidade de uma fonte de energia que ilumina o objeto a ser observado;
- (B) Radiação e atmosfera** - a energia se desloca da sua fonte para os objetos monitorados e sofrerá interações com a atmosfera pela qual passa. Ao atingir a superfície, a energia será refletida pelos objetos e parte dela retornará ao sensor após interagir novamente com a atmosfera;
- (C) Interação com o objeto monitorado** - após a passagem pela atmosfera, a energia interage com o objeto. As interações dependem das características do objeto e da radiação;
- (D) Registro da energia pelo sensor** - após a reflexão da energia pelo objeto ou após a emissão pelo objeto, é utilizado um sensor para coletar e mensurar o fluxo da radiação eletromagnética;

- (E) **Transmissão, recepção, e processamento** - a energia registrada pelo sensor deve ser transmitida, geralmente na forma eletrônica, para uma estação de recepção e processamento. Nessa estação ocorrerá o processamento dos dados e, posteriormente, a geração de uma imagem;
- (F) **Interpretação e Análise** - a imagem processada é interpretada, visualmente e/ou digitalmente, para extração de informações sobre os objetos de interesse;
- (G) **Aplicação** - as informações geradas a partir das imagens processadas são utilizadas em diferentes tipos de estudos e aplicações.

**Figura 1 – Elementos de Sensoriamento Remoto.**



Fonte: Adaptado de CCRS (2007)

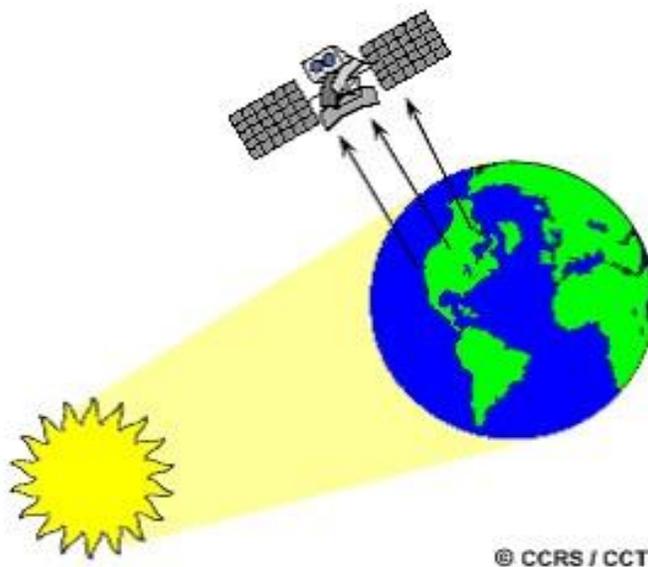
Os sensores remotos são dispositivos capazes de detectar a energia eletromagnética (em determinadas faixas do espectro

eletromagnético) proveniente de um objeto, transformá-las em um sinal elétrico e registrá-las, de tal forma que este possa ser armazenado ou transmitido em tempo real para posteriormente ser convertido em informações que descrevem as feições dos objetos que compõem a superfície terrestre. As variações de energia eletromagnética da área observada podem ser coletadas por sistemas sensores imageadores ou não-imageadores (MORAES, 2002).

Os sensores são classificados como ativos e passivos. Na Figura 2 é apresentado o sensor passivo, que não possui fonte própria de energia eletromagnética, como por exemplo os sensores do satélite LANDSAT-5, os radiômetros e espectrorradiômetros.

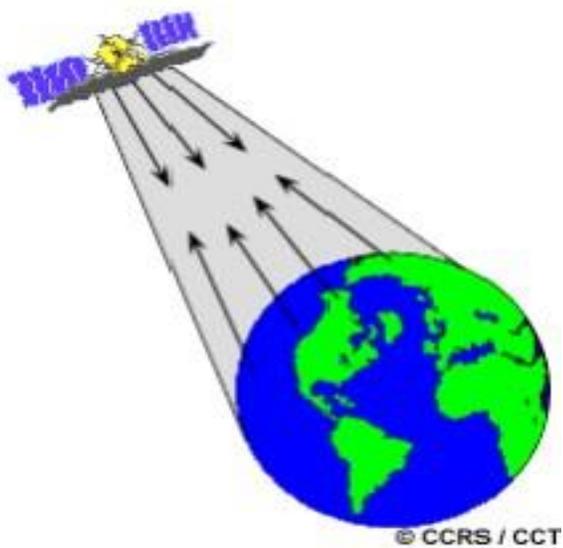
Os sensores ativos possuem uma fonte própria de energia eletromagnética, como podemos observar na Figura 3. Os mesmos emitem energia eletromagnética para os objetos terrestres a serem imageados e detectam parte desta energia que é refletida por estes na direção destes sensores. Como exemplo pode-se citar o radar e qualquer câmara fotográfica com flash. Os sistemas fotográficos foram os primeiros equipamentos a serem desenvolvidos e utilizados para o sensoriamento remoto de objetos terrestres (MORAES, 2002).

**Figura 2 – Demonstração dos sensores passivos.**



Fonte: (CCRS, 2007)

**Figura 3 – Demonstração dos sensores ativos.**



Fonte: (CCRS, 2007)

Outro ponto importante é a aquisição dos dados. Ao nível do solo, a aquisição é realizada em campo ou em laboratório cujas medidas são obtidas utilizando-se equipamentos radiômetros ou espectroradiômetros, essas informações podem ser comparadas com as informações recolhidas a partir de sensores em aeronaves ou satélites.

Ao nível de aeronaves, os dados são obtidos por meio de sistemas sensores de varredura óptico-eletrônico, sistemas fotográficos ou radares. Aviões ou helicópteros permitem capturar imagens mais detalhadas e também facilitam a captura de dados em qualquer parte da superfície terrestre a qualquer momento. A altura do voo do instrumento utilizado pode influenciar na resolução espacial dos dados, esse ponto será abordado mais adiante. Ao nível orbital, os sistemas sensores estão localizados em satélites, e este tipo de sensoriamento permite a repetição das informações, assim como um melhor monitoramento de grandes áreas da superfície terrestre.

Esses satélites podem ser orbitais ou estacionários (MORAES, 2002; CCRS, 2007). Na Figura 4 são apresentados os níveis orbitais e suborbital de coleta de dados de sensoriamento remoto.

**Figura 4 – Níveis de coleta de dados.**



Fonte: (MOREIRA, 2001)

Segundo CCRS (2007) e Meneses e Almeida (2012) no processo de aquisição de imagem, quatro conceitos são fundamentais: Resolução espacial, espectral, radiométrica e temporal. Descreveremos cada um desses conceitos a seguir:

1. A resolução espacial ou geométrica refere-se ao campo de visada instantânea (do inglês, *Instantaneous Field Of View* - IFOV), e indica o tamanho do pixel no terreno ou a menor área possível a ser abrangida por uma unidade de imagem. Quanto menor a resolução espacial maior será o detalhamento de objetos em cena, resultando em um conjunto maior de informações. Então se um sensor tem uma resolução espacial de vinte metros (20m), cada pixel representará uma área de 20m x 20m (20m<sup>2</sup>). A Figura

5 ilustra o contraste visual entre diferentes resoluções em áreas urbanas;

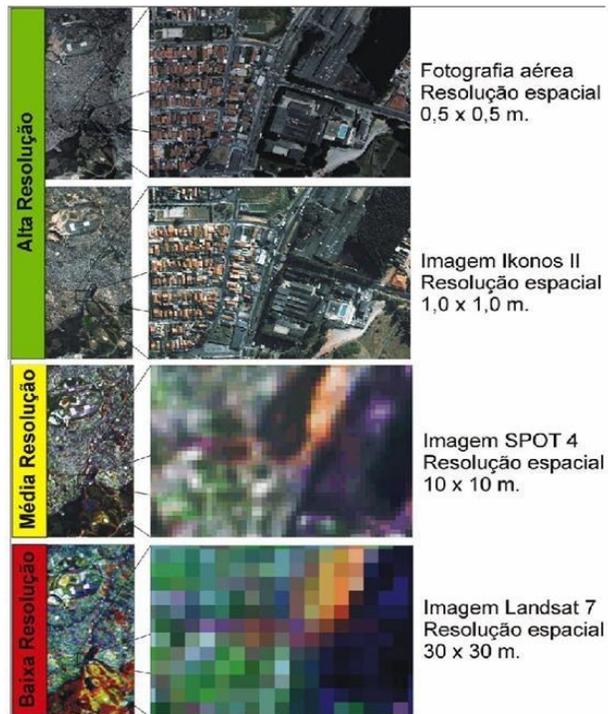
2. A resolução espectral determina a capacidade do sensor para definir curtos intervalos de comprimento de onda, ou seja, é uma medida da largura das faixas espectrais e da sensibilidade do sensor em distinguir entre dois níveis de intensidade do sinal de retorno; quanto melhor a resolução espectral, maior a quantidade de canais e mais estreita a faixa de comprimento de onda para um canal ou banda particular;
3. A resolução radiométrica descreve a capacidade do detector para medir as diferenças de intensidades dos níveis de radiação, definindo o número desses níveis que ele pode discriminar. Em termos práticos, a quantização do sinal é medida pelo intervalo de número de valores digitais usados para expressar os valores de radiação medidos pelo detector. A quantização é normalmente expressa em termos de números de dígitos binários (bits). Quanto maior é a quantização, maior a discriminação dos objetos e, conseqüentemente, maior será a qualidade visual da imagem, como apresentado na Figura 6. Nessa figura, observa-se que a imagem de 8 bits ( $2^8 = 256$ ) tem detalhes visuais melhores que as imagens de menores resoluções radiométricas de 6, 4 e 2 bits;
4. A resolução temporal determina o tempo necessário para o sensor voltar a um alvo específico e obter a imagem da mesma área, isto é, o tempo necessário para completar um ciclo de órbita. Este ciclo está relacionado às características orbitais da plataforma (altura, velocidade, inclinação) e ao ângulo total de abertura do sensor. Se um sensor é capaz de obter uma imagem de uma área a cada 16 dias, então sua resolução temporal é este período. A resolução temporal é de grande interesse

especialmente em estudos relacionados a mudanças na superfície terrestre e no seu monitoramento.

Imagens obtidas através da aquisição e processamento de sensoriamento remoto são usadas em muitos campos de pesquisa, uma vez que as informações fornecidas por elas são fortemente exigidas em muitas áreas. Para Sausen (2005) essas imagens são usadas por diversos profissionais em áreas como:

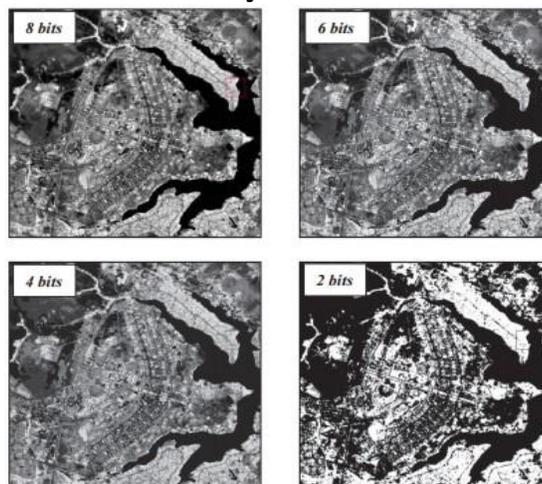
- Meteorologia: previsão do tempo, acompanhamento de mudanças atmosféricas, controle de poluentes, medição do efeito estufa e do buraco na camada de ozônio;
- Defesa Civil: previsão de catástrofes naturais que permitem a tomada de medidas preventivas;
- Planejamento e acompanhamento de culturas agrícolas;
- Planejamento e monitoramento do crescimento urbano;
- Monitoramento de áreas florestais para detecção de queimadas e outras formas de desmatamento;
- Usos militares: espionagem, acompanhamento de movimentação de inimigos e planejamento estratégico do posicionamento de tropas.

**Figura 5 – Imagens de diferentes sensores e resoluções em uma área urbana.**



Fonte: (MELO, 2002)

**Figura 6 – Exemplos de imagens com diferentes níveis de quantização ou de resolução radiométrica.**



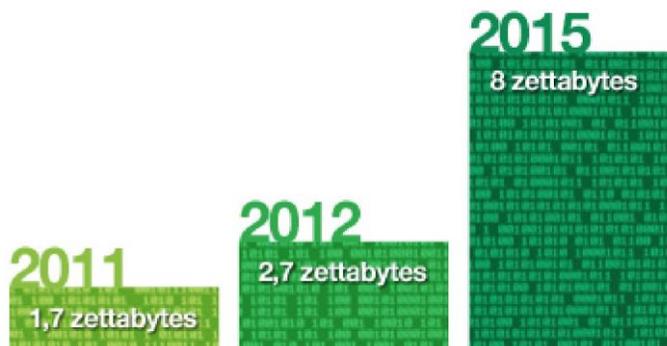
Fonte: (MENESES; ALMEIDA, 2012)

## 2.2 MINERAÇÃO DE IMAGENS DE SENSORIAMENTO REMOTO

Nos últimos anos, a quantidade de dados produzida por sistemas computacionais tem crescido acentuadamente. Este fato pode ser evidenciado claramente na quantidade de informações disponíveis e compartilhadas na Internet. Neste cenário, grandes bases de dados se tornaram comuns. Podemos citar as bases de dados de instituições financeiras, cartões de crédito, hospitais e governo (SILVA, 2011).

A Figura 7 apresenta a quantidade de informações armazenadas nos anos de 2011 e 2012, assim como uma estimativa da quantidade armazenada em 2015.

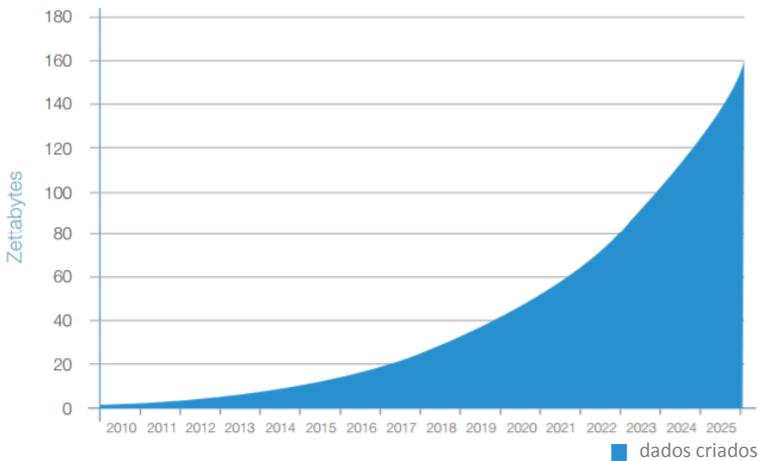
**Figura 7 – Volume de informação digital.**



Fonte: Adaptado de IBM (2012)

Segundo Reinsel, Gantz e Rydning (2017), em 2016 foram gerados cerca de 16 zettabytes de dados e em 2015 estima-se que serão gerados 163 zettabytes, cerca de 10 vezes mais que em 2016. Em 2025, aproximadamente 20% de todos os dados criados serão oriundos de uma variedade de dispositivos, como câmeras de segurança, automóveis, barcos, trens, máquinas de venda, leitores de digitais, brinquedos, entre outros. É possível observar esse crescimento na geração de dados na Figura 8.

**Figura 8 – Tamanho anual da esfera global de dados.**



Fonte: Adaptado de Reinsel, Gantz e Rydning (2017)

Outro fato relevante é a grande quantidade de informações que os satélites de sensoriamento remoto adquirem com o passar do tempo. Os bancos de dados dessas imagens estão entre os que mais crescem. O volume de dados armazenados dificulta a extração, pelo homem, de determinadas informações estratégicas sem o suporte de ferramentas específicas.

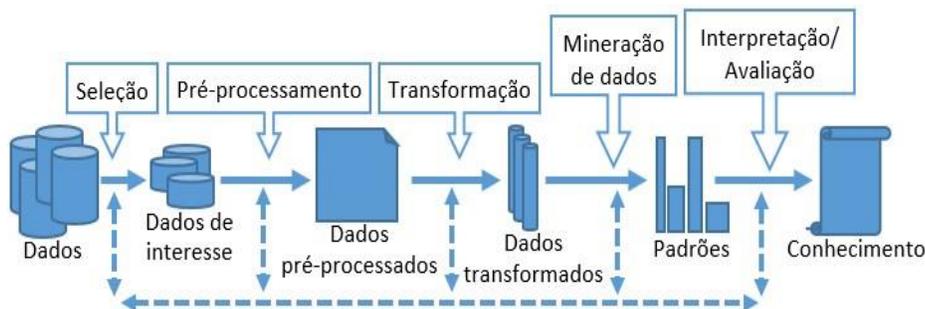
Assim, há uma demanda para uma nova geração de teorias e ferramentas computacionais para auxiliar os seres humanos a extrair padrões de grandes volumes de dados digitais. Estas teorias e ferramentas são o objeto de estudo da Descoberta de Conhecimento em Bancos de Dados - DCBD (do inglês *Knowledge Discovery in Databases* - KDD) (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

### 2.2.1 Mineração de dados

Para Fayyad (1997), DCBD é o processo não trivial de identificar padrões válidos, potencialmente úteis e compreensíveis nos dados. Ou seja, todo o processo geral, que envolve várias etapas interativas e

iterativas. Mineração de Dados (MD) é apenas uma etapa dentro de todo esse processo. As etapas no processo de DCBD são: seleção, limpeza e integração dos dados (pré-processamento), transformação dos dados, mineração dos dados e avaliação (interpretação) e apresentação dos resultados. A Figura 9 descreve o processo de DCBD.

**Figura 9 – Processo de descoberta de conhecimento em banco de dados.**



Fonte: (FAYYAD, 1997)

Nos termos de DCBD, dados são os elementos puros, quantificáveis sobre um determinado evento. Ou seja, são fatos, números ou textos, que individualmente não oferecem embasamento para o entendimento da situação. Informações são os dados analisados e contextualizados. Envolve interpretação de um conjunto de dados, portanto são constituídas por padrões, associações ou relações que todos aqueles dados acumulados podem proporcionar. E conhecimento é a combinação de estruturas de dados e procedimentos interpretáveis que levam a um comportamento conhecido. Esse conhecimento fornece base para previsão com determinado grau de certeza e auxilia na tomada de decisões.

Segundo Ramos e Lobo (2004), no processo de DCBD deve-se atentar para algumas preocupações e desafios, como por exemplo:

- Sociais: em relação à privacidade de dados, que podem pertencer a pessoas individuais ou a empresas;
- Instrumento de interação homem-máquina: a descoberta de conhecimento só é importante se for perceptível para o utilizador;
- Fontes de dados: diferentes tipos de dados requerem diferentes tipos de algoritmos e de métodos para os processar;
- Dados perdidos e com ruído: incluir estratégias estatísticas mais sofisticadas para identificar variáveis escondidas e dependências.

A mineração de dados consiste em extrair ou "minerar" conhecimento a partir de grandes repositórios de dados (FACELI et al., 2011). Em parte da literatura, a mineração de dados também é tratada pelos estatísticos e analistas de dados como sinónimo para outro termo, o processo de DCBD. Outros autores consideram a mineração de dados uma etapa no processo de DCBD, responsável pela aplicação de algoritmos específicos para extrair padrões de dados (FAYYAD, 1997).

Muitos métodos de mineração de dados são baseados em conceitos de aprendizagem de máquinas, reconhecimento de padrões e de estatística (RAMOS; LOBO, 2004).

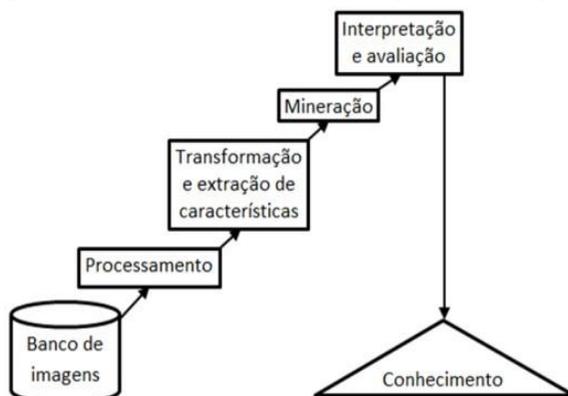
Na prática, os dois níveis principais dos objetivos primários de mineração de dados tendem a ser de previsão ou de descrição. De fato, os tipos de padrões a se descobrir dependem da tarefa de mineração de dados utilizada. A importância relativa da previsão e da descrição para uma determinada aplicação de mineração de dados, pode variar consideravelmente. Seus objetivos são concretizados recorrendo às tarefas primárias de mineração de dados: classificação; regressão; agrupamento ou *clustering*; sumarização; modelização das dependências; e alterações e detecção de desvios (RAMOS; LOBO, 2004).

Através do processo de DCBD é possível extrair dos bancos de dados informações e conhecimentos que podem ser usados em diversos tipos de aplicações, que vão desde a análise das informações de compras dos consumidores até o controle de produção e a exploração científica.

### 2.2.2 Mineração de imagens

O processo de mineração de dados em imagens é apresentado na Figura 10. As imagens de um acervo (banco de imagens) são recuperadas segundo critérios inerentes à aplicação. A seguir, uma fase de pré-processamento realça 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. Os 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 em outras aplicações estratégicas (SILVA, 2006).

**Figura 10 – Processo de mineração de imagens.**



Fonte: Adaptado de Zhang, Lee e Hsu (2002)

Segundo Silva (2006) a mineração de dados em imagens não consiste simplesmente no uso de técnicas de mineração de dados aplicadas em bancos “convencionais” ao domínio de imagens. Diferenças importantes entre estes bancos convencionais e os de imagens incluem:

- Textura: numa imagem, cada elemento (pixel) está muito relacionado com seus vizinhos, muitas vezes fazendo parte de uma região homogênea. Se tratarmos os elementos de imagens como entidades isoladas, perdemos a capacidade de capturar a informação de textura presente no contexto;
- Processamento em vários níveis: estudos clássicos sobre a visão humana (MARR, 1982) mostram que o processamento visual ocorre em vários níveis. A princípio ocorre a detecção de bordas, geometrias e estruturas dos objetos, até chegar à identificação de objetos na cena, contextualizando-os de acordo com as intenções e conhecimento do observador, associando os elementos perceptuais (borda, geometria, estrutura) a padrões, protótipos e eventos segundo a cognição do indivíduo;
- Ambiguidade de interpretação: diferentes observadores podem interpretar a mesma imagem de forma distinta, dependendo da natureza do estudo e dos métodos de análise empregados. O nível de conhecimento e experiência do intérprete influenciam diretamente na compreensão e avaliação dos elementos retratados;
- Dependência de domínio: cenários e atividades do mundo real pertencentes a determinado domínio possuem características e

elementos específicos. A identificação dos elementos, suas classes e relacionamentos está ligada ao contexto em si, podendo uma mesma imagem possuir informações distintas e inerentes a diferentes domínios em questão.

Dessa forma, a mineração de imagens de sensoriamento remoto trata especificamente do desafio de captar padrões, processos e agentes presentes no espaço geográfico, a fim de extrair conhecimento específico para entender ou para tomar decisões relativas a um conjunto de temas relevantes, incluindo a mudança da terra, variações climáticas e estudos de biodiversidade. Eventos como padrões de desmatamento, correlações de mudança climática e dinâmica de espécies são exemplos de preciosos conhecimentos contidos em repositórios de imagens de sensoriamento remoto (SILVA; CÂMARA; ESCADA, 2008).

Tais abordagens de mineração de imagens integram bancos de dados espaciais e questões de mineração de dados, trazendo recursos valiosos para entender fatos e processos representados em dados espaciais, descobrindo relacionamentos espaciais, criando bases de conhecimento espacial, e revelando padrões espaciais e processos contidos em repositórios espaciais (SILVA; CÂMARA; ESCADA, 2008).

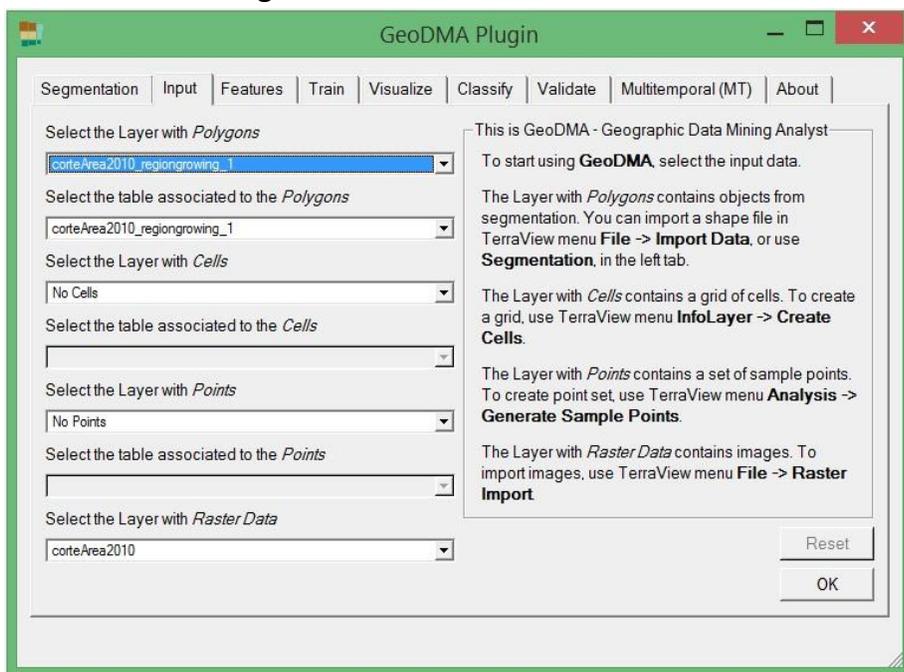
Aplicações desta tecnologia incluem, além de sensoriamento remoto, sistemas de informações geográficas, imagens médicas, geomarketing, navegação, controle de tráfego, estudos ambientais, e muitas outras áreas onde os dados espaciais são usados (SILVA; CÂMARA; ESCADA, 2008).

### 2.2.3 GeoDMA - Geographical Data Mining Analyst

A Figura 11 apresenta a ferramenta utilizada neste trabalho para minerar imagens, o Geographical Data Mining Analyst (GeoDMA).

Essa ferramenta surgiu a partir das ideias propostas por Silva et al. (2005), mas que ao longo do seu desenvolvimento se tornou mais robusta, podendo ser usada em diversas aplicações, incluindo agricultura e áreas urbanas. O GeoDMA 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 (KORTING et al., 2009).

**Figura 11 – Interface do GeoDMA.**



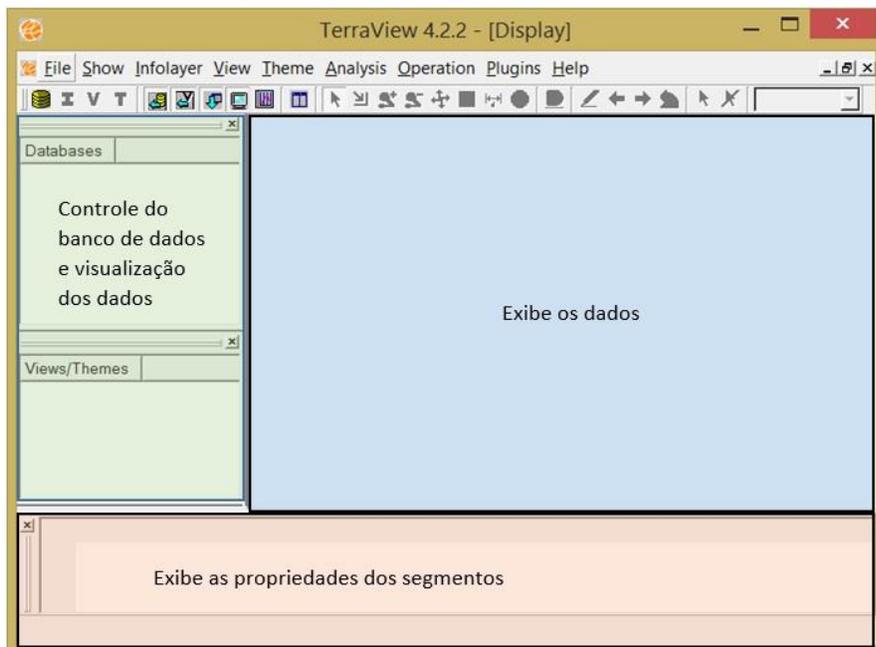
Fonte: Autoria própria

O GeoDMA é um plugin do sistema TerraView. Isto significa que todas as funcionalidades de manipulação e visualização de bancos de dados geográficos são provenientes do TerraView. Ambos os softwares são executados em conjunto e produzem os resultados que são exibidos na tela principal (KORTING et al., 2009).

O TerraView é um Sistema de Informação Geográfica (SIG) livre de código aberto desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (INPE), que proporciona fácil manuseio e entendimento por parte do usuário, portanto um aplicativo que apresenta os princípios básicos das técnicas de geoprocessamento, sensoriamento remoto e banco de dados geográficos. O SIG tem por alicerce uma biblioteca de geoprocessamento chamada TerraLib (CÂMARA et al., 2008).

O objetivo de sua construção foi apresentar à comunidade um fácil visualizador de dados geográficos com recursos de consulta e análise desses dados. O aplicativo permite a manipulação de dados geográficos representados na forma de vetor ou matriz. O armazenamento dos dados é feito em SGBDs como Access, MySQL e PostgreSQL (INPE, 2014). A Figura 12 permite a visualização da interface do TerraView, onde é possível visualizar, controlar e manipular os dados de imagens.

**Figura 12 – Interface do TerraView.**



A Figura 13 apresenta o diagrama geral do GeoDMA. As etapas de processamento começam a partir da definição dos dados de entrada, passando pela extração de características e a aplicação de algoritmos de mineração de dados para extrair e fornecer informações sobre observação da Terra (KORTING, 2012).

**Figura 13 – GeoDMA: diagrama das principais etapas de processamento para a análise de imagens.**



Fonte: Adaptado de Korting (2012).

Na etapa de definição de dados de entrada, o GeoDMA importa uma variedade de dados geoespaciais, que são armazenados em um banco de dados no formato *raster* ou vetorial. As regiões extraídas pela operação de segmentação, pontos e células definem os objetos que são armazenados no banco de dados.

Atualmente, o GeoDMA possui 4 algoritmos para realizar a segmentação: crescimento de regiões baseado em Bins et al. (1996); crescimento de regiões e um procedimento multiresolução baseado em Baatz e Schape (2000); algoritmo baseado em Korting, Fonseca e Câmara (2011) que utiliza mapas auto-organizáveis; e uma técnica de resegmentação aplicado às imagens urbanas com base em Korting, Dutra e Fonseca (2011).

Na segunda etapa, as características extraídas são divididas em três grupos:

- As características baseadas em segmentação são propriedades obtidas a partir dos objetos segmentados, integrando dados do tipo raster e vetores. São divididas em características espaciais e espectrais;
- As características baseadas em paisagem são as métricas de paisagem, usando tipo de dados vetoriais. A ecologia da paisagem considera *patches* como áreas, ou categorias, contendo habitat;
- Ciclos obtidos a partir de séries temporais *raster* são usados para extrair características multi-temporais. Esse conjunto de características foi dividido em dois grupos: básicas e polares.

A Tabela 1 apresenta algumas dessas características extraídas pelo GeoDMA. Mais informações sobre essas e outras características extraídas por esse programa são encontradas em Korting (2012) no Apêndice A.

Na etapa de extração de dados os objetos armazenados numa base de dados são utilizados pelo intérprete para selecionar amostras representativas para cada padrão de interesse. Os diferentes padrões compõem a tipologia da cobertura da terra, e alguns algoritmos automaticamente constroem um modelo de classificação com base em amostras de treinamento, que podem ser colhidas sempre que necessário. Este modelo deve ser usado para classificar o banco de dados inteiro, ou bancos de dados diferentes com a mesma tipologia.

O GeoDMA possui três algoritmos para a classificação: Mapas Auto-Organizáveis (do inglês *Self-Organizing Map* - SOM), Árvores de Decisão e Redes Neurais.

A saída do processo de mineração de imagens no GeoDMA é um mapa temático obtido pela aplicação do modelo de classificação para o banco de dados. Como o processo de DCBD sugere, a partir da saída o

intérprete pode fazer um número de inferências além da avaliação do mapa (KORTING, 2012).

Tabela 1 – Algumas das características extraídas pelo GeoDMA.

<b>Características espectrais baseadas em segmentação.</b>	
<b>Nome</b>	<b>Descrição</b>
<i>Amplitude</i>	Define a amplitude de um pixel dentro do segmento. A amplitude significa a diferença entre o valor máximo do pixel menos o valor mínimo.
<i>Mean</i>	Retorna o valor médio para todos n pixels dentro do segmento.
<i>Mode</i>	Retorna o valor mais ocorrente (moda) para todos n pixels dentro do segmento. Quando o segmento é multimodal, o primeiro valor é assumido.
<b>Características espaciais baseadas em segmentação</b>	
<b>Nome</b>	<b>Descrição</b>
<i>Angle</i>	Representa o ângulo principal de um segmento.
<i>Area</i>	Retorna a área do segmento. Quando medida em pixels é igual a N.
<i>Box area</i>	Retorna a área da caixa delimitadora de um segmento, medido em pixels.
<i>Length</i>	É a altura da caixa delimitadora do segmento.
<i>Width</i>	É a largura da caixa delimitadora do segmento.
<i>Perimeter</i>	É a quantidade de pixels na borda do segmento.
<b>Características baseadas em paisagem</b>	
<b>Nome</b>	<b>Descrição</b>
<i>Class area</i>	Significa a soma das áreas de uma célula.
<i>Percent land</i>	Igual à soma das áreas (m <sup>2</sup> ) de todos os patches do tipo patch correspondente, divididos por área de paisagem total (m <sup>2</sup> ). É igual ao percentual a paisagem composta do tipo patch correspondente.
<i>Landscape shape index</i>	É igual à soma do limite da paisagem e todos os segmentos da borda (m) dentro do limite. Essa soma envolve o tipo de patch correspondente (incluindo as fronteiras), dividido pela raiz quadrada da área de paisagem total (m <sup>2</sup> ).
<i>Mean perimeter area ratio</i>	É igual à soma da relação entre perímetros e áreas, dividido pelo número de patches do mesmo tipo.
<b>Características temporais para descrever eventos cíclicos.</b>	
<b>Nome</b>	<b>Descrição</b>
<i>Amplitude</i>	Tipo básico. A diferença entre os valores máximos e mínimos do ciclo.

	Uma pequena amplitude significa um ciclo estável.
<i>Circle</i>	Tipo polar. Retorna valores próximos a 1 quando a forma é mais semelhante a um círculo. Na visualização polar, um círculo significa uma característica constante.
<i>Cycle's mean</i>	Tipo básico. Valor médio da curva ao longo de um ciclo.
<i>Cycle's std</i>	Tipo básico. O desvio padrão dos valores do ciclo.

Fonte: Adaptado de Korting (2012).

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 a uma determinada classe. Com as amostras selecionadas nesta etapa e um modelo gerado (e.g. árvore de decisão) o próximo passo é a classificação, onde o modelo gerado é capaz de classificar dados de segmentos novos através do modelo construído na etapa de treinamento. O GeoDMA pode carregar um modelo já existente ou salvar um modelo construído para que ele possa ser usado em futuras classificações.

### 2.3 AGENTES DE MINERAÇÃO DE DADOS

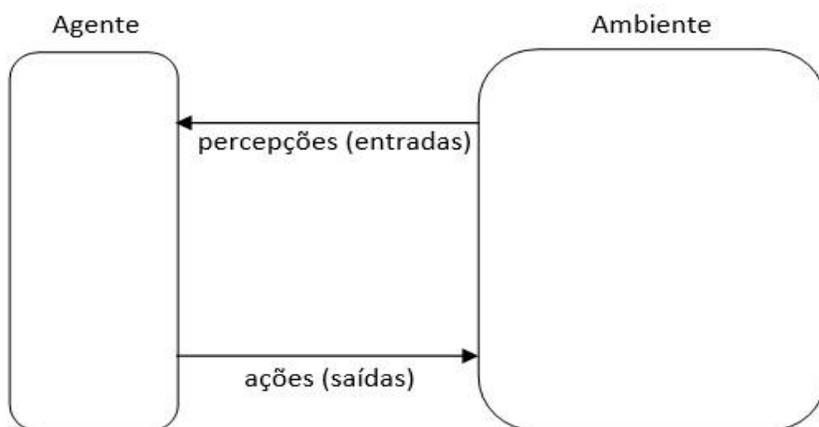
Um agente é algo que age (a palavra agente vem do latim *agere*, que significa fazer). No entanto, espera-se que um agente computacional tenha outros atributos que possam distingui-lo de meros “programas”, tais como operar sob controle autônomo, perceber seu ambiente, persistir por um período de tempo prolongado, adaptar-se a mudanças e ser capaz de assumir metas de outros (RUSSELL; NORVING, 2010).

Então pode-se dizer que um agente é tudo o que pode ser considerado capaz de perceber seu ambiente por meio de sensores e

de agir sobre esse ambiente por intermédio de atuadores (RUSSELL; NORVING, 2010). A Figura 14 ilustra essa ideia. Em outras palavras, um agente tem a capacidade de analisar um ambiente e atuar nesse ambiente de forma autônoma, modificando ou não o ambiente, para alcançar suas metas. As características de um agente são: autonomia, interação, reação e iniciativa.

O termo percepção faz referência às entradas perceptivas do agente em qualquer momento. Em geral, a escolha da ação de um agente em qualquer instante pode depender da sequência inteira de percepções observadas até o momento. Um agente é completamente especificado pela “função do agente”, que mapeia a sequência de percepções em ações (RUSSELL; NORVING, 2010).

**Figura 14 – Agentes interagem com ambientes por meio de sensores e atuadores.**



Fonte: Adaptado de Russell e Norving (2010)

Russell e Norving (2010) descrevem quatro tipos básicos de agentes:

- Agentes reativos simples: esses agentes selecionam ações com base na percepção atual, ignorando o restante do histórico de percepções. Os agentes reativos simples têm a admirável

propriedade de serem simples, mas se caracterizam por terem inteligência muito limitada;

- Agente reativo com estado: é um agente reativo mais complexo, pois mantém um estado interno que depende de seu histórico de percepções e ações. A cada nova percepção, o agente atualiza seu estado interno e suas ações agora dependem não só da percepção recebida, mas também do seu estado interno. Esse agente também é chamado de agente baseado em modelo;
- Agentes baseados em objetivos: esses agentes precisam saber quais são os seus objetivos para que assim possam alcançá-los. Então, esse agente mantém um registro interno do estado do mundo baseado em suas percepções e também um conjunto de metas que ele tenta alcançar. Esses objetivos são alcançados através de busca e planejamento;
- Agente baseado em utilidade: além dos objetivos, este agente é capaz de mensurar o grau de felicidade a partir do estado em que se encontra. Isto é feito através da função de utilidade que o agente mantém internamente. Essa função de utilidade mapeia um estado (ou uma sequência de estados) em um número real, que descreve o grau de felicidade associado.

Segundo Chauhan (1997 apud XAVIER JÚNIOR, 2012), as funções desempenhadas pelos agentes podem ser classificadas nos seguintes tipos:

- Agentes de Entretenimento: são agentes que tem como propósito o entretenimento. Dentro desse grupo, estão aqueles que são desenvolvidos especialmente para os mais diversos tipos de jogos;
- Agentes de Informação: são agentes que têm acesso e são capazes de coletar e manipular informações de muitas fontes. As

informações obtidas normalmente são oriundas de consultas propostas por usuários e/ou agentes;

- **Agentes de Interface:** são agentes que apoiam e fornecem assistência ao usuário. Eles observam e monitoram as ações do usuário, aprendem com ele e sugerem melhores formas para realizar uma tarefa. Seu aprendizado pode ser através de observação e imitação do usuário, ou por feedback positivo ou negativo, por instruções explícitas do usuário ou através de outros agentes;
- **Agentes Colaborativos:** são agentes que trabalham com autonomia e cooperação entre si e com o objetivo de realizar suas tarefas. Seus atributos-chaves incluem autonomia, habilidade social, correspondência e pró-atividade. Visando obter uma organização coordenada, eles podem negociar para obter, mutuamente, acordos aceitáveis em alguns assuntos;
- **Agentes Móveis:** são agentes que possuem a habilidade de se mover de um host a outro de uma rede durante suas execuções, carregando consigo seus estados de execução. Questões como segurança, discrição, mecanismo de transporte e autenticação são bastante relevantes neste tipo de agente;
- **Agentes Reativos:** são agentes que não possuem modelo interno do seu ambiente. Em vez disso, eles respondem de acordo com o estado do ambiente em que estão inseridos;
- **Agentes Híbridos:** são agentes que combinam dois ou mais tipos de agentes com o objetivo de maximizar as capacidades e minimizar as deficiências de cada um deles.

Os agentes de software, inteligentes ou não, são programas que executam tarefas específicas em favor de um usuário, de forma independente ou com alguma orientação. Um agente inteligente efetua, de forma reativa e/ou proativa, tarefas interativas adaptadas à necessidade de usuário humano ou de outros agentes. Para realizar estas tarefas, é necessário possuir um conjunto de características: independência, aprendizado, cooperação, raciocínio e inteligência (BUI; LEE, 1999 apud KANEKO, 2005).

O projeto de um agente inteligente deve especificar sua percepção, suas ações, objetivos e qual o seu ambiente, ou seja, o *Performance Environment Actuators Sensors* (PEAS). Para Wooldridge (2009) as arquiteturas de agentes são, dessa maneira, arquiteturas de softwares para sistemas de tomada de decisão que estão embarcados em um ambiente.

Para Russell e Norving (2010) o ambiente onde o agente atua é onde o problema se encontra, para o qual o agente é a solução. Um dos ambientes mais importantes para agentes inteligentes é a internet. As tecnologias de inteligência artificial servem de base para muitas ferramentas na Internet, como mecanismos de pesquisa, sistemas de recomendação e sistemas de construção de Web sites. A seguir, algumas propriedades de ambiente:

- Completamente Observável vs. Parcialmente Observável: Um ambiente é completamente observável se os sensores dos agentes detectam todos os aspectos que são relevantes para a escolha da ação. E pode ser parcialmente observável por causa dos ruídos ou sensores não apurados, ou simplesmente porque nem todos os estados do ambiente podem ser percebidos pelos sensores do agente;

- Determinístico vs. Estocástico: Se o próximo estado do ambiente é completamente determinado pelo atual estado e a ação executada pelo agente, então dizemos que o ambiente é determinístico, caso contrário, estocástico;
- Episódico vs. Sequencial: Num ambiente episódico, a experiência do agente é dividida em episódios atômicos. Cada episódio consiste do agente percebendo e então executando uma única ação. O próximo episódio não depende das ações tomadas nos episódios anteriores. No ambiente episódico a escolha da ação em cada episódio depende unicamente do próprio episódio. Dessa forma o agente não precisa pensar à frente. Já no ambiente sequencial, a decisão atual pode afetar as decisões futuras, como num jogo de xadrez, por exemplo;
- Estático vs. Dinâmico: Se o ambiente pode mudar enquanto um agente está deliberando, então dizemos que o ambiente é dinâmico para o agente; caso contrário, ele é estático e o agente não precisa continuar observando o ambiente enquanto está tomando uma decisão e não precisa se preocupar com o passar do tempo;
- Discreto vs. Contínuo: A distinção entre ambiente discreto e contínuo pode ser aplicada ao estado do ambiente, considerando o tempo, as percepções e ações do agente;
- Agente único vs. Multiagente: Um agente resolvendo palavras cruzadas sozinho está nitidamente num ambiente com agente único, enquanto que um agente jogando xadrez está em um ambiente com mais de um agente. Essa distinção de ambientes faz diferença na visão do agente, já que em um ambiente

multiagente ele precisa considerar a presença dos outros agentes para maximizar sua medida de desempenho. Em um ambiente multiagente competitivo, para que um agente maximize seu desempenho, ele precisa minimizar a medida de performance dos agentes concorrentes. Isso acontece também em um ambiente multiagente cooperativo, onde o agente precisará se preocupar com o desempenho dos demais agentes para que todos possam atingir o objetivo em conjunto. Os problemas nos ambientes de agente único e multiagente são muitas vezes completamente diferentes. A comunicação, por exemplo, frequentemente emerge como um comportamento racional em sistemas multiagente. Em ambientes competitivos, o comportamento estocástico é racional, porque evita que sejam criadas armadilhas explorando a previsibilidade. As características do agente e do ambiente devem ser consideradas no projeto de agentes em um Sistema Multiagente (SMA).

Sistemas multiagente são um subcampo relativamente novo na Inteligência Artificial. A área só começou a ser estudada nos anos 80 e só ganhou reconhecimento em meados dos anos 90. Mas, desde então, o interesse científico e industrial no campo cresceu. Este rápido crescimento tem sido impulsionado pela constatação de que agentes são um paradigma de software apropriado para explorar necessidades apresentadas pelas plataformas modernas de computação e de informação: distribuídas, abertas, de grande volume e heterogêneas (WOOLDRIDGE, 2009).

Para Wooldridge (2009), um sistema multiagente consiste de um número de agentes que interagem uns com os outros, tipicamente através de troca de mensagens ou através de alguma infraestrutura de rede de computador. No caso mais geral, os agentes em um sistema multiagente estarão representando ou agindo em nome dos usuários

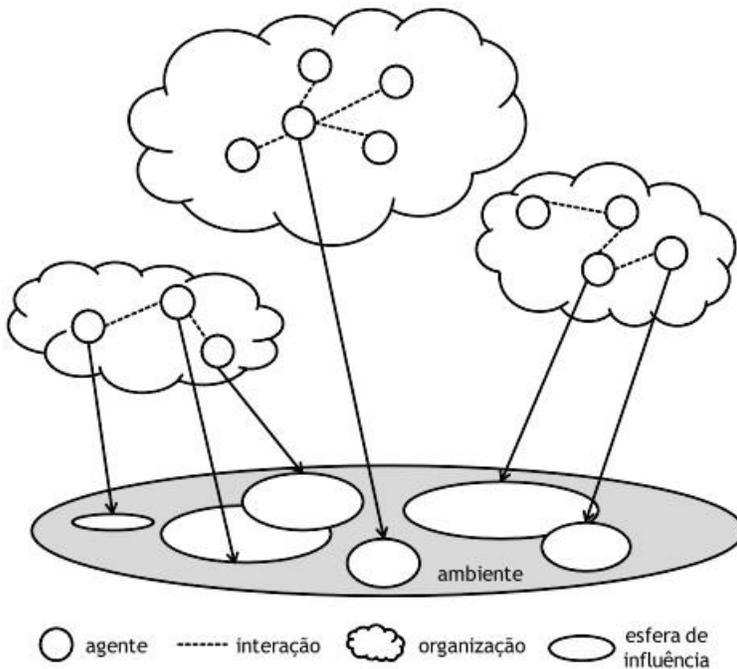
ou proprietários com os mais diversos objetivos e motivações. A fim de interagir com sucesso, esses agentes possuem assim a capacidade de cooperar, coordenar e negociar uns com os outros, da mesma maneira que nós cooperamos, coordenamos e negociamos com outras pessoas em nossas vidas.

Segundo Sycara (1998), as características de SMA são:

- Cada agente possui uma visão limitada;
- Não há controle global do sistema;
- Trabalham com dados descentralizados; e
- A computação é assíncrona, de maneira a permitir a comunicação entre agentes heterogêneos.

Para Wooldridge (2009), SMA contém um número de agentes que interagem por comunicação, são capazes de agir em um ambiente, possuem diferentes esferas de influência, e são ligados uns aos outros por relações organizacionais, como ilustrado na Figura 15. Cada agente possui diferentes controles sobre diferentes partes do ambiente, podendo ou não existir sobreposição de nicho entre eles. O fato dessas esferas poderem coincidir faz surgir dependências entre os agentes. Os agentes também estão ligados por outros relacionamentos hierárquicos.

**Figura 15 – Estrutura típica de um sistema multiagente.**



Fonte: Adaptado de Wooldridge (2009)

Segundo Bellifemine, Caire e Greenwood (2007), coordenação é um processo no qual agentes se comprometem em ajudar a garantir que uma comunidade de agentes individuais aja de maneira coerente. São várias as razões pelas quais o SMA necessita ser coordenado, tais como:

- Os objetivos dos agentes podem causar conflitos entre as suas ações;
- Os objetivos dos agentes podem ser interdependentes;
- Agentes podem ter diferentes capacidades e diferentes conhecimentos; e
- Metas globais do sistema podem ser alcançadas mais rapidamente se diferentes agentes agirem em cada subproblema.

No livro *Os Dragões do Éden*, Sagan (1977) fala sobre os cães de caça, que possuem a capacidade de seguir trilhas utilizando o olfato. Então, em um vasto emaranhado de odores, esses cães conseguem distinguir um determinado padrão, seja o odor de uma pessoa desaparecida ou de um animal. O cão de caça serve como exemplo real de um agente minerador de dados.

Os sistemas de agentes/multiagente e o processo de descoberta de conhecimento em banco de dados são duas áreas de pesquisa cada vez mais inter-relacionadas, que buscam encontrar benefícios para ambas as partes. Ambos os campos enfrentam desafios que a outra tecnologia pode ajudar.

Os problemas típicos em agentes que podem encontrar soluções satisfatórias em Mineração de Dados incluem aprendizagem multiagente, adaptação, evolução e análise de comportamento. Por exemplo, o conhecimento extraído através da Mineração de Dados poderia fornecer modelos mais estáveis, previsíveis e controláveis para o planejamento ou pode ajudar na auto-organização e evolução dos sistemas multiagente em direções aceitáveis (CAO; GORODETSKY; MITKAS, 2009).

Para Seydim (1999), os agentes podem ser usados para automatizar as tarefas individuais em vários passos através do processo de descoberta de conhecimento em banco de dados, tais como:

- Na preparação dos dados, o uso dos agentes pode ser especialmente na sensibilidade aos parâmetros de aprendizagem, aplicando, por exemplo, alguns gatilhos para atualizações de banco de dados e manipulação de dados em falta ou inválidos;
- Na preparação dos dados em mineração de dados, que envolve a seleção dos dados, limpeza dos dados, pré-processamento e representação, os agentes podem ser utilizados para automatizar

vários desses passos. Uma possibilidade para automatizar a etapa de seleção dos dados é a realização de análise de sensibilidade automática para determinar quais os parâmetros que devem ser utilizados na aprendizagem. Isto reduziria a dependência de ter especialistas do domínio para examinar o problema cada vez que algo mudar no ambiente;

- A etapa de limpeza dos dados pode ser automatizada através da utilização de um agente inteligente com uma base de regras, por exemplo. Quando um registro é adicionado ou atualizado em um banco de dados relacional, um gatilho poderia chamar o agente inteligente para examinar os dados da transação. As regras em sua base de regra poderiam especificar a forma de limpar os dados faltantes ou inválidos;
- Pré-processamento de dados também requer o conhecimento do domínio, uma vez que não há nenhuma maneira de saber a semântica dos atributos e relacionamentos, como campos calculados ou derivados. No entanto, os passos de pré-processamento e representação de dados, como escala ou redução de dimensionalidade, mapeamento de símbolos e normalização, que geralmente são especificados pelo perito em mineração de dados, podem ser automatizados usando regras e informações estatísticas básicas sobre variáveis;
- Na busca por padrões de interesse, utilizando aprendizagem e inteligência na classificação, *clustering*, sumarização e generalização. Um agente pode aprender a partir de um perfil ou a partir de exemplos. Comentários de usuários podem ser usados para refinar a confiança nas previsões do agente. Um agente inteligente pode usar o conhecimento de domínio com regras

simples incorporado e utilizando os dados de treinamento que pode aprender e reduzir a necessidade de especialistas de domínio. No entendimento do que se aprende, o uso do agente pode ser apenas como um agente fixo ou simplesmente um programa na visualização.

A principal vantagem do uso de agentes inteligentes na automação da mineração de dados é indicada como seu possível apoio para mineração de dados em transações online. Quando novos dados são adicionados ao banco de dados, um alarme ou um agente gatilho pode enviar eventos para a principal aplicação de mineração e para a tarefa de aprendizagem, para que os novos dados possam ser avaliados com os dados já minerados (SEYDIM, 1999).

Esta integração pode auxiliar em vários aspectos o processo de descoberta de conhecimento, tais como problemas de administração dos dados, interação e supervisão humana no processo de MD, seleção de dados, enriquecimento do conhecimento através da combinação de técnicas distintas de MD, entre outros (CAO; GORODETSKY; MITKAS, 2009).

Segundo Cao, Gorodetsky e Mitkas (2009), essa ideia de agentes de mineração, que une as duas áreas para lidar melhor com um problema em vez de usar uma área de forma unilateral, é impulsionada por crescentes aplicações em diversas áreas, como por exemplo:

- E-commerce, incluindo leilões eletrônico, negociações e trading;
- Inteligência de negócios, incluindo a cadeia de suprimentos e gestão de relacionamento com o cliente;
- Mineração de Dados em aplicações que necessitam de apoio à decisão e de ação rápida, como o mercado financeiro, assistência médica e segurança;

- Alocação de recursos em ambientes paralelos e distribuídos, tais como computação em grid, serviços de rede e computação paralela;
- Mineração de Web 1.0, 2.0 e Web semântica.

## 2.4 TRABALHOS RELACIONADOS

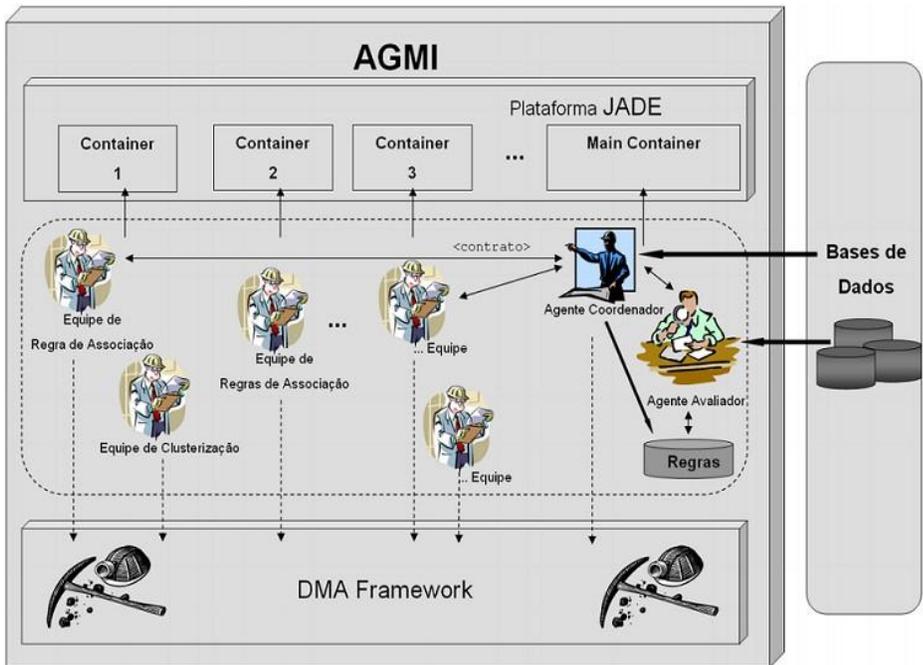
A interação entre MD e agentes é aplicada em diferentes campos de pesquisa. São apresentados em seguida alguns trabalhos relacionados ao tema proposto nesta pesquisa:

Silva (2011) apresenta um protótipo do *AGent 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.

O AGMI apresentou bons resultados quanto ao desempenho, capacidade autônoma de melhorar o conhecimento descoberto e quanto à qualidade do conhecimento apresentado. Comparando com a abordagem testada, utilizando apenas o algoritmo de Regras de Associação, os experimentos com AGMI mostraram um aumento médio de 170% na qualidade média das 10 melhores regras encontradas e de 350% na qualidade média das 100 melhores regras encontradas.

Além disso, AGMI aumentou a qualidade de 193 regras, através de heurística aplicada autonomamente pelo agente Avaliador. As regras descobertas nos experimentos foram analisadas por especialistas da Controladoria Geral da União e apresentaram fortes indícios de irregularidades em licitações tais como cartéis, simulação de concorrência e direcionamento de editais. A Figura 16 mostra o AGMI numa visão estrutural.

Figura 16 – Visão estrutural do AGMI.

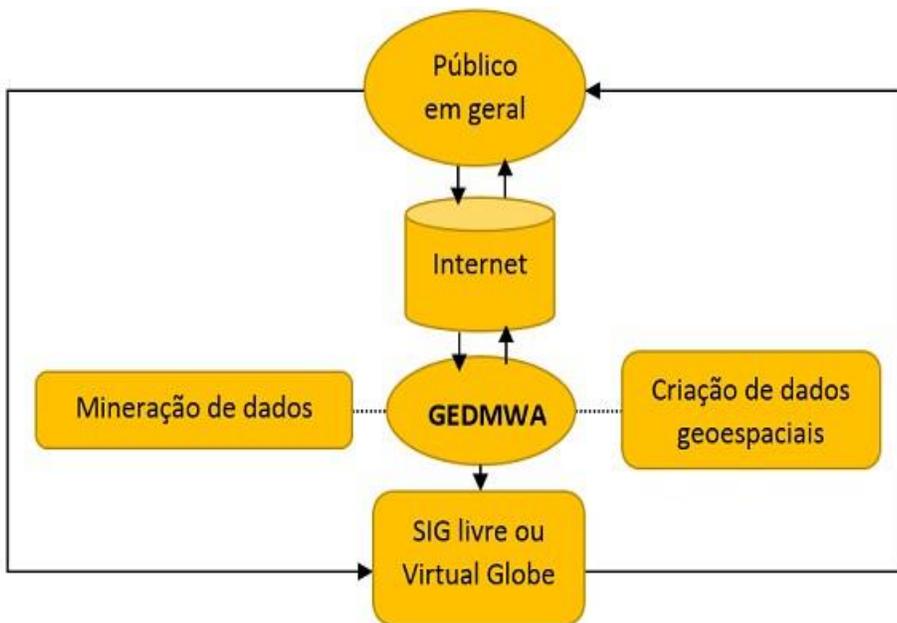


Fonte: Silva (2011).

Pultar, Raubal e Goodchild (2008) trazem o projeto e a implementação de um protótipo do *Geospatial Exploratory Data Mining Web Agent* (GEDMWA). Esse protótipo intercepta os dados de páginas na web e segue links para adquirir conhecimento, a fim de agregar valor à geoinformação utilizável em um Sistema de Informação Geográfica (SIG).

O agente cria um banco de dados com textos de páginas da web, minera e explora informações de localização e então converte para um formato adequado de dados geoespaciais. Isso proporciona para diversas comunidades de usuários uma ferramenta que utiliza uma variedade de fontes distribuídas para descobrir conhecimento adicional sobre as suas áreas de interesse. Pode-se observar na Figura 17: o fluxo global, os métodos e as ferramentas utilizadas.

**Figura 17 – GEDMWA - Entrada e saída de dados.**



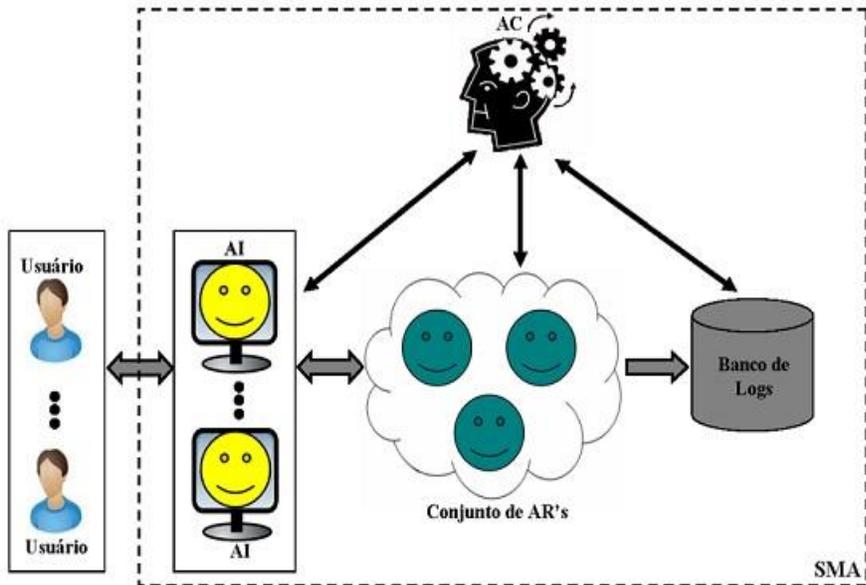
Fonte: Pultar, Raubal e Goodchild (2008).

Xavier Júnior (2012) propôs o NatalGIS, que é um sistema multiagente para recomendação de informações geográficas. O sistema permite que os usuários possam guardar as informações visualizadas em seus próprios históricos de acesso, estruturados para tal, em um banco de dados relacional.

Baseado nesse histórico, o sistema forma agrupamentos de informações relevantes através do uso de métodos de agrupamento de dados relacionais. Através das informações guardadas no perfil dos usuários, os agentes inteligentes podem interagir e recomendar informações geográficas relevantes aos usuários.

A Figura 18 apresenta a arquitetura do sistema, onde é possível observar os Agentes de Interface (AI), os Agentes de Controle (AC) e o conjunto de Agentes de Recomendação (AR).

Figura 18 – Arquitetura do sistema NatalGIS



Fonte: Xavier Júnior (2012).

Todos esses trabalhos, bem como outros trabalhos existentes na literatura e que não foram citados aqui, exemplificam o uso de agentes e sistemas multiagente na mineração de dados, mas nenhuma das metodologias já foi aplicada à mineração de imagens de satélite.

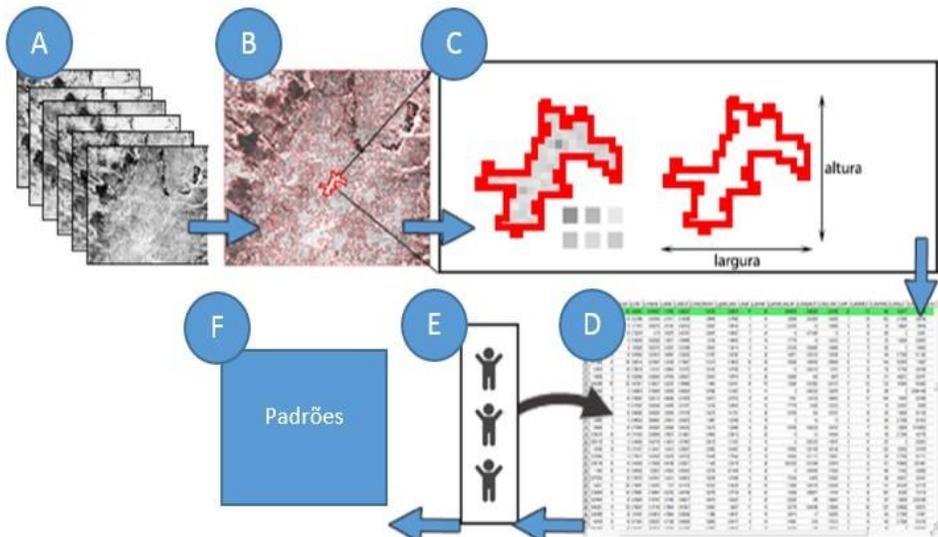
## CAPÍTULO 3 METODOLOGIA

Conforme apresentado na Seção 2.2, uma das etapas do processo de mineração de imagens é a extração de características, que detecta informações importantes sobre as imagens. A seguir, a etapa de mineração propriamente dita tem a finalidade de descobrir padrões relevantes. Esses padrões serão avaliados e interpretados para que se possa obter o conhecimento que será aplicado no entendimento de problemas.

Na Seção 2.3, foram apresentados os conceitos de agentes e sistemas multiagente, que podem ser utilizados em diversas etapas do processo de DCBD, inclusive na etapa de mineração de dados, como visto na Seção 2.4. A proposta deste livro é elaborar uma metodologia e aplicá-la com os conceitos de Agentes e Sistemas Multiagente para avançar na Mineração de Imagens de Sensoriamento Remoto.

O processo, como se pode observar na Figura 19, foi dividido nas seguintes fases: (A) a primeira etapa é composta pela aquisição das imagens, ou seja, a seleção de dados no repositório de imagens de satélites. (B) A seguir, essas imagens são segmentadas para extrair regiões e objetos da paisagem. (C) Em seguida, os atributos espaciais e espectrais de cada um dos segmentos são extraídos. (D) Esses atributos são armazenados no banco de dados e servirão como base de conhecimento e parâmetro para os agentes. (E) Depois, com os agentes previamente treinados e já com um modelo interno, pode-se minerar os dados desses ou de novos segmentos em busca de determinados padrões. (F) Esses padrões resultantes são então avaliados e interpretados por um especialista para a obtenção do conhecimento final.

Figura 19 – Diagrama do processo.



Fonte: Autoria própria

As seções a seguir detalham as etapas desse processo e também a ferramenta desenvolvida chamada AgentGeo, que é responsável pela etapa de mineração envolvendo os agentes.

### 3.1 SELEÇÃO DE DADOS DO REPOSITÓRIO DE IMAGENS DE SATÉLITES

Essa é a primeira etapa do processo, que consiste na aquisição de imagens de satélite da área de estudo selecionada. Esses dados são inseridos no TerraView, que é responsável pela parte de manipulação de dados geográficos. Neste trabalho foram realizados dois estudos de caso (Capítulo 4), um com cenas do satélite LANDSAT-5 e outro com cenas do satélite LANDSAT-8.

As cenas utilizadas no primeiro estudo de caso foram adquiridas no Banco de Imagens da Divisão de Geração de Imagens (DGI) do INPE (DGI, 2011).

O satélite responsável pela captura das cenas foi o LANDSAT-5 TM, lançado em 01 de março de 1984 e desativado em 5 de junho de 2013. As cenas obtidas cobrem, cada uma, uma área de 170 x 185 Km (USGS, 2014d).

Esse satélite possui sete bandas espectrais, onde cada uma representa uma resposta da reflectância do solo em uma determinada faixa espectral. Em DGI (2009b), é possível observar as principais características e aplicações de cada banda, como:

- Banda 1 (0,45 a 0,52  $\mu\text{m}$  - azul) - Apresenta grande penetração em corpos de água, com elevada transparência, permitindo estudos batimétricos. Sofre absorção pela clorofila e pigmentos fotossintéticos auxiliares (carotenóides). Apresenta sensibilidade a plumas de fumaça oriundas de queimadas ou atividade industrial. Pode apresentar atenuação pela atmosfera;
- Banda 2 (0,52 a 0,60  $\mu\text{m}$  - verde) - Apresenta grande sensibilidade à presença de sedimentos em suspensão, possibilitando sua análise em termos de quantidade e qualidade. Boa penetração em corpos de água;
- Banda 3 (0,63 a 0,69  $\mu\text{m}$  - vermelho) - A vegetação verde, densa e uniforme, apresenta grande absorção, ficando escura, permitindo bom contraste entre as áreas ocupadas com vegetação (ex.: solo exposto, estradas e áreas urbanas). Apresenta bom contraste entre diferentes tipos de cobertura vegetal (ex.: campo, cerrado e floresta). Permite análise da variação litológica em regiões com baixa cobertura vegetal. Permite o mapeamento da drenagem através da visualização da mata galeria e entalhe dos cursos dos rios em regiões com pouca cobertura vegetal. Esta é a banda mais

utilizada para delimitar a mancha urbana, incluindo identificação de novos loteamentos. Permite a identificação de áreas agrícolas;

- Banda 4 (0,76 a 0,90  $\mu\text{m}$  - infravermelho próximo) – Os corpos de água absorvem muita energia nesta banda e ficam escuros, permitindo o mapeamento da rede de drenagem e delineamento de corpos de água. A vegetação verde, densa e uniforme, reflete muita energia nesta banda, se mostrando clara nas imagens. Apresenta sensibilidade à rugosidade da copa das florestas (dossel florestal). Apresenta sensibilidade à morfologia do terreno, permitindo a obtenção de informações sobre Geomorfologia, Solos e Geologia. Serve para análise e mapeamento de feições geológicas e estruturais, separar e mapear áreas ocupadas com pinus e eucalipto, além de ser usada para mapear áreas ocupadas com vegetação que foram queimadas. E permite a visualização de áreas ocupadas com macrófitas aquáticas (ex.: aguapé). Permite a identificação de áreas agrícolas;
- Banda 5 (1,55 a 1,75  $\mu\text{m}$  - infravermelho médio) - Apresenta sensibilidade ao teor de umidade das plantas, servindo para observar estresse na vegetação, causado por desequilíbrio hídrico. Esta banda pode sofrer perturbações em caso de ocorrer excesso de chuva antes da obtenção da cena pelo satélite;
- Banda 6 (10,4 a 12,5  $\mu\text{m}$  - infravermelho termal) - Apresenta sensibilidade aos fenômenos relativos aos contrastes térmicos, servindo para detectar propriedades termais de rochas, solos, vegetação e água;
- Banda 7 (2,08 a 2,35  $\mu\text{m}$  - infravermelho distante) - Apresenta sensibilidade à morfologia do terreno, permitindo obter

informações sobre Geomorfologia, Solos e Geologia. Esta banda serve para identificar minerais com íons hidroxilas. Potencialmente favorável à discriminação de produtos de alteração hidrotermal.

Cada pixel da imagem tem uma resolução espacial de 30 metros (isto é, representa uma área no formato de um quadrado no solo de 30 metros de lado), com exceção da banda 6, que tem uma resolução espacial de 120 metros (USGS, 2014d).

Cada uma destas sete bandas compõe uma imagem em 256 tons de cinza, enquanto uma imagem colorida é gerada com a composição de 3 bandas, ou seja, associa-se para cada uma das bandas as cores vermelha, verde e azul, de acordo com o objetivo de tal combinação. Segundo a recomendação do DGI (2009a), as principais combinações para imagens coloridas são:

- Bandas 3, 2 e 1: imagens em "cor natural", com boa penetração na água, realçando as correntes, a turbidez e os sedimentos. A vegetação aparece em tonalidades esverdeadas;
- Bandas 2, 3 e 4: define melhor os limites entre o solo e a água, ainda mantendo algum detalhe em águas pouco profundas, e mostrando as diferenças na vegetação que aparece em tonalidades de vermelho;
- Bandas 3, 4 e 5: mostra mais claramente os limites entre o solo e a água, com a vegetação mais discriminada, aparecendo em tonalidades de verde e rosa;
- Bandas 2, 4 e 7: mostra a vegetação em tons verdes e permite discriminar a umidade tanto na vegetação como no solo.

Já as cenas utilizadas no segundo estudo de caso foram adquiridas no EarthExplorer (USGS, 2014a). O satélite responsável pela captura dessas cenas foi o LANDSAT-8, que está em órbita desde 11 de fevereiro de 2013. A resolução temporal do LANDSAT-8 é de 16 dias, ou seja, o satélite leva este período para realizar outra passagem em uma mesma região. O mesmo possui 11 bandas e diferentes resoluções espaciais, como se pode observar na Tabela 2 (USGS, 2014e).

Tabela 2 - Bandas e respectivas resoluções espaciais do LANDSAT-8.

<b>Bandas</b>	<b>Tamanho da onda (micrometros)</b>	<b>Resolução (Metros)</b>
Banda 1 - Ultra Azul	0.43 – 0.45	30
Banda 2 - Azul	0.45 – 0.51	30
Banda 3 - Verde	0.53 – 0.59	30
Banda 4 - Vermelho	0.64 – 0.67	30
Banda 5 - Infravermelho Próximo	0.85 – 0.88	30
Banda 6 - Infravermelho de Ondas Curtas 1	1.57 – 1.65	30
Banda 7 - Infravermelho de Ondas Curtas 2	2.11 – 2.29	30
Banda 8 - Pancromática	0.50 – 0.68	15
Banda 9 - Cirrus	1.36 – 1.38	30
Banda 10 - Infravermelho Termal 1	10.60 – 11.19	100
Banda 11 - Infravermelho Termal 2	11.50 – 12.51	100

Fonte: Adaptado de USGS (2014b).

O nível de detalhes (resolução espacial) não é o único aspecto interessante de se analisar numa imagem de satélite, mas também a forma como as mudanças de energia refletida por diferentes tipos de materiais da superfície é utilizada para identificar características de interesse. Segundo USGS (2014c), cada uma dessas bandas possui características peculiares, sendo mais indicada para determinados tipos de estudos. A seguir, é possível observar algumas dessas características ou indicação de estudo:

- Banda 1 - Útil para estudos costeiros e aerossol;
- Banda 2 - Indicada para mapeamento batimétrico, distinção do solo da vegetação e da vegetação decídua da conífera;
- Banda 3 - Útil para a avaliação do vigor das plantas;
- Banda 4 - Discrimina encostas de vegetação;
- Banda 5 - Enfatiza teor de biomassa e linhas costeiras;
- Banda 6 - Discrimina o teor de umidade do solo e da vegetação. Penetra nuvens finas;
- Banda 7 - Discrimina melhor o teor de umidade do solo e da vegetação. Penetra nuvens finas;
- Banda 8 - Possui resolução espacial de 15 metros e uma definição de imagem mais nítida;
- Banda 9 - Útil para a detecção de nuvens;
- Banda 10 - Indicada para mapeamento térmico e estima a umidade do solo;
- Banda 11 - Melhor mapeamento térmico e estimativa da umidade do solo.

Uma cena do LANDSAT-8 cobre, aproximadamente, 170 x 183 quilômetros. A resolução radiométrica é de 12 bits, fornecendo 4096 potenciais níveis de cinza em cada imagem, diferentemente dos 256 níveis de cinza dos 8 bits de versões anteriores de sua família, como o LANDSAT-5 (USGS, 2014e; USGS, 2014b).

### 3.2 SEGMENTAÇÃO E EXTRAÇÃO DE CARACTERÍSTICAS

Após a construção do banco de dados, as imagens adquiridas precisam passar por um processo de segmentação para extrair regiões elementares. Nesse processo, a imagem é particionada em regiões homogêneas, denominadas de segmento ou objeto e cada um desses segmentos possuem características espectrais e espaciais. Todos esses segmentos e as características que são extraídas de cada um deles são armazenados no banco de dados.

Para a etapa da segmentação, utilizamos o algoritmo de Crescimento de Regiões baseado em Bins et al. (1996) que está implementado no GeoDMA (Subseção 2.2.3), o qual agrupa pixels formando regiões. Para realizar este agrupamento o algoritmo utiliza dois limiares: Similaridade e Área Mínima.

Segundo Barbosa et al. (2000), a similaridade define a diferença mínima entre o valor de um pixel e o valor médio da região contígua a ele, para que este pixel possa ser agrupado a esta região. Se a diferença entre o valor do pixel e a média da região vizinha for maior que o valor de similaridade definida pelo usuário, o pixel não será agrupado àquela região. Por outro lado, a área mínima define o menor tamanho de região de acordo com o limiar definido pelo usuário, ou seja, não haverá regiões com área em pixels menor do que o valor definido pelo usuário.

O GeoDMA também é o responsável por extrair as características de cada um dos segmentos. Essas características baseadas em segmentação podem ser espectrais e espaciais. Segundo Korting (2012), as características espectrais estão relacionadas a todos os valores dos pixels dentro de uma região e incluem, portanto, métricas para valores mínimos e máximos de pixel, ou para valores médios. As características espaciais retratam aspectos morfológicos dos segmentos, incluindo a altura, largura e área.

### 3.3 AGENTGEO

O AgentGeo foi desenvolvido para realizar a etapa de mineração de imagens através de agentes, que implementa funções requeridas pela proposta deste trabalho: criação, edição e seleção de agentes, seleção e criação do ambiente e uso dos agentes para minerar as imagens.

Este sistema foi desenvolvido em Java utilizando o ambiente NetBeans IDE 8.0, que é um projeto de código-fonte aberto dedicado a prover recursos de desenvolvimento de software (NETBEANS, 2015). Java é uma linguagem de programação de alto nível orientada a objetos, que funciona em qualquer plataforma que possua uma máquina virtual Java. Além de ser uma linguagem robusta e facilitadora para o desenvolvedor, Java é *multithread*, o que possibilita várias tarefas em apenas um processo. O software desenvolvido está em sua versão 1.0 e é compatível com o Sistema de Gerenciamento de Banco de Dados (SGBD) PostgreSQL.

O sistema possui dois tipos de agentes:

- Agentes simples: são criados pelos usuários do sistema. Possuem uma descrição (nome do agente), um perfil, que é o modelo interno do agente e que será usado para minerar os dados, e um campo para metadados, onde o usuário pode inserir informações sobre o agente (que tipo de dados ele busca e para qual tipo de imagem ele é adequado, por exemplo);
- Agente controlador: imerso no sistema, esse agente monitora os agentes simples. Se houver um impasse entre dois agentes em um dado segmento de uma imagem, o controlador é o responsável por identificar e marcar esse segmento para o usuário.

Cada agente simples no sistema é representado por uma *thread*. Dessa forma, eles poderão atuar juntos no ambiente. Todos os agentes simples são armazenados em um único arquivo denominado de "agentes.dat". Sendo assim, tem-se uma fácil manipulação dos agentes, dando a oportunidade aos usuários de compartilharem o arquivo com seus agentes com outros usuários do AgentGeo.

A Figura 20 apresenta a tela de conexão com o banco de dados, onde o usuário irá conectar o AgentGeo ao banco de dados gerenciado pelo TerraView. Para criar o ambiente, ou seja, o local onde os agentes irão realizar a mineração, tem-se a tela para seleção das tabelas que possuem os dados das imagens, como observa-se na Figura 21.

**Figura 20 – AgentGeo: tela de conexão com o banco de dados.**

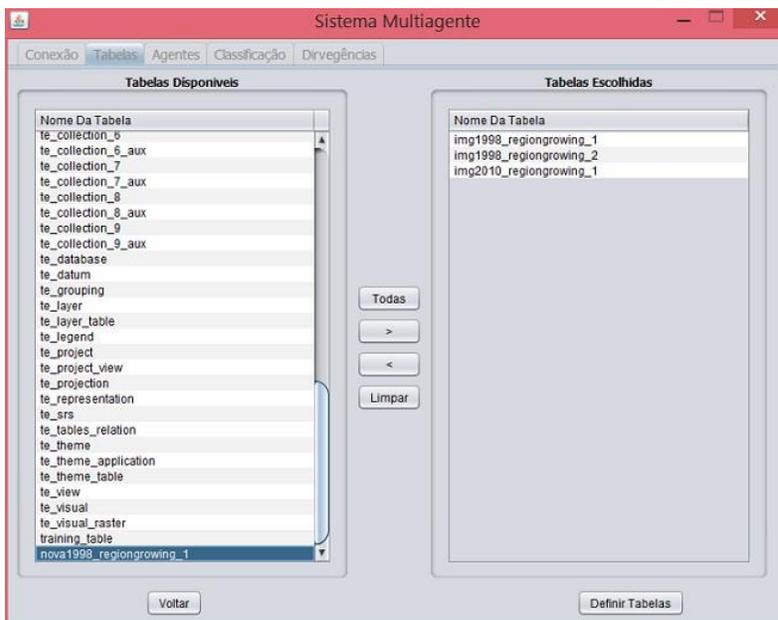
A imagem mostra a interface de usuário do sistema "Sistema Multiagente". No topo, há uma barra de menu com as opções "Conexão", "Tabelas", "Agentes", "Classificação" e "Divergências". O menu "Conexão" está selecionado. O conteúdo principal da tela é um formulário intitulado "Conexão com o Banco De Dados". Este formulário contém os seguintes campos e botões:

- Banco de Dados: [Menu suspenso]
- Servidor: [Campo de texto]
- Porta: [Campo de texto]
- Usuário: [Campo de texto]
- Senha: [Campo de texto]
- Botão "Conectar" (localizado à direita dos campos de usuário e senha)
- Botão "Selecionar Banco" (localizado à esquerda do menu suspenso)
- Nome do Banco: [Campo de texto]

Fonte: Autoria própria.

Após a criação do ambiente (seleção das tabelas com os dados das imagens), o usuário seleciona os agentes que irão atuar. A Figura 22 apresenta a tela para realizar essa seleção dos agentes. Na Figura 23, observa-se a tela onde o usuário pode incluir, editar e excluir um agente.

**Figura 21 – AgentGeo: tela de seleção das tabelas para criar o ambiente.**



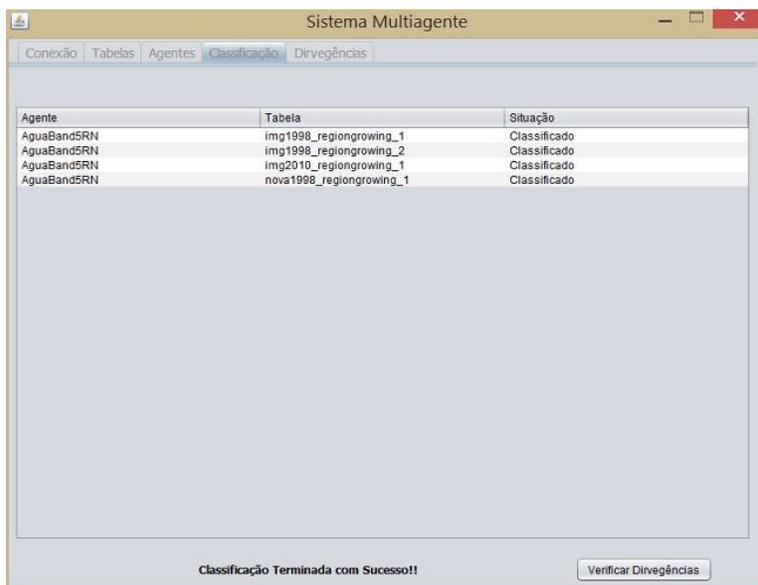
Fonte: Autoria própria.



A Figura 24 apresenta a tela que exibe os resultados da interação dos agentes com o ambiente. Caso haja divergência entre os Agentes Simples, essa será apresentada em outra tela, como observa-se na Figura 25. São exibidos nessa tela: os agentes, as tabelas (ambiente) e a quantidade de segmentos marcados pelo agente controlador que aponta as divergências.

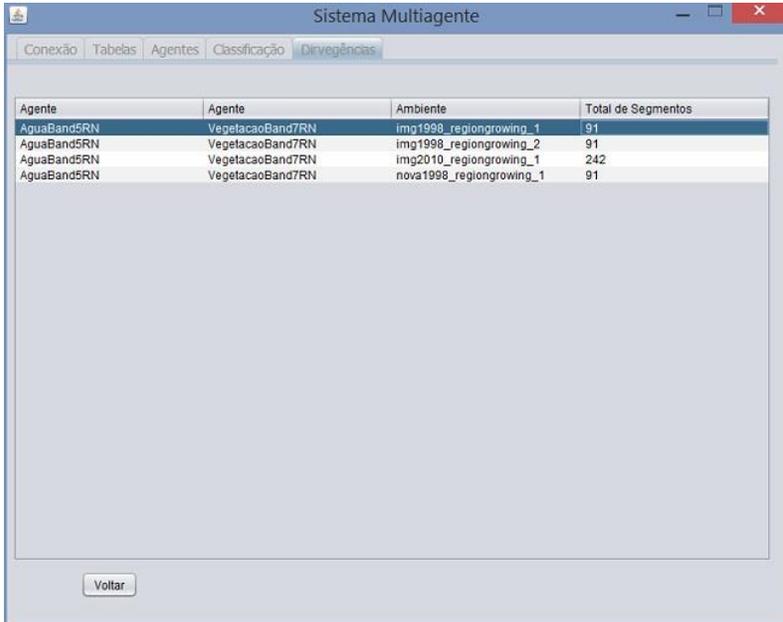
A Figura 26 apresenta todas as telas do sistema segundo o esquema de funcionamento: (A) Conexão com o banco de dados; (B) Seleção das tabelas que possuem os dados das imagens (Criação do ambiente); (C) Seleção dos agentes que irão atuar no ambiente; (D) Inclusão, edição e exclusão dos agentes; (E) Interação dos agentes com o ambiente; (F) Resultados das divergências entre os Agentes Simples. Pode-se perceber que após iniciar a interação dos agentes com o ambiente, da tela (C) para a (E), não há como retornar no processo e modificar o ambiente ou os Agentes Simples.

**Figura 24 – AgentGeo: tela com resultados da interação entre os agentes com o ambiente.**



Fonte: Autoria própria.

**Figura 25 – AgentGeo: tela com divergência entre os Agentes Simples.**



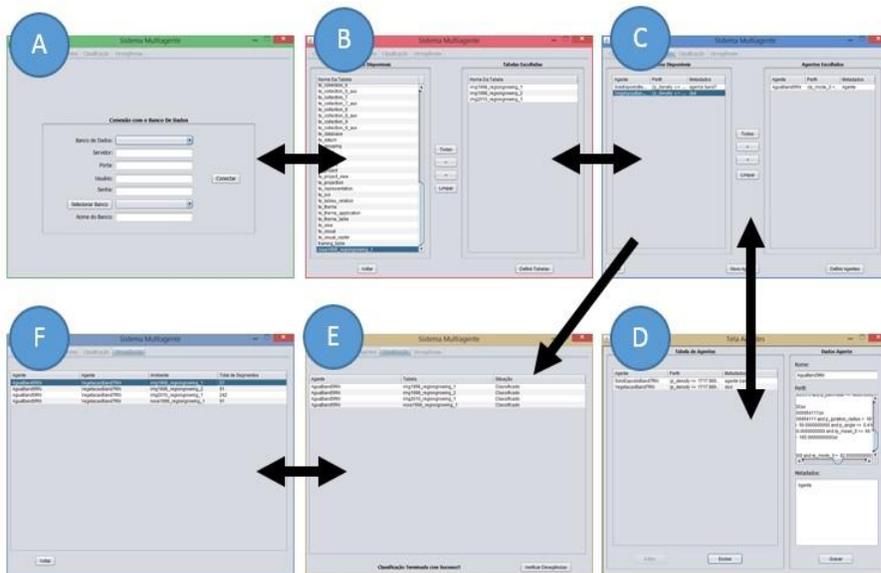
The screenshot shows a window titled 'Sistema Multiagente' with a menu bar containing 'Conexão', 'Tabelas', 'Agentes', 'Classificação', and 'Divergências'. The 'Divergências' tab is active, displaying a table with the following data:

Agente	Agente	Ambiente	Total de Segmentos
AguaBand5RN	VegetacaoBand7RN	img1998_regiongrowing_1	91
AguaBand5RN	VegetacaoBand7RN	img1998_regiongrowing_2	91
AguaBand5RN	VegetacaoBand7RN	img2010_regiongrowing_1	242
AguaBand5RN	VegetacaoBand7RN	nova1998_regiongrowing_1	91

At the bottom of the window, there is a button labeled 'Voltar'.

Fonte: Autoria própria.

**Figura 26 – Telas do sistema: esquema de funcionamento do AgentGeo.**



Fonte: Autoria própria.

Os agentes do sistema são do tipo reativo simples, sua decisão é com base na percepção atual do ambiente, ignorando seu histórico de percepção (RUSSELL; NORVING, 2010). Esses agentes são competitivos, se preocupando somente em realizar as suas tarefas. Cabe ao agente controlador receber os modelos dos agentes simples e verificar se há impasses entre eles. Sobre o ambiente onde esses agentes estão trabalhando, pode-se dizer que ele é completamente observável, determinístico, episódico, estático, discreto e multiagente.

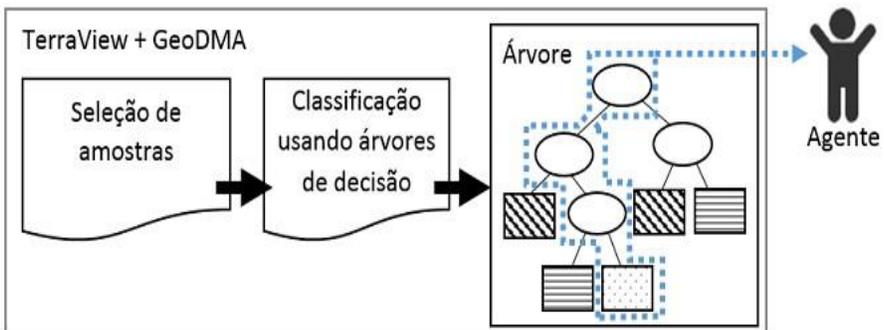
O processo de criação dos agentes simples possui duas etapas:

- **Aprendizagem do agente:** Elaboração do perfil do agente, ou seja, qual será o modelo interno que o agente utilizará para realizar a mineração dos dados. Neste livro, os modelos foram criados através do GeodMA. Após realizar as etapas de: seleção da imagem, segmentação, extração de características e treinamento,

é realizada a classificação usando, neste trabalho, árvores de decisão. Esta árvore fornece o modelo que será utilizado pelo agente. A Figura 27 apresenta essa etapa.

- Validação do agente: O usuário pode usar o agente em alguns testes a fim de validar o modelo. Caso os resultados não sejam satisfatórios, ele pode retornar à etapa de aprendizagem, ou seja, usando outra vez o GeoDMA, mais amostras podem ser selecionadas na etapa de treinamento ou amostras selecionadas erroneamente podem ser retiradas. Se o modelo do agente for válido e não precisar de mais modificações, o agente permanece armazenado no arquivo "agentes.dat" e pode ser usado sempre que necessário.

**Figura 27 – Etapa de aprendizagem do agente.**



Fonte: Autoria própria.

Pode-se observar que o AgentGeo interage com o TerraView e o GeoDMA, o qual realiza todo o processo para minerar dados em uma única imagem. Essa proposta visa superar esta limitação através de 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 AgentGeo tem como objetivo realizar a mineração desses dados através do uso de agentes.

A Figura 28 apresenta o diagrama do processo usando as duas ferramentas: (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. Também é realizada a etapa de aprendizagem do agente; (3) Esses atributos são armazenados no banco de dados que é gerenciado pelo TerraView; (4-A) A Ferramenta é conectada ao mesmo banco de dados gerenciado pelo TerraView. (4-B) Como forma de criar o ambiente para os agentes, o usuário seleciona as tabelas do banco de dados que estão armazenando os atributos extraídos. (4-C e D) O usuário seleciona, cria e/ou edita os agentes que serão utilizados para a mineração visando atingir os resultados esperados. (4-E) Nessa etapa há a interação dos agentes com o ambiente. Também entra em ação o Agente Controlador, que verifica a ocorrência de impasses entre os Agentes Simples. Caso ocorra impasse, o Agente Controlador irá identificar o segmento, que pode ser visto pelo usuário no TerraView. (4-F) Resultado dos impasses entre os agentes Simples. (5) Por fim, o usuário pode visualizar os resultados das imagens processadas através do TerraView.

### 3.4 AVALIAÇÃO E INTERPRETAÇÃO

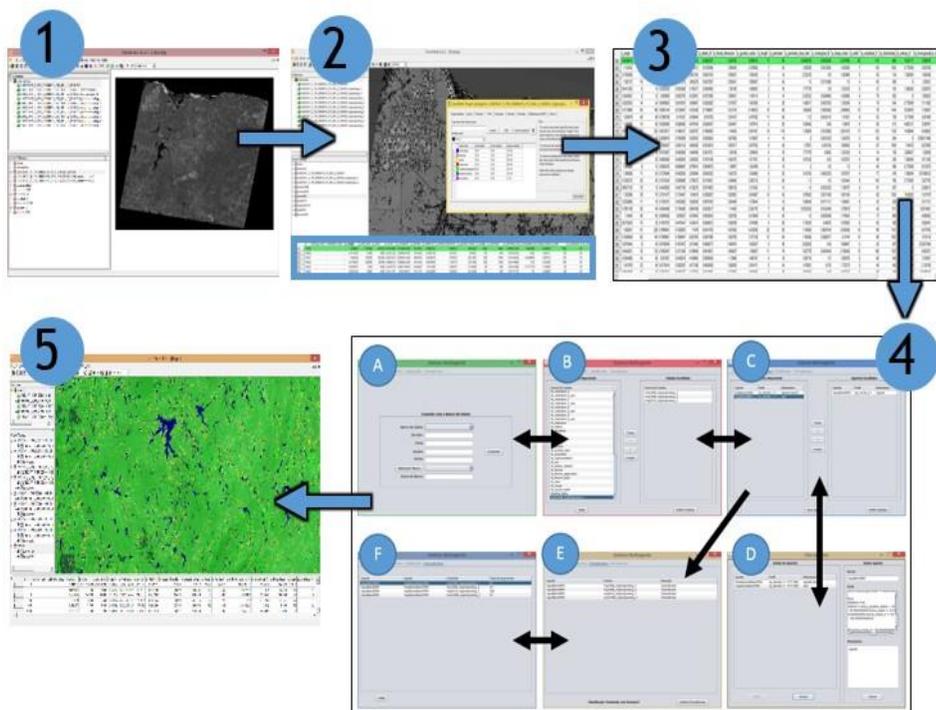
Após a etapa de mineração usando os agentes, os dados podem, finalmente, ser exibidos no TerraView. Para que isso ocorra, por enquanto o usuário deve, no TerraView, editar as legendas das imagens que possuem os segmentos e as características desses segmentos.

Alteram-se os Parâmetros de Agrupamento para modo Valor Único, seleciona-se o atributo que possui o resultado da etapa de mineração e, caso o usuário assim deseje, modifica-se a aparência dos dados que serão exibidos. Versões futuras do software automatizarão este procedimento.

A partir dessas informações exibidas no TerraView, o intérprete pode realizar uma avaliação e, se os resultados não forem satisfatórios, ele pode retornar e executar novamente alguma das etapas do processo.

No Capítulo 4 serão apresentados dois estudos de caso que exemplificam o uso da metodologia proposta e a utilização do software desenvolvido.

**Figura 28 – Diagrama do processo usando as ferramentas: TerraView, GeoDMA e AgentGeo.**



Fonte: Autoria própria.

## CAPÍTULO 4 ESTUDOS DE CASO E RESULTADOS

Foram realizados dois estudos de caso, um com várias cenas da região nordeste do Brasil e outro com o estado do Rio Grande do Norte (RN), que exemplificam e avaliam o uso de agentes na mineração de imagens em diferentes tipos de estudos.

A Tabela 3 apresenta as áreas avaliadas e os agentes criados e utilizados para minerar os dados das imagens:

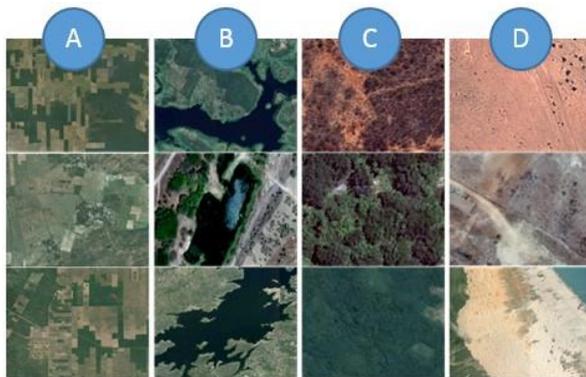
Tabela 3 – Áreas de estudo e agentes utilizados.

Área	Agente			
	Corpos D'água	Solo Exposto	Área de Cultivo	Vegetação
Região Nordeste	x	x	x	
Rio Grande do Norte	x	x		x

Fonte: Autoria própria.

Para estes estudos, cada agente busca por determinados padrões referentes à paisagem. A Figura 29 apresenta algumas amostras do que foi considerado nesses estudos como sendo (A) áreas de cultivo, (B) corpos d'água, (C) vegetação e (D) solo exposto.

**Figura 29 – Imagens exemplificando os tipos de objetos buscados: (A) áreas de cultivo, (B) corpos d'água, (C) vegetação e (D) solo exposto.**



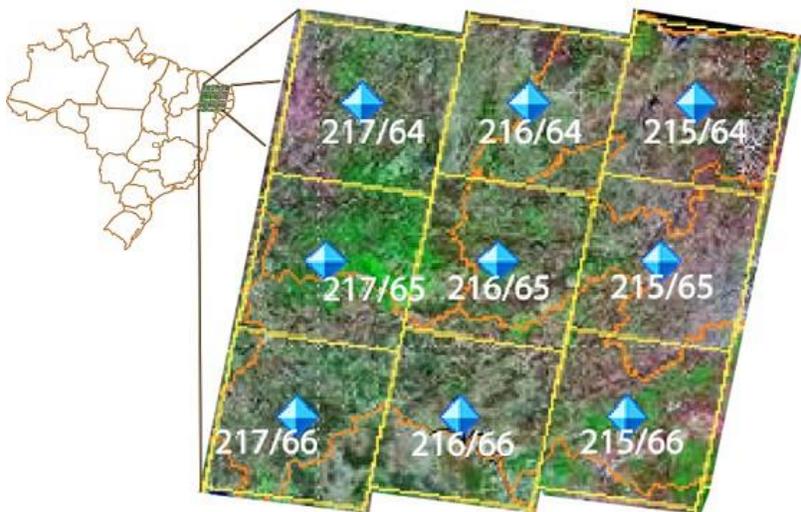
Fonte: (GOOGLE, 2015).

#### 4.1 ESTUDO DE CASO 1: NORDESTE

O primeiro estudo de caso foi realizado com várias cenas do LANDSAT-5. O objetivo deste estudo de caso é encontrar corpos d'água, áreas onde o solo está exposto e também áreas de cultivo em várias regiões do nordeste. Ou seja, um usuário utilizando a metodologia proposta, poderia buscar um objeto de interesse em diversas cenas simultaneamente ao invés de buscar esse objeto em uma cena por vez.

Do LANDSAT-5, foi selecionada a banda 5 (infravermelho médio) que, mesmo sofrendo perturbações caso haja chuva antes da obtenção da cena, é a mais indicada para o estudo. Então, foi utilizada apenas essa banda. A Figura 30 apresenta os locais (órbitas e pontos) das cenas. Pode-se perceber que elas abrangem uma boa parte da região Nordeste, com áreas dos seguintes estados: Rio Grande do Norte, Ceará, Paraíba, Pernambuco, Alagoas, Piauí e Bahia. A Tabela 4 indica os anos de cada uma destas cenas.

**Figura 30 – Localização das cenas utilizadas.**



Fonte: Autoria própria.

Tabela 4 – Cenas do LANDSAT-5 utilizadas.

<b>Cena</b>	<b>Mês/Ano</b>
215/64	07/2005
216/64	07/2005
217/64	08/2011
215/65	08/2010
216/65	09/2009
217/65	08/2011
215/66	09/2010
216/66	10/2010
217/66	11/2009

Fonte: Autoria própria.

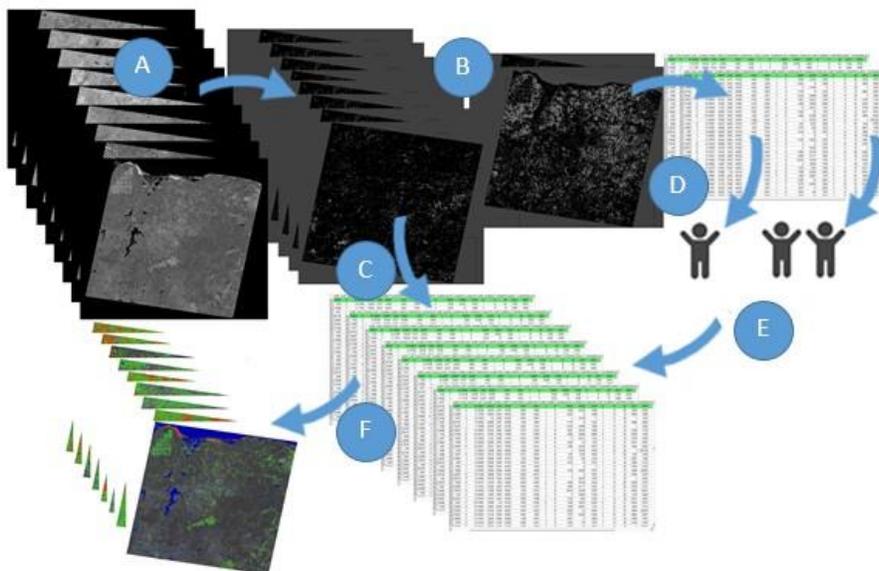
Usando a metodologia apresentada neste trabalho, todas as cenas foram segmentadas usando o algoritmo de Crescimento de Regiões e depois, as características de cada um dos segmentos foram extraídas. Usando a cena 216/64, foram selecionados alguns segmentos como amostras e, em seguida, foi criada a árvore de decisão que serviu de modelo para o Agente Corpos D'água. Utilizando a cena 215/64, também foram selecionados alguns segmentos como amostras e criados dois modelos com base nessa árvore. Este procedimento fornece modelos para criar dois agentes: Agente Áreas de Cultivo e Agente Solo Exposto.

Os três agentes podem ser usados de forma individual ou em conjunto, ou seja, podem apenas buscar por objetos que tenham o padrão referente a corpos d'água dentro de uma imagem ou procurar objetos de diferentes padrões de uma vez. Eles também podem atuar em todo o ambiente, ou seja, em todas as tabelas do banco de dados que possuem os dados dos segmentos, ou em parte dele, isto é, apenas em algumas das tabelas.

As etapas do desenvolvimento desse estudo podem ser observadas na Figura 31: (A) Imagens inseridas no TerraView; (B)

Imagens segmentadas usando Crescimento de Regiões - os valores para área mínima e similaridade foram 16 e 8, respectivamente; (C) As características dos segmentos foram extraídas; (D) Treinamento e geração das árvores de decisão, cada qual com base em uma cena; Geração do modelo para o Agente Corpos D'água e dos modelos para os Agentes Áreas de Cultivo e Solo Exposto; (E) Uso do AgentGeo para minerar os dados de todas as outras imagens; (F) Apresentação dos resultados no TerraView.

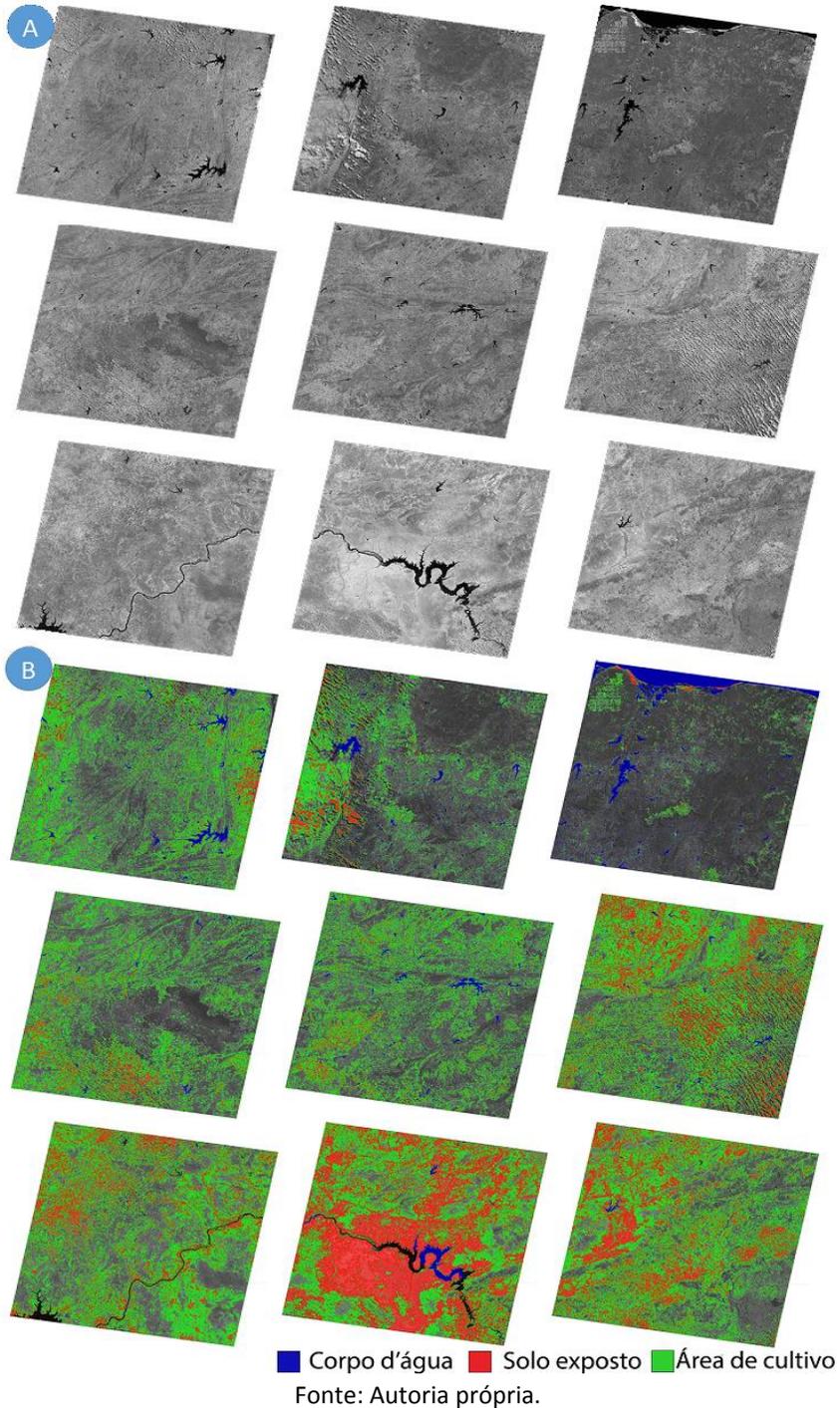
**Figura 31 – Etapas do desenvolvimento do primeiro estudo de caso.**



Fonte: Autoria própria.

A Figura 32 apresenta (A) as cenas originais e (B) as cenas após a mineração utilizando os agentes. Esse processo poderia ter sido realizado no GeoDMA, mas em uma imagem por vez. Utilizando a metodologia proposta neste trabalho, todos os dados das imagens foram minerados simultaneamente pelos agentes.

**Figura 32 – Resultado do estudo de caso: (A) Cenas originais e (B) cenas após a mineração utilizando os agentes.**



Para validar esse estudo de caso, foi realizado um processo de inspeção visual utilizando o Google Maps (GOOGLE, 2015). Para cada uma das nove cenas, 20 amostras foram selecionadas para cada um dos três agentes, totalizando assim 540 amostras avaliadas. Cada amostra era comparada com a mesma região apresentada no Google Maps. A Tabela 5 apresenta o resultado desta avaliação. Já na Tabela 6, pode-se observar os resultados de cada agente.

O agente que obteve mais sucesso foi o Agente Corpo D'água, pois em apenas uma das nove cenas o modelo usado pelo agente não foi eficaz e isso resultou em nove amostras classificadas de forma errada. Já os Agentes Solo Exposto e Áreas de Cultivo classificaram, de forma errada, várias amostras que na verdade são zonas urbanas ou sombra de nuvens.

Durante a coleta de amostras para criar o modelo do Agente Solo Exposto, não foram coletadas amostras que excluíssem as nuvens da classe solo exposto. Dessa forma, a presença de nuvens na imagem ocasionou erros na classificação. O mesmo aconteceu com o Agente Áreas de Cultivo e com amostras relacionadas a zonas urbanas. Esta etapa da aprendizagem do agente, onde é elaborado seu modelo, é uma das mais importantes do processo proposto nesta abordagem.

Esse estudo de caso mostra a metodologia proposta colocada em prática. Mesmo com 90.37% das amostras classificadas corretamente, é possível obter melhores resultados. Para isso, mais amostras relevantes devem ser selecionadas na etapa de treinamento. Assim, os modelos para os agentes terão melhor qualidade. A falta de amostras relevantes pode levar segmentos que não são daquela classe específica a serem classificados de forma errada.

Tabela 5 – Resultados da avaliação em todas as cenas através do processo de inspeção visual.

Cena	Agente	Total de amostras	Amostras classificadas corretamente	Amostras classificadas erroneamente
215/64	Corpo D'água	20	20	0
	Solo Exposto	20	19	1
	Áreas de Cultivo	20	18	2
	<b>Total</b>	<b>60</b>	<b>57 (95%)</b>	<b>3 (5%)</b>
216/64	Corpo D'água	20	20	0
	Solo Exposto	20	14	6
	Áreas de Cultivo	20	17	3
	<b>Total</b>	<b>60</b>	<b>51 (85%)</b>	<b>9 (15%)</b>
217/64	Corpo D'água	20	20	0
	Solo Exposto	20	16	4
	Áreas de Cultivo	20	17	3
	<b>Total</b>	<b>60</b>	<b>53 (88%)</b>	<b>7 (12%)</b>
215/65	Corpo D'água	20	20	0
	Solo Exposto	20	15	5
	Áreas de Cultivo	20	20	0
	<b>Total</b>	<b>60</b>	<b>55 (92%)</b>	<b>5 (8%)</b>
216/65	Corpo D'água	20	20	0
	Solo Exposto	20	17	3
	Áreas de Cultivo	20	18	2
	<b>Total</b>	<b>60</b>	<b>55 (92%)</b>	<b>5 (8%)</b>
217/65	Corpo D'água	20	20	0
	Solo Exposto	20	17	3
	Áreas de Cultivo	20	18	2
	<b>Total</b>	<b>60</b>	<b>55 (92%)</b>	<b>5 (8%)</b>
215/66	Corpo D'água	20	20	0
	Solo Exposto	20	18	2
	Áreas de Cultivo	20	20	0
	<b>Total</b>	<b>60</b>	<b>58 (97%)</b>	<b>2 (3%)</b>
216/66	Corpo D'água	20	20	0
	Solo Exposto	20	18	2
	Áreas de Cultivo	20	17	3
	<b>Total</b>	<b>60</b>	<b>55 (92%)</b>	<b>5 (8%)</b>
217/66	Corpo D'água	20	11	9
	Solo Exposto	20	18	2
	Áreas de Cultivo	20	20	0
	<b>Total</b>	<b>60</b>	<b>49 (82%)</b>	<b>11 (18%)</b>

Fonte: Autoria própria.

Tabela 6 – Resultados da avaliação para cada agente.

<b>Agente</b>	<b>Total de amostras avaliadas</b>	<b>Amostras classificadas corretamente</b>	<b>Amostras classificadas erroneamente</b>
Corpo D'água	180	171	9
Solo Exposto	180	152	28
Áreas de Cultivo	180	165	15
<b>Total</b>	<b>540</b>	<b>488 (90%)</b>	<b>52 (10%)</b>

Fonte: Autoria própria.

#### 4.2 ESTUDO DE CASO 2: ESTADO DO RIO GRANDE DO NORTE

Para o segundo estudo de caso, será utilizada como área de estudo todo o estado do Rio Grande do Norte. O objetivo é utilizar a metodologia proposta neste trabalho para identificar corpos d'água, solo exposto e vegetação em todo o RN. Então, com base em amostras de diversas áreas do estado, serão usados agentes para identificar estes objetos na área escolhida.

Primeiramente foi criado um mosaico, que tem por objetivo unir em uma única imagem duas ou mais cenas. Para compor o mosaico, foram utilizadas oito cenas do satélite LANDSAT-8.

Dentre as onze bandas do LANDSAT-8, foi selecionada a banda 5 para realizar a parte do estudo que envolve a identificação dos corpos d'água, pois ela mede o infravermelho próximo enfatizando o teor de biomassa. Nesta banda, os corpos d'água possuem uma resposta espectral mais baixa em relação a outros objetos da paisagem (USGS, 2014c).

A banda 7 foi selecionada para realizar a outra parte do estudo, que envolve a identificação de áreas com solo exposto e também áreas

com vegetação, pois esta banda discrimina melhor o teor da umidade do solo e da vegetação (USGS, 2014c).

Pode-se observar na Tabela 7 as cenas utilizadas para formar o mosaico. Em todas elas, foi realizada uma conversão radiométrica de 12 bits para 8 bits. O software utilizado para este procedimento foi o *Geospatial Data Abstraction Library* (GDAL), que é uma biblioteca geoespacial para tradução de formatos de dados raster. Apesar de não possuir interface gráfica de usuário, o GDAL possui muitas funcionalidades relevantes (GDAL, 2014).

Após realizar o processo de elaboração do mosaico, foi utilizado um shape do RN para extrair do mosaico apenas a área do Estado, já que o mosaico abrangia não só o RN, mas também áreas de estados vizinhos. Shape é um tipo de arquivo digital que representa uma feição ou elemento vetorial, seja ela em formato de ponto, linha ou polígono e que contém uma referência espacial (coordenadas geográficas) de qualquer que seja o elemento mapeado (SEMACE, 2010).

Tabela 7 - Cenas do LANDSAT-8 utilizadas.

<b>Cena</b>	<b>Ano</b>
215/063	2013
216/063	2014
214/064	2014
215/064	2013
216/064	2014
214/065	2014
215/065	2013
216/065	2014

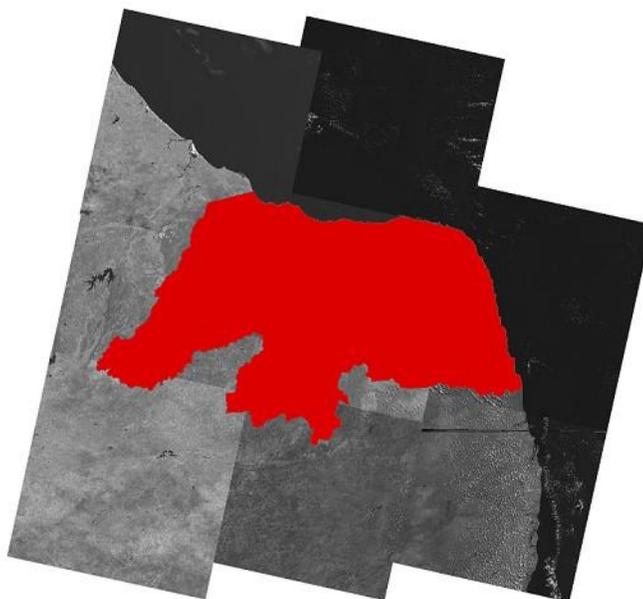
Fonte: Autoria própria.

A Figura 33 apresenta o mosaico com a sobreposição do *shape* do RN, delimitando assim a área que será utilizada no estudo. Na Figura 34 pode-se observar o resultado com a banda 5 do LANDSAT-8: todo o

estado do Rio Grande do Norte. A Figura 35 apresenta o mesmo resultado, mas com a banda 7 do LANDSAT-8.

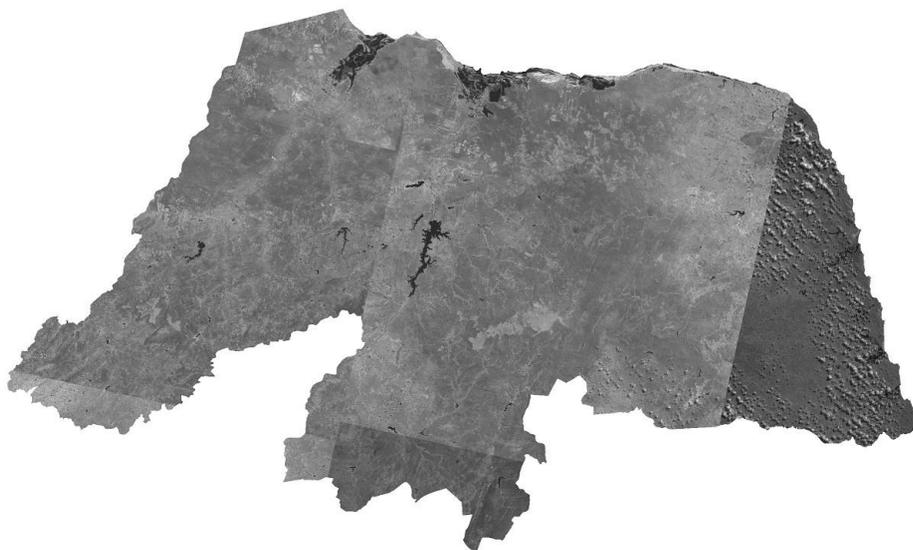
Seguindo as etapas da metodologia proposta, as cenas e o *shape* do RN foram inseridos no TerraView, foi realizada a elaboração do mosaico e, utilizando o *shape* do RN, foi extraída apenas a área desse estado. Então, usando o GeoDMA, as duas imagens do estado foram segmentadas com o algoritmo de Crescimento de Regiões. Os valores de similaridade e área mínima foram, 4 e 38, respectivamente. Vários testes foram realizados até se chegar a esses valores, que apresentaram melhor os detalhes dos segmentos da imagem. O resultado dessa segmentação pode ser observado na Figura 36.

**Figura 33 – Mosaico com cenas do LANDSAT-8 e Shape do Rio Grande do Norte.**



Fonte: Autoria própria.

**Figura 34 – Área de estudo com a banda 5 do LANDSAT-8.**



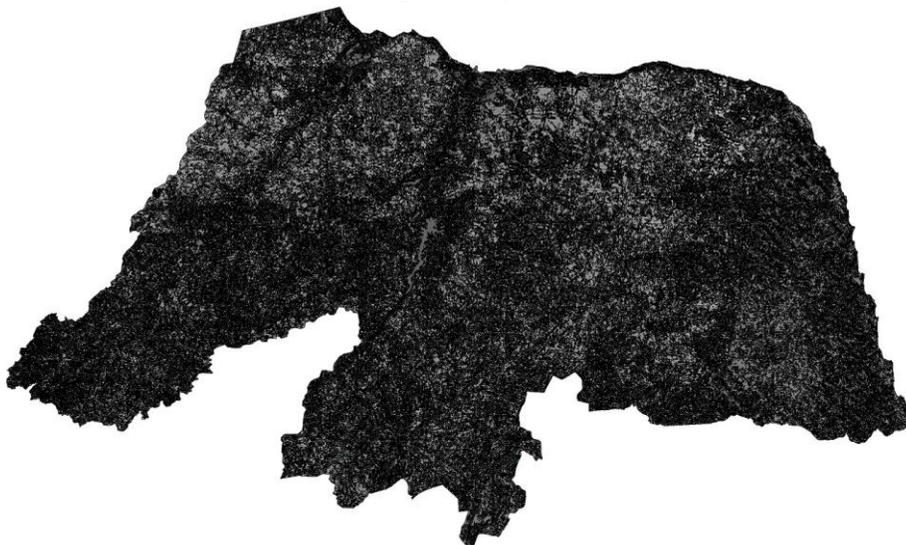
Fonte: Autoria própria.

**Figura 35 – Área de estudo com a banda 7 do LANDSAT-8.**



Fonte: Autoria própria.

**Figura 36 – Área de estudo: Estado do Rio Grande do Norte após a segmentação.**



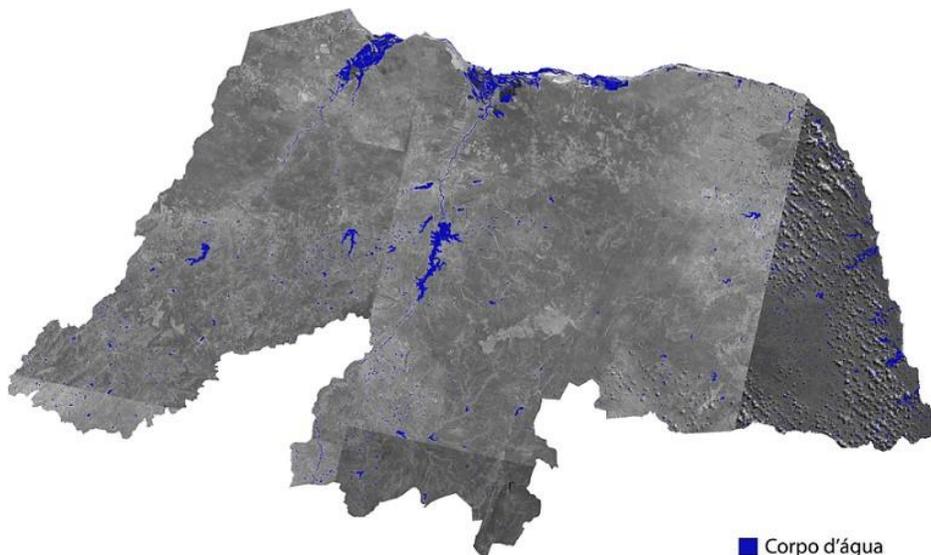
Fonte: Aatoria própria.

Após a segmentação, as características de cada um dos segmentos foram extraídas. Logo depois, os modelos dos agentes foram criados. Primeiro para o Agente Corpos D'água, que interage com o ambiente criado pela imagem do RN com a banda 5 do LANDSAT-8, e em seguida os Agentes Solo Exposto e Vegetação, que interagem com o a imagem do RN, mas com a banda 7 do LANDSAT-8.

Para o Agente Corpos D'água, que tem como objetivo identificar corpos d'água no RN através da banda 5 do LANDSAT-8, a base de conhecimento é composta por 2357 amostras, selecionadas na etapa de aprendizagem, onde 1875 são amostras de áreas que não são água e 482 amostras de áreas que são corpos d'água.

Após essa etapa de treinamento, onde algumas amostras foram coletadas, uma árvore de decisão foi gerada. Esta árvore serviu de modelo para o agente, que foi o responsável por classificar todo o restante da imagem. A Figura 37 apresenta o resultado desse processo com o Agente Corpos D'água, que atua somente na banda 5.

**Figura 37 – Resultado do processo de mineração utilizando o Agente  
Corpos D'água.**

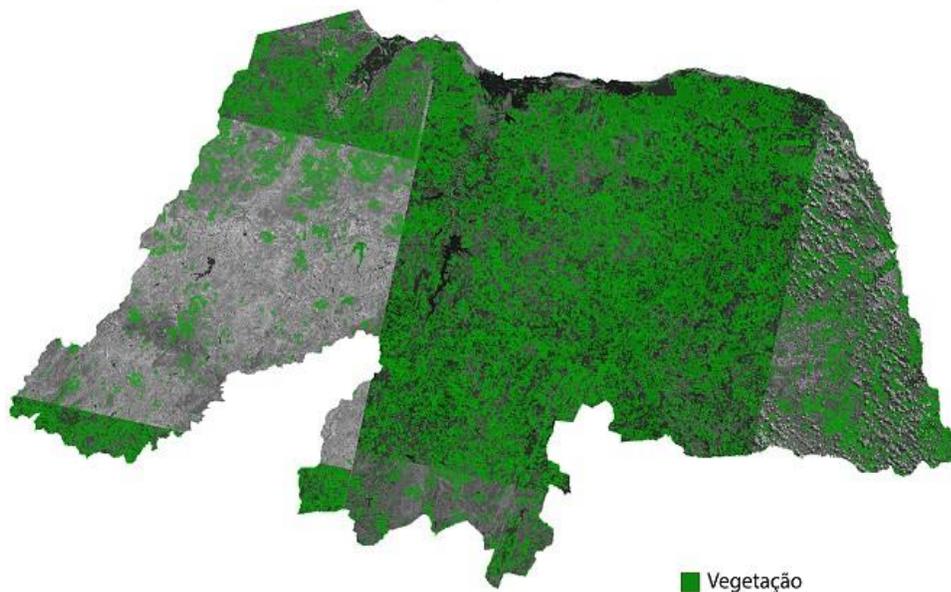


Fonte: Autorial própria.

Para o Agente Vegetação, que tem como objetivo identificar a vegetação no RN utilizando a banda 7 do LANDSAT-8, a base de conhecimento é composta por 1811 amostras, sendo que 1658 são amostras de áreas que não são vegetação e 153 amostras correspondem à vegetação. E para o Agente Solo Exposto, que tem como objetivo encontrar áreas onde o solo está exposto no RN utilizando a banda 7 do LANDSAT-8, a base de conhecimento é composta por 1811 amostras. Dessas amostras, 1555 não são áreas de solo exposto e 256 são.

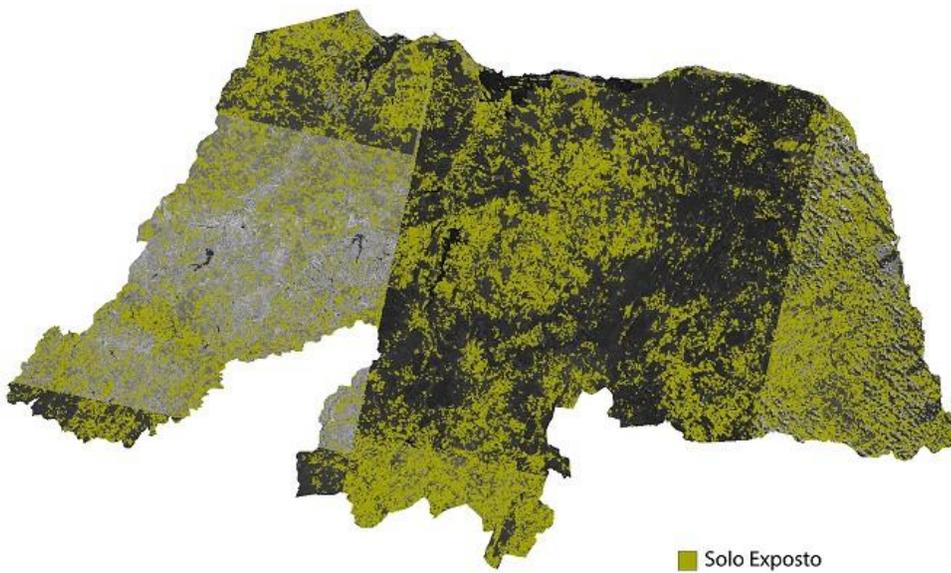
Os agentes Vegetação e Solo Exposto utilizam a banda 7. Na Figura 38 pode-se observar o resultado da classificação do Agente Vegetação e na Figura 39 o resultado da classificação do Agente Solo Exposto. Mesmo com dois agentes no mesmo ambiente, os resultados são apresentados separadamente para uma melhor avaliação e análise.

**Figura 38 – Resultado do processo de mineração utilizando o Agente Vegetação.**



Fonte: Autoria própria.

**Figura 39 – Resultado do processo de mineração utilizando o Agente Solo Exposto.**

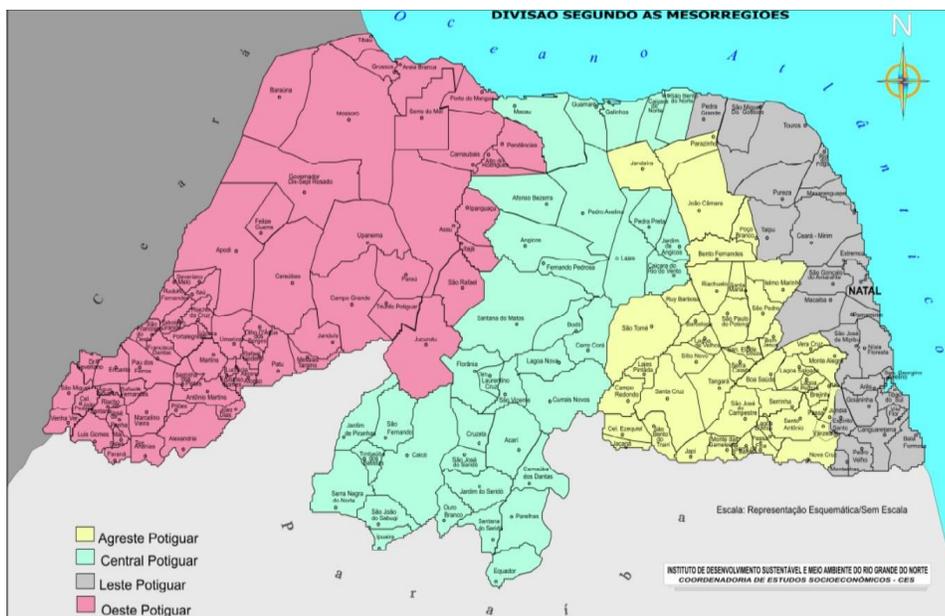


Fonte: Autoria própria.

Embora a classificação com dois ou mais agentes possa gerar conflitos entre estes, nesse estudo de caso não houve nenhuma divergência. Desta forma, não houve interação do agente controlador com o ambiente, houve somente com os agentes simples.

Como forma de validar esse estudo de caso, foi realizada uma avaliação de algumas áreas classificadas pelos agentes. Primeiro o Estado foi dividido em áreas menores. Para isso, foram usadas as mesorregiões geográficas do Rio Grande do Norte que compõem o RN em quatro partes: Agreste potiguar, Central potiguar, Leste potiguar e Oeste potiguar. A Figura 40 apresenta o mapa do Rio Grande do Norte e suas mesorregiões geográficas.

**Figura 40 – Mesorregiões do Rio Grande do Norte.**



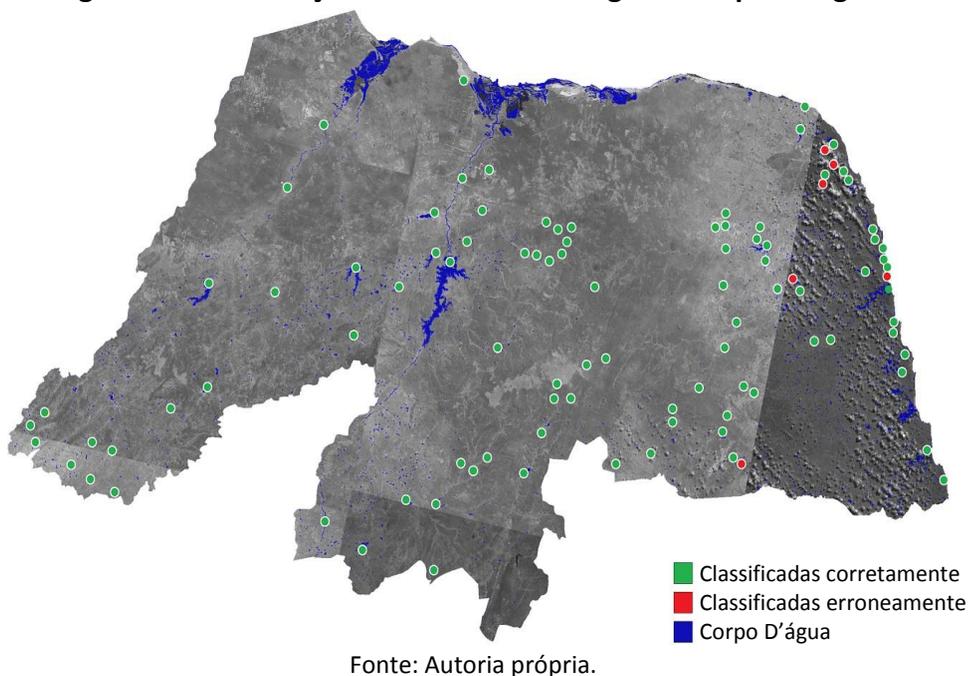
Fonte: (IDEMA, 2014)

De cada uma dessas mesorregiões, foram selecionadas 25 amostras. Em seguida, uma avaliação através de inspeção visual utilizando o Google Maps (GOOGLE, 2015). Isso foi feito para cada um

dos três agentes, ou seja, foram avaliadas 100 amostras para cada um dos agentes.

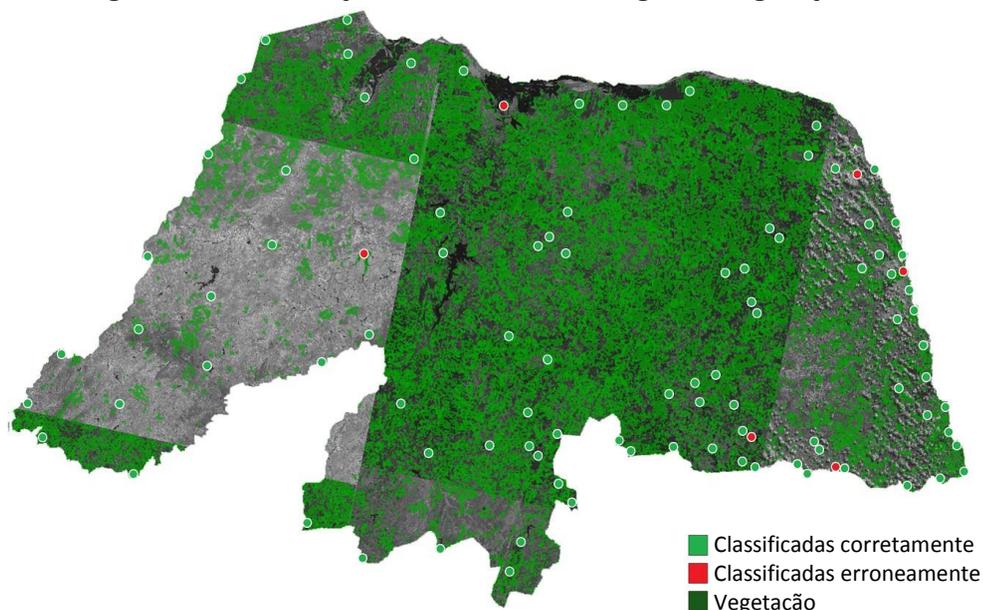
As Figuras 41, 42 e 43, apresentam os resultados da classificação realizada por cada um dos agentes e também a localização das amostras avaliadas. Em verde figuram as amostras classificadas corretamente. Já em vermelho aparecem as amostras que foram classificadas erroneamente.

**Figura 41 – Localização das amostras do Agente Corpo D’água.**



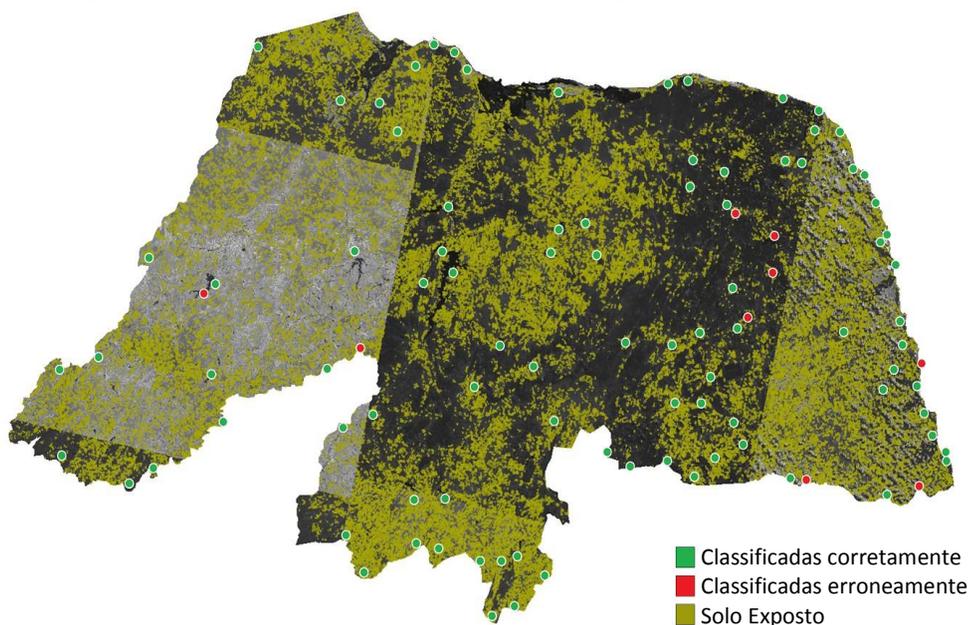
A Tabela 8 apresenta os resultados da avaliação através do processo de inspeção visual. Pode-se perceber que das 100 amostras para o Agente Corpo D’água, apenas seis foram classificadas de forma errada, dentre as quais, cinco são sombras de nuvens que foram confundidas com corpos d’água.

**Figura 42 – Localização das amostras do Agente Vegetação.**



Fonte: Autoria própria.

**Figura 43 – Localização das amostras do Agente Solo Exposto.**



Fonte: Autoria própria.

Tabela 8 – Resultados da avaliação através do processo de inspeção visual.

Agente	Mesorregião	Quantidade de amostras	Amostras classificadas corretamente	Amostras classificadas erroneamente
<b>Corpos D'água</b>	Leste Potiguar	25	21	4
	Agreste Potiguar	25	23	2
	Central Potiguar	25	25	0
	Oeste Potiguar	25	25	0
	<b>Total</b>	<b>100</b>	<b>94</b>	<b>6</b>
<b>Vegetação</b>	Leste Potiguar	25	23	2
	Agreste Potiguar	25	23	2
	Central Potiguar	25	24	1
	Oeste Potiguar	25	24	1
	<b>Total</b>	<b>100</b>	<b>94</b>	<b>6</b>
<b>Solo Exposto</b>	Leste Potiguar	25	22	3
	Agreste Potiguar	25	20	5
	Central Potiguar	25	25	0
	Oeste Potiguar	25	23	2
	<b>Total</b>	<b>100</b>	<b>90</b>	<b>10</b>

Fonte: Autoria própria.

O Agente Vegetação também classificou erroneamente seis amostras. Algumas amostras eram corpos d'água, mas hoje apresentam mata ciliar. Mesmo assim, tais amostras foram marcadas como classificadas incorretamente. Outras áreas eram sombras de nuvens ou áreas de salinas e foram classificadas erroneamente como vegetação.

Já o Agente Solo Exposto classificou dez amostras de forma errada. Estas amostras, em sua maioria, eram áreas de zona urbana e foram classificadas como áreas de solo exposto.

Através desta avaliação, constata-se que os Agentes Corpos D'água e Vegetação obtiveram 94% de acerto, enquanto o Agente Solo Exposto classificou 90% das amostras de forma correta.

Esse estudo mostra, mais uma vez, a metodologia proposta sendo colocada em prática. Os agentes podem ser usados de diversas formas, uma vez que foi utilizado um agente para o ambiente criado pela imagem do RN com a banda 5 do LANDSAT-8 e dois agentes para o ambiente criado pela imagem do RN, mas dessa vez com a banda 7 do LANDSAT-8.

## CAPÍTULO 5 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

O uso de agentes e sistemas multiagente no processo de descoberta de conhecimento é um campo novo e promissor que vem sendo aplicado em diversas áreas de pesquisa. Usando conceitos das duas áreas, este livro apresentou uma abordagem que utiliza sistemas multiagente para minerar imagens de sensoriamento remoto. Para aplicar a metodologia proposta, foram utilizadas duas ferramentas: o GeoDMA e o AgentGeo, desenvolvido no contexto da pesquisa abordada neste livro.

O processo de mineração de imagens pode ser realizado no GeoDMA, mas utilizando apenas uma imagem por vez. Já na metodologia proposta neste livro, todos os dados das imagens foram minerados simultaneamente pelos agentes. Como as imagens são armazenadas no banco de dados e os agentes agem nesse banco, o tempo para um agente realizar a mineração é proporcional ao tempo de uma consulta no banco de dados.

Foram realizados dois estudos de caso. No primeiro foram utilizados agentes para encontrar corpos d'água, solo exposto e áreas de cultivo em diversas cenas do satélite LANDSAT 5 que cobrem uma parte da região nordeste do Brasil. Já no segundo, foi possível identificar corpos d'água, áreas com vegetação e áreas com solo exposto em todo o estado do Rio Grande do Norte.

Para validar esta proposta, alguns dos segmentos classificados pelos agentes foram comparados com imagens do Google Maps. Foi observado que a metodologia atende ao propósito de usar agentes para minerar as imagens em diferentes áreas e aplicações.

Algumas limitações foram encontradas durante o desenvolvimento desta proposta e dos estudos de caso. Uma delas é o desafio de se criar um modelo excelente para um agente, pois isto

depende da etapa de treinamento do mesmo e, neste procedimento, é muito importante a seleção de amostras pertinentes e relevantes. Caso o usuário não selecione amostras realmente representativas ou que não pertençam à classe alvo, o agente pode não realizar a mineração da maneira adequada.

De modo geral, o resultado da pesquisa apresentada neste livro tem como principais contribuições a definição de uma metodologia que utiliza um sistema multiagente para minerar dados de imagens de sensoriamento remoto, a implementação do AgentGeo que, em conjunto com o GeoDMA, viabilizam a abordagem proposta e a realização de estudos de caso que permitiram avaliar todo o processo.

Como modificações futuras, espera-se realizar novos estudos de caso, em áreas distintas usando outras imagens (de diferentes satélites, resoluções espaciais ou tamanhos da área de estudo). Para resultados mais precisos, amostras adicionais podem ser selecionadas na etapa de treinamento do agente. Também pode-se expandir os estudos de caso através da criação de outros agentes com objetivos distintos, bem como a implementação de novas funcionalidades no AgentGeo, que certamente permitirão uma ampliação do uso das contribuições nessa área de pesquisa.

## REFERÊNCIAS

BAATZ, M.; SCHAPE, A. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: *Wichmann-Verlag (ed.), XII Angewandte Geographische Informationsverarbeitung*. Heidelberg, Alemanha: Wichmann Verlag, 2000.

BARBOSA, C. et al. *Mapping amazon basin wetlands through region growing segmentation and segmented-based classification jers-1 data*. In: *IX Simposio Latinoamericano de Percepción Remota*. Puerto Iguazú, Argentina: [s.n.], 2000.

BELLIFEMINE, F.; CAIRE, G.; GREENWOOD, D. *Developing Multi-Agent Systems with JADE*. 1. ed. [S.l.]: Wiley, 2007.

BINS, L. S. et al. Satellite imagery segmentation: a region growing approach. In: *VII Simpósio Brasileiro de Sensoriamento Remoto*. Salvador, Brasil: Instituto Nacional de Pesquisas Espaciais - INPE, 1996. p. 677-680.

BUI, T.; LEE, J. An agent-based framework for building decision support systems. *Decis. Support Syst.*, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 25, n. 3, p. 225-237, abr. 1999.

CAO, L.; GORODETSKY, V.; MITKAS, P. A. Agent mining: The synergy of agents and data mining. *Intelligent Systems, IEEE*, v. 24, n. 3, p. 64-72, 2009.

CCRS. *Tutorial: Fundamentals of Remote Sensing*. Paris: Natural Resources Canada Canada Centre for Remote Sensing, 2007.

CHAUHAN, D. *JAFMAS: A Java-based Agent Framework for Multiagent Systems Development and Implementation*. Tese (Doutorado) - University of Cincinnati, Cincinnati, 1997.

CÂMARA, G. et al. *Terralib: An open source gis library for large-scale environmental and socio-economic applications*. In: *Open Source Approaches in Spatial Data Handling*. [s.n.], 2008. v. 2, p. 247-270. Disponível em: <<http://www.terralib.org/docs/papers/TerraLib-OSBook-versionJanuary2008.pdf>>.

DGI. *Imagens LANDSAT TM e ETM*. 2009. Divisão de Geração de Imagens (DGI/INPE). Disponível em: <[http://www.dgi.inpe.br/siteDgi/ATUS\\_LandSat.php](http://www.dgi.inpe.br/siteDgi/ATUS_LandSat.php)>. Acesso em: 31.01.2015.

DGI. *Os satélites LANDSAT 5 e 7*. 2009. Divisão de Geração de Imagens (DGI/INPE). Disponível em: <[http://www.dgi.inpe.br/Suporte/files/Cameras-LANDSAT57\\_PT.php](http://www.dgi.inpe.br/Suporte/files/Cameras-LANDSAT57_PT.php)>. Acesso em: 20.01.2015.

DGI. *Catálogo de Imagens*. 2011. Divisão de Geração de Imagens (DGI/INPE). Disponível em: <<http://www.dgi.inpe.br/CDSR/>>. Acesso em: 20.01.2015.

DIÓGENES, T. C. F. *Mineração de imagens de sensoriamento remoto para detecção de padrões fluviais no rio Mossoró*. Dissertação (Mestrado) - Universidade do Estado do Rio Grande do Norte - UERN e Universidade Federal Rural do Semi-Árido - UFERSA, Mossoró, 2011

FACELI, K. et al. *Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina*. 1. ed. Paris: LTC, 2011. 394 p.

FAYYAD, U. Data mining and knowledge discovery in databases: Implications for scientific datadatabase. *Proc. of the 9th International Conference on Scientific and Statistical Database Management*, p. 2-11, 1997.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI Magazine*, v. 17, p. 37-54, 1996.

GDAL. *Geospatial Data Abstraction Library (GDAL)*. 2014. Disponível em: <<http://www.gdal.org/>>. Acesso em: 08.11.2014.

GOOGLE. *Google Maps*. 2015. Disponível em: <<https://www.google.com.br/maps/>>. Acesso em: 01.01.2015.

IBM. *Saiba o que é o Big Data e os desafios que as empresas enfrentam*. 2012. Disponível em: <[http://www.ibm.com/midmarket/br/pt/infografico\\_bigdata.html](http://www.ibm.com/midmarket/br/pt/infografico_bigdata.html)>. Acesso em: 31.10.2013.

IDEMA. *Perfil do Rio Grande do Norte*. 2014. Instituto de Desenvolvimento Econômico e Meio Ambiente (IDEMA). Disponível em: <<http://www.seplan.rn.gov.br/arquivos/download/PERFIL%20DO%20RN%202014.pdf>>. Acesso em: 18.01.2015.

INPE. *TerraView*. 2014. Instituto Nacional de Pesquisas Espaciais. Disponível em: <<http://www.dpi.inpe.br/terraview>>. Acesso em: 11.12.2014.

KANEKO, C. M. *Uso de Agentes de Software para Busca de Informações em Nível Gerencial*. Dissertação (Mestrado) - Universidade Federal de Santa Catarina - UFSC, Florianópolis, 2005.

KORTING, T. S. *GeoDMA: a toolbox integrating data mining with objectbased and multi-temporal analysis of satellite remotely sensed imagery*. 97 p. Tese (Doutorado) - Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos, 2012.

KORTING, T. S.; DUTRA, L. V.; FONSECA, L. M. G. *A resegmentation approach for detecting rectangular objects in high-resolution imagery*. *IEEE Geoscience and Remote Sensing Letters*, v. 8, n. 4, p. 621-625, 2011. ISSN 1545-598X. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5692808>>.

KORTING, T. S.; FONSECA, L. M. G.; CÂMARA, G. *A geographical approach to self-organizing maps algorithm applied to image segmentation*. In: . [S.l.: s.n.], 2011, (Lecture Notes in Computer Science, v. 6915). p. 162-170.

KORTING, T. S. et al. GeoDMA - um sistema para mineração de dados de sensoriamento remoto. In: *XIV Simpósio Brasileiro de Sensoriamento*

- Remoto. Natal, Brasil: Instituto Nacional de Pesquisas Espaciais - INPE, 2009.
- MARR, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York, NY, USA: Henry Holt and Co., Inc., 1982. ISBN 0716715678.
- MELO, D. H. C. T. B. *Uso de dados Ikonos II na análise urbana: testes operacionais na Zona Leste de São Paulo*. 146 p. Dissertação (Mestrado) - Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos, 2002.
- MENESES, P. R.; ALMEIDA, T. de. *Introdução ao Processamento de Imagens de Sensoriamento Remoto*. Brasília: CNPq, Universidade de Brasília - UnB, 2012. 266 p.
- MORAES, E. C. de. *Fundamentos de Sensoriamento Remoto*. São José dos Campos: Instituto Nacional de Pesquisas Espaciais - INPE, 2002. 281 p.
- MOREIRA, M. A. *Fundamentos do sensoriamento remoto e metodologias de aplicação*. São José dos Campos: Instituto Nacional de Pesquisas Espaciais - INPE, 2001. 208 p.
- MOURA, C. D. G. de; FREITAS, N. C. A. de; SILVA, M. P. dos S. Agentes de mineração de imagens de satélite. In: *XV Brazilian Symposium on Geoinformatics, Campos do Jordão, São Paulo, Brazil*. [s.n.], 2014. p. 162-167. Disponível em: <[http://www.geoinfo.info/proceedings\\_geoinfo2014.split/Paper09-S-p14.pdf](http://www.geoinfo.info/proceedings_geoinfo2014.split/Paper09-S-p14.pdf)>. Acesso em: 30.12.2014.
- NETBEANS. *Sobre: Bem-vindo a comunidade netbeans!* 2015. Oracle Corporation e/ou suas afiliadas. Disponível em: <[https://netbeans.org/about/index\\_pt\\_BR.html](https://netbeans.org/about/index_pt_BR.html)>. Acesso em: 20.01.2015.
- PULTAR, E.; RAUBAL, M.; GOODCHILD, M. F. Gedmwa: geospatial exploratory data mining web agent. In: *GIS*. [S.l.]: ACM, 2008. ISBN 978-1-60558-323-5.

RAMOS, C.; LOBO, F. Descoberta de conhecimento em bases de dados. *Dos Algarves - A Multidisciplinary E-Journal Revista da ESGHT-UALG*, Campus da Penha: UALG, n. 12, p. 37-54, 2004. ISSN 0873-7347.

REINSEL, D.; GANTZ, J.; RYDNING, J.; *Data Age 2025: The Evolution of Data to Life-Critical Don't Focus on Big Data; Focus on the Data That's Big*. 2017. IDC White Paper. Disponível em: <<http://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>>. Acesso em: 04.07.2017.

RUSSELL, S.; NORVING, P. *Artificial Intelligence: A Modern Approach*. 3. ed. [S.l.]: Prentice Hall, 2010. ISBN 0136042597.

SAGAN, C. *The Dragons of Eden: Speculations on the Evolution of Human Intelligence*. [S.l.]: Random House, 1977. 263 p.

SAUSEN, T. *Sensoriamento Remoto - Tópicos em meio ambiente e ciências atmosféricas*. São José dos Campos: Instituto Nacional de Pesquisas Espaciais - INPE, 2005.

SEMACE. *Shape - Definições e Conversão*. 2010. Superintendência Estadual do Meio Ambiente (SEMACE). Disponível em: <<http://www.semace.ce.gov.br/2011/06/shape-definicoes-e-conversao/>>. Acesso em: 08.01.2015.

SEYDIM, A. Y. *Intelligent Agents: A Data Mining Perspective*. [S.l.], 1999.

SILVA, C. V. S. *Agentes de Mineração e sua Aplicação no Domínio de Auditoria Governamental*. Dissertação (Mestrado) - Universidade de Brasília - UnB, Brasília, 2011.

SILVA, M. P. dos S. *Mineração de dados em bancos de imagens*. 123 p. Tese (Doutorado) - Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos, 2006.

SILVA, M. P. dos S.; CÂMARA, G.; ESCADA, M. I. S. Image mining: Detecting deforestation patterns through satellites. In: *Data Mining Applications for Empowering Knowledge Societies*. [S.l.]: IGI Global, 2008. p. 55-75.

SILVA, M. P. dos S. et al. Mining patterns of change in remote sensing image databases. In: *Proceedings of the Fifth IEEE International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society, 2005.

SYCARA, K. P. Multiagent systems. *AI Magazine*, v. 19, n. 2, p. 72-92, 1998.

USGS. *EarthExplorer*. 2014. U.S. Geological Survey (USGS). Disponível em: <<http://earthexplorer.usgs.gov/>>. Acesso em: 20.08.2014.

USGS. *Frequently Asked Questions about the Landsat Missions: Question: What are the band designations for the landsat satellites?* 2014. U.S. Geological Survey (USGS). Disponível em: <[http://landsat.usgs.gov/band\\_designations\\_landsat\\_satellites.php](http://landsat.usgs.gov/band_designations_landsat_satellites.php)>. Acesso em: 17.01.2015.

USGS. *Frequently Asked Questions about the Landsat Missions: Question: What are the best spectral bands to use for my study?* 2014. U.S. Geological Survey (USGS). Disponível em: <[http://landsat.usgs.gov/best\\_spectral\\_bands\\_to\\_use.php](http://landsat.usgs.gov/best_spectral_bands_to_use.php)>. Acesso em: 17.01.2015.

USGS. *Landsat 5 History*. 2014. U.S. Geological Survey (USGS). Disponível em: <[http://landsat.usgs.gov/about\\_landsat5.php](http://landsat.usgs.gov/about_landsat5.php)>. Acesso em: 20.08.2014.

USGS. *Landsat 8*. 2014. U.S. Geological Survey (USGS). Disponível em: <<http://landsat.usgs.gov/landsat8.php>>. Acesso em: 12.01.2015.

WOOLDRIDGE, M. *An Introduction to MultiAgent Systems*. 2. ed. Chichester, England: John Wiley and Sons, LTD, 2009.

XAVIER JÚNIOR, J. C. *NatalGIS: Um Sistema Multiagente de Recomendação de Informações Geográficas baseado em Agrupamento de Dados Relacionais*. 189 p. Tese (Doutorado) - Universidade Federal do Rio Grande do Norte - UFRN, Natal - RN, 2012.

ZHANG, J.; LEE, M. L.; HSU, W. Image mining: Trends and developments. *Journal of Intelligent Information Systems*, v. 19, p. 7-23, 2002.



## **Sobre os autores**

### **Ciro Daniel Gurgel de Moura**

Possui graduação em Ciência da Computação pela Universidade do Estado do Rio Grande do Norte (2012) e mestrado em Ciência da Computação pela Universidade do Estado do Rio Grande do Norte e pela Universidade Federal Rural do Semi-Árido (2015). Áreas de interesse: Banco de Dados, Mineração de Dados, Geoinformática, Processamento de Imagens e Desenvolvimento Web.

Contato: [ciro.dgm@gmail.com](mailto:ciro.dgm@gmail.com)

Currículo Lattes: <http://lattes.cnpq.br/6573088805666433>

### **Marcelino Pereira dos Santos Silva**

Possui graduação em Ciência da Computação pela Universidade Federal de Campina Grande (1992), mestrado em Ciência da Computação pela Universidade Federal de Pernambuco (2002) e doutorado em Computação Aplicada pelo Instituto Nacional de Pesquisas Espaciais (2006). Atualmente é Professor Adjunto da Universidade do Estado do Rio Grande do Norte. Tem experiência na área de computação com ênfase em Banco de Dados, atuando principalmente nas seguintes áreas: Mineração de Dados, Geoinformática, Inteligência Computacional, Processamento de Imagens.

Contato: [prof.marcelino@gmail.com](mailto:prof.marcelino@gmail.com)

Currículo Lattes: <http://lattes.cnpq.br/7817033448036400>

### **Nicksson Ckayo Arrais de Freitas**

Especialista em Desenvolvimento de sistemas com JAVA (2017). Mestrando em Ciência da Computação pela Universidade do Estado do Rio Grande do Norte / Universidade Federal Rural do Semi Árido. Bacharel em Ciência da Computação pela Universidade do Estado do Rio Grande do Norte (2015).

Contato: [nickssonarrais@gmail.com](mailto:nickssonarrais@gmail.com)

Currículo Lattes: <http://lattes.cnpq.br/3965021702433315>

