



**11ª Jornada Científica e
Tecnológica do IFSULDEMINAS**

**& 8º Simpósio de
Pós-Graduação**

ESTUDO COMPARATIVO DE ALGORITMOS DE SELEÇÃO DE ATRIBUTOS EM IMAGEM DE ALTA RESOLUÇÃO

Micaela A. SILVA¹; Daniel R. ARCANJO²; Camila S. dos ANJOS³

RESUMO

A segmentação é o processo em que uma imagem é dividida em segmentos, de com as características espectrais do alvo de interesse. Para a classificação de uma imagem tais segmentos são associados a uma classe espectral. Diante disto, este trabalho teve como objetivo realizar a classificação da imagem de alta resolução WorldView-3 (0,30m) sobre a área de Tripoli - Líbia. A metodologia baseou-se em realizar os tratamentos das imagens e, partindo disto efetuar a segmentação, classificação e validação, seguido da identificação de parâmetros que melhor discriminam as classes levantadas. Os algoritmos de seleção de atributos *GainRatio* e *InfoGain* foram comparados ao classificar a imagem pelo algoritmo *RandomForest*. O algoritmo *InfoGain* apresentou maior acurácia global se mostrando superior na classificação da cena.

Palavras-chave: Seleção de Atributos; Processamento de Imagens; Sensoriamento Remoto.

1. INTRODUÇÃO

O conjunto de elementos presentes na superfície terrestre podem ser observados e analisados em uma imagem de satélite, de acordo a similaridade entre as regiões.

Auxiliando nestes estudos, o processamento digital de imagens, é fonte essencial para extração de atributos por meio de imagens de alta resolução. Isso propicia geração de mapas, visualização de todos os segmentos espalhados pela área urbana como construções, calçamento, vegetação, transportes urbanos, entre outros (JENSEN, 2011). Para isso, são necessárias diversas técnicas de segmentação e classificação das imagens.

Diante disto, o objetivo deste trabalho é avaliar o potencial da classificação gerada por dois algoritmos de seleção de atributos aplicada a uma imagem de alta resolução espacial.

2. FUNDAMENTAÇÃO TEÓRICA

Entre os meios de armazenamento de dados existentes como banco de dados, torna-se necessário efetuar a mineração de informações, onde estas serão de grande importância para o ganho de conhecimento sobre o local que está sendo estudado. Após este processo, são aplicadas técnicas de mineração de dados (*Data Mining*) que irão identificar padrões que se destacam e possibilitam novos caminhos para a tomada de decisão em uma base de dados (CAMILO; SILVA, 2009).

1 Discente, IFSULDEMINAS – *Campus* Inconfidentes. E-mail: micaela.allmeida@gmail.com

2 Discente, IFSULDEMINAS – *Campus* Inconfidentes. E-mail: daniel_arcanjo15@hotmail.com

3 Orientador, IFSULDEMINAS – *Campus* Inconfidentes. E-mail: camila.lacerda@ifsuldeminas.edu.br

Uma das técnicas utilizadas consiste em segmentar a imagem em unidades semelhantes de pixels, transformando a imagem de estrutura matricial em uma imagem de estrutura vetorial. O algoritmo de segmentação também pode ser baseado pelo crescimento de regiões, resultando em um processo iterativo (ESPINDOLA, 2005)

Neste caso, os vetores são objetos que permitem relacionamentos de herança, vizinhança e forma, permitindo introduzir características importantes, utilizando elementos principais como: cor, tonalidade, padrão, textura, forma, sombra, altura, localização para interpretação, entre outros (CRÓSTA, 1999).

A classificação de imagens de satélites é um processo em que alvos de uma mesma cena podem ser diferenciados e identificados, entretanto, existe uma enorme complexidade para extrair e simplificar essas informações de forma que garanta a qualidade e utilidade do conhecimento através dos dados (MARTIN; ODELL, 1995).

Ainda podem ser utilizados outros algoritmos de seleção de atributos, que para este trabalho serão essenciais. Como exemplo, tem-se a taxa de ganho de informação (*GainRation*), sendo caracterizado pelo uso de uma métrica para ranquear todos os atributos de uma base de dados, calculada através do ganho de informações de um atributo contra o número de saídas testes, com a maior possibilidade de valores (NETTO, 2013). E o *InfoGain* utilizado para medir a semelhança entre o atributo de uma classe, com todas as outras classes, diminuindo o grau de incerteza (ANJOS, 2019).

3. MATERIAL E MÉTODOS

Para a realização deste trabalho foi escolhida uma imagem do satélite WorldView-3, disponibilizada gratuitamente pela *Digital Globe*. As cenas retratam a região da cidade de Tripoli - Líbia. A imagem possui 8 bandas, sendo uma banda pancromática com resolução espacial de 0,30 m e as bandas multiespectrais com 1,30 m de resolução espacial. Com o intuito de melhorar a resolução espacial da imagem foi realizado a fusão entre as bandas multiespectrais e a pancromática por meio do *software ENVI* (Figura 1).

Em seguida, foram realizadas transformações multiespectrais para a extração de atributos decorrentes da imagem original (Figura 3, Figura 4 e Figura 5, respectivamente):



Figura 01. Imagem da Fusão de 31/08/2015. Composição Colorida 5R3G2B.



Figura 02. Imagem da Análise das 31/08/2015. Componentes Principais (PCA)



Figura 03. Imagem da Fração Mínima de Ruído (MNF) 31/08/2015.



Figura 04. Imagem do Índice de Vegetação (NDVI) de 31/08/2015.

As imagens originais e as imagens decorrentes das transformações foram inseridas no *software eCognition* para realizar a segmentação utilizando o algoritmo *Multiresolution Segmentation*, com parâmetro de escala igual a 25 e critérios de forma e compacidade igual a 0,9 e 0,8 respectivamente e o algoritmo *spectral difference* com critério 10 para realizar diminuição do número de segmentos.

Depois de realizar a segmentação foi feita duas exportações para o *software Weka*, a primeira das classes das amostras, lembrando-se de inserir a classe *class name* e a segunda exportação de todos os segmentos e gerou-se atributos, sendo eles: *Área; Border Index; Border to PPO, Brightness; Compactness; Max. Diff.; Mean Layer; Mode*. Já no *software Weka*, realizou-se a seleção de atributos com os algoritmos *GainRatio* e *InfoGain* e a classificação por meio do algoritmo *RandomForest* para cada um dos métodos de seleção de atributos (Figura 7 e 8) para posteriores comparações.

4. RESULTADOS E DISCUSSÕES

Os algoritmos analisados classificaram cada segmento que foram renomeados na tabela de atributos e exportados para o *software QGis*, onde foi possível visualizar a cena gerada por intermédio da classificação.

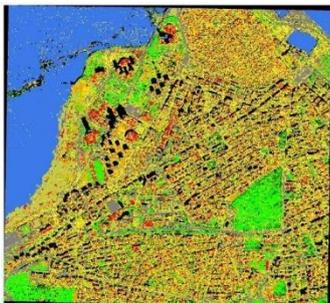


Figura 05. Classificação Gerada a partir da Seleção de Atributos com Algoritmo *Info Gain*. Maior contraste e detalhes.

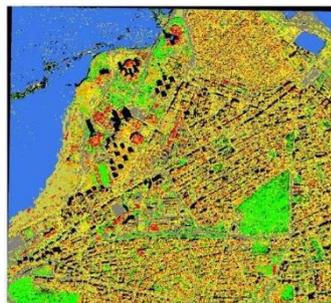


Figura 06. Classificação Gerada a partir da Seleção de Atributos com Algoritmo *Gain Ration*.

Comparando-se os métodos utilizados sabe-se que a proporção do ganho de informação de cada atributo modifica-se com relação a classe, quanto maior este ganho menor será a incerteza. Em uma árvore de decisão, o atributo com maior ganho é ranqueado pelo valor da entropia (incerteza) e quanto maior este valor, melhor será a descrição dos dados (HAN; KAMBER; JIAN, 2011).

Por não possuir a verdade de campo a validação foi gerada pelo método *Croos-Validation*, utilizando o conjunto amostral. Assim, gerou-se a classificação com parte das amostras e foram guardadas outras amostras para utilizá-las como teste.

O algoritmo *InfoGain* apresentou por meio da classificação *RandomForest*, um coeficiente

kappa no valor de 0,91. Já a seleção de atributos através do algoritmo *GainRatio* por meio da classificação *RandomForest*, obteve um coeficiente *kappa* no valor de 0,80.

5. CONCLUSÕES

A metodologia apresentada, proporcionou a identificação das características de interesse por meio da segmentação e classificação. O que tornou possível a análise o comportamento dos atributos de acordo a classificação. O melhor algoritmo de seleção de atributos foi o *InfoGain*, que gerou uma classificação com maior acurácia e coeficiente *kappa* de 0,91, sendo classificado como excelente. Já para o algoritmo *GainRation*, o coeficiente *kappa* foi de 0,80.

Diante dos resultados encontrados neste trabalho demonstrou-se a viabilidade de se utilizar imagens digitais para identificar materiais presentes no ambiente urbano, permitindo extrair informações úteis e relevantes para avaliar a influência dos atributos na classificação de imagens.

REFERÊNCIAS

- ANJOS, C.S. **Extração e Seleção de Atributos**. Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais, Inconfidentes, 2019.
- CRÓSTA, Á. P. **Processamento Digital de Imagens de Sensoriamento Remoto**. Campinas: IG/UNICAMP. 170 p. 1999.
- ESPINDOLA, G. M. **Emprego da estatística espacial na avaliação da Segmentação de imagens de sensoriamento remoto**. 30F. Trabalho de curso (Mestrado em sensoriamento Remoto) - INPE, São José dos Campos, 2005.
- HAN, J.; KAMBER, M.; JIAN, P. **Data Mining: Concepts and Techniques**. Morgan Kaufmann Publishers, San Francisco, CA, USA, 2011.
- JENSEN, J.R. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. São José dos Campos. 2011.
- CAMILO, C. O; SILVA, J. C. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas**. Instituto de Informática, Universidade Federal de Goiás. 2009.
- MARTIN, J; ODELL, J. **Análise e Projeto Orientados a Objeto**. São Paulo: ed. Makron Books, p.624, 1995.
- NETTO, O. P. **Um Filtro Interativo Utilizando Árvores de Decisão**. Universidade de São Paulo, 2013.