

Método Estatístico de Seleção de Canais Aplicado a Classificação de Imagens

Joanito de Andrade Oliveira ¹

Luciano Vieira Dutra ²

Camilo Daleles Rennó ²

Dionísio Costa Cruz Junior ³

Universidade Federal do Recôncavo da Bahia – UFRB ¹
Campus Universitário da UFRB – 44380-000 - Cruz das Almas – BA, Brasil
joanito@ufrb.edu.br

Instituto Nacional de Pesquisas Espaciais – INPE ²
Caixa Postal 515 - 12201-970 - São José dos Campos - SP, Brasil
{camilo,dutra}@dpi.inpe.br

Escola de Engenharia de Agrimensura – EEA ³
Rua Teixeira de Barros, 40.050-001 - Salvador – Ba, Brasil
dionisiojunior@hotmail.com

Abstract. With the steady increase in the number of features available from remote sensing sources, there is a growing necessity to reduce the complexity of the classification task. When data dimensionality is very high, a search strategy should be used to select the subset of features that gives the minimum classification error, considering the limited size of training data. Particularly when one deals with the very complex environment that is the urban scene, shape feature extraction is necessary to distinguish different classes of objects which have similar spectral signature. The objective of the work is to extract feature of the regions of a Landsat5 /TM image and to reduce the dimensionality, of form that does not occur loss in the efficiency in the classification process. As there is no a deterministic relation between feature selection methods and classification error, it is possible to conclude that all search strategies should be used to narrow the number of choices assessments based on classification error. The system was developed in IDL/ENVI. The results present the viability to use the methods of selection of attributes, objectifying the reduction of the dimensionality without losing the discriminatory power between the class. Further studies are being done aiming to implement evaluation methods for classification.

Palavras Chaves: feature selection, image processing, multispectra image, seleção de atributos, processamento de imagens, imagens multiespectrais.

1. Introdução

O foco principal do desenvolvimento de sensores espaciais é a geração de imagens que possam auxiliar na classificação do uso da terra, no planejamento e monitoramento do ambiente rural e urbano. A geração de mapas cresceu paralelamente com a disponibilidade das imagens de satélites e com o desenvolvimento e aprimoramento das técnicas de manipulação de imagens, constituindo um ponto de partida para muitas aplicações em diferentes domínios de planejamento do espaço (Donnay, 1999).

Em decorrência do aprimoramento da resolução espacial, temporal e espectral dos sensores espaciais, está sendo possível monitorar diferentes fenômenos com um nível maior de detalhes. A alta resolução espectral das imagens implica num grande potencial para a identificação de diferentes tipos de superfícies, baseado nas características de suas respostas espectrais específicas. Atualmente, os dados provenientes dos sensores multiespectrais e hiperespectrais são utilizados para extrair informações que auxiliem no processo de separabilidade entre as classes. Pode-se pensar que, quanto maior o número de atributos (canais) para representar um padrão (classe), maior o poder discriminatório do classificador. Porém, nem sempre isso é verdade. Alguns sistemas podem sofrer a maldição da dimensionalidade (degradação na acurácia dos resultados da classificação com o aumento da dimensionalidade dos dados, mantendo constante o número de amostras de treinamento).

Segundo Jain et al. (2000), existem duas razões para reduzir a dimensionalidade: diminuir o custo de processamento e aumentar a acurácia da classificação. Outro método utilizado para reduzir a dimensionalidade dos dados é a transformação por componente principal. A Transformação por Principais Componentes (TPC) é uma aplicação de uma transformação de bandas que reuni em um conjunto de novas bandas o que representa a maioria dos dados contidos nas bandas originais. Ou seja, os valores das bandas originais são modificados a fim de estabelecer uma melhor separabilidade entre as classes.

Este trabalho tem por objetivo aplicar métodos estatísticos para selecionar os melhores canais, provenientes de sensores multiespectrais. O algoritmo de seleção de canais é aplicado para reduzir a dimensionalidade sem que ocorra perda na eficiência do poder de distinção das classes pelo classificador.

2. Seleção de Canais

Um número alto de canais não implica em uma boa classificação, ao contrário, pode existir uma alta correlação entre os canais, fazendo com que o custo de processamento aumente, gerando problemas matemáticos (inversão de matrizes), sendo necessário escolher os canais mais representativos, ou seja, que melhor discriminam os objetos de interesse.

Métodos automáticos de seleção de canais são importantes em muitas situações em que se tem disponível um conjunto grande de atributos e deseja-se selecionar um subconjunto adequado para solucionar o problema. A seleção de canais é uma técnica de otimização que, dado um conjunto de d atributos, tenta selecionar um subconjunto de tamanho requerido que maximiza uma função critério.

A maioria dos métodos de seleção de atributos utiliza princípios estatísticos, ou seja, medidas estatísticas dos atributos entre pares de classes, selecionando os atributos que contém informações relevantes para reproduzir o problema. Há várias formas de medir a distância entre conjuntos de classes diferentes no espaço de atributos (Theodoridis e Koutroumbas, 1999; Kohn, 1998). Dentre elas, pode-se citar: distância Euclidiana, distância de Mahalanobis e a distância de Bhattacharyya (Tabela 1).

Tabela 1: Exemplos de distâncias entre conjuntos de classes

Distância Euclidiana: distância entre dois pontos. No espaço de atributos d -dimensional a distância euclidiana entre um ponto x qualquer e o centróide (média) da classe k .	Distância de Mahalanobis: distância entre um ponto e o centróide de uma distribuição. A distância de Mahalanobis entre um ponto x e a classe k .	Distância de Bhattacharyya: medida da similaridade entre duas distribuições estatísticas (Bhattacharyya, 1943). Na seleção de atributos, a distância de Bhattacharyya entre a distribuição x e da classe k .
$E_{xk} = \sqrt{\sum_{i=1}^d (x_i - \mu_{ik})^2}$	$M_{xk} = (\mu_k - x)^t [\sum_k]^{-1} (\mu_k - x)$	$B_{ik} = \frac{1}{8} (\mu_k - \mu_i)^t \left[\frac{\sum_k + \sum_i}{2} \right]^{-1} (\mu_k - \mu_i) + \frac{1}{2} \ln \frac{\frac{ \sum_k + \sum_i }{2}}{\sqrt{ \sum_i \sum_k }}$

Para executar a seleção de atributos com base em distâncias entre classes, é necessário definir uma função critério que possa avaliar a separabilidade entre todas as classes de uma maneira global. Ao utilizar uma função critério em métodos de seleção de atributos deve-se maximizar a distância entre padrões de classes diferentes no espaço de atributos, visando otimizar o conjunto de atributos para minimizar a probabilidade de erro na classificação.

A distância Jeffreys-Matusita (distância JM) é uma função critério mais utilizada, que expressa as medidas estatísticas de separabilidade entre duas distribuições (Richards, 1993). A distância JM entre as classes x e k é dada por:

$$JM_{xk} = \sqrt{2(1 - e^{-B_{xk}})} \quad , \quad JM \in [0, \sqrt{2}) \quad (1)$$

onde B_{ki} representa a distância de Bhattacharyya entre as classes k e i .

Segundo Boz (2002), os algoritmos de seleção de atributos são divididos em três grupos: exponenciais (busca exaustiva), randômicos (algoritmo genético) e seqüenciais. De acordo com Jain e Zongker (1997), tais abordagens podem ser agrupadas nas categorias de reconhecimento estatístico de padrões (REP) e em redes neurais.

2.1 Seleção de Atributos em Redes Neurais Artificiais

O método *Node Pruning* é o método mais utilizado na literatura em aplicação de redes neurais (Mao et al., 1994). Também chamado de “corte de nós”, funciona através de uma rede neural multi-camadas com retro-alimentação, utilizando um algoritmo de aprendizado baseado em retro-propagação (*backpropagation*). Segundo Campos (2001), esse método é definido por uma medida de “significância de nós” e utiliza um algoritmo que elimina os nós menos significativos. Dessa forma, reduz-se o conjunto de treinamento da rede, diminuindo a complexidade.

2.2 Seleção de Atributos em Reconhecimento Estatístico de Padrões

O método de seleção de atributo utilizado para extrair um subconjunto ótimo do conjunto global de atributos é a busca exaustiva. O número de possibilidades desse método cresce exponencialmente, fazendo a busca exaustiva pouco utilizada para um número alto de atributos. O que se faz é extrair um subconjunto ótimo do conjunto global de n atributos para um número fixo de k atributos. A esses métodos dá-se o nome de método de seleção de atributos exaustivo de escolha fixa.

Segundo Narendra e Fukunaga (1994), o algoritmo *branch-and-bound* de seleção de atributos extrai um subconjunto ótimo sem efetuar a busca exaustiva. A principal característica do algoritmo é a propriedade monotônica da função critério, cujo valor aumenta a medida que se adiciona um atributo no conjunto global de atributos.

Os métodos que extrai subconjuntos sub-ótimos utilizam duas abordagens: determinística e estocástica. As duas abordagens proporcionam a seleção de uma única solução (único caminho) ou de várias soluções (múltiplos caminhos) para o problema de seleção de atributos. Os métodos seqüenciais são os mais utilizados para selecionar um único subconjunto. Dos métodos que fornecem múltiplos caminhos para a solução do problema, os algoritmos genéticos destacam-se por utilizar um conjunto de soluções a cada iteração.

2.2.1 Algoritmos Seqüenciais de Seleção de Atributos

Os métodos seqüenciais de seleção de atributos têm sido estudados em vários trabalhos (Bittencourt et al., 2004; Jain e Zongker, 1997). Segundo Dutra e Huber (1999), os dois métodos seqüenciais mais proeminentes da seleção de atributos são: *Sequential Forward Feature Selection* (SFS) e *Sequential Backward Feature Selection* (SBS).

O algoritmo de seleção seqüencial para frente (SFS) vem sendo utilizado na tentativa de resolver a tarefa de seleção de atributos (Dutra e Huber, 1999; Jain et al., 2000; Campos, 2001; Oliveira, 2005). O método SFS é fácil de implementar, possui baixo custo computacional e muito utilizado na solução de problemas de otimização e busca. Aplicado em sistema, onde pretende obter um subconjunto de atributos com eficiência e de forma rápida, o método SFS apresenta uma aceitável solução para o que se propõe.

Nos sistemas de processamento de imagens, o método SFS é utilizado principalmente na seleção dos canais que melhor discriminam as classes. Outra utilização é na escolha dos melhores atributos que representam uma classe (variância, média, entre outros), e que serão utilizados no processo de classificação.

O algoritmo SFS inicia com um conjunto vazio ($Y_0 = \phi$) e, conforme o algoritmo é executado, o melhor atributo x^+ é inserido no subconjunto, resultante do máximo valor da função critério $J(Y_K + x^+)$, quando combinado com os atributos Y_K que já tenha sido selecionado. Esse método tem menor custo computacional quando se deseja obter conjuntos pequenos em relação ao total de atributos (Jain e Zongker, 1997). Uma vez que um atributo tenha sido selecionado, ele não pode ser descartado do subconjunto, o que pode provocar o chamado efeito *nesting*.

O algoritmo SBS, assim com método SFS vem sendo utilizado em trabalho direcionando ao estudo de seleção de atributos em sistemas de processamento de imagens (Dutra e Huber, 1999; Jain et al., 2000; Campos, 2001; Oliveira, 2005). Esse algoritmo inicia com um conjunto de atributos completo ($Y_0 = X$) e, nas iterações do algoritmo, remove-se o atributo com o mínimo valor da função critério $J(Y_K - x^-)$, onde Y_K é o conjunto de atributos já selecionado e x^- é o atributo a ser removido (Figura 2). O método SBS tem menor custo computacional, quando se deseja obter conjuntos grandes em relação ao total de atributos (Oliveira, 2005).

Uma desvantagem desse método é que a primeira interação é executada com o número global de atributos. Se existir uma alta correlação entre os atributos em estudo, o método pode eliminar o atributo que representa melhor a separabilidades entre as classes. Isso por que, a discriminação das classes por esse atributo, pode não representar uma boa separabilidades entre as classes quando se uni ao um subconjunto maior. Uma vez eliminado o atributo, ele não retornará ao subconjunto de atributos selecionado para a classificação.

3. Materiais e Métodos.

Para a execução do trabalho foi desenvolvido um sistema de classificação de imagens na linguagem de programação IDL (*Interactive Data Language*), utilizando algumas funções do ENVI. O classificador desenvolvido utiliza o método da Máxima Verossimilhança.

Com objetivo de apresentar a potencialidade dos métodos de seleção proposto para a classificação em imagens de sensoriamento remoto foi utilizada uma imagem Landsat 5/TM, com as 7 bandas, do Estado do de Mato Grosso (286031 E a 301031 E; 8695883 N a 8680883N; fuso 22S). A imagem Landsat/TM utilizada refere-se à órbita-ponto 225/068, de 11 de julho de 2008. Executou-se um recorte na imagem original para a região de estudo.

A interface inicial do sistema oferece funções de inserção dos canais a serem utilizados na etapa de extração das regiões de interesse (ROI) e no processo de seleção (Figura 1).

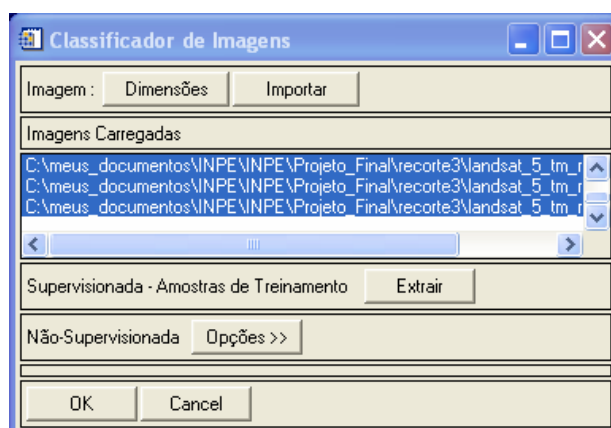


Figura 1. Interface Inicial do Sistema.

O próximo passo é definir as amostras de treinamento representativas para cada classe (Figura 2). Para a classificação usaram-se três classes (mata, desmatamento e rio).

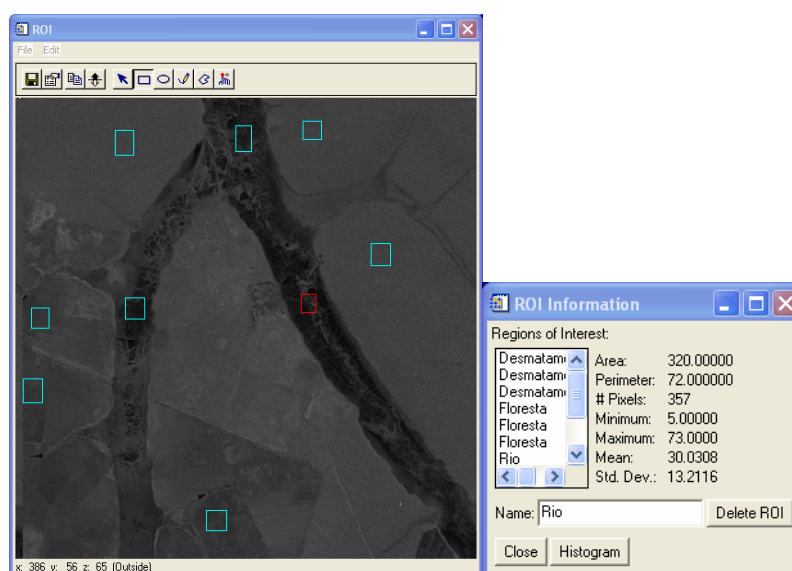


Figura 2. Interface para Extração das Amostras de Treinamento.

A partir das regiões de treinamento, fez-se a seleção dos canais pelo método SFS. Executou-se uma modificação no algoritmo SFS para selecionar apenas os pares de canais que oferecem a melhor separabilidade entre as classes. O método SFS é utilizado para a seleção de um subconjunto de canais, que pode ser formado por um canal ou N-1 canais, onde N é número total de canais. A escolha pela seleção de pares de canais se deu pelo fato de reduzir a dimensionalidade a dois planos.

Como utilizou os 7 canais da imagem Landsat 5, gerou-se as combinações dois a dois desses canais, totalizando 21 combinações. A partir de cada par calculou-se a distância *JM*, atribuindo a melhor combinação para o par que obteve a maior média *JM*. A Figura 3 apresenta os melhores pares de canais que oferecem a melhor separabilidade entre as classes.

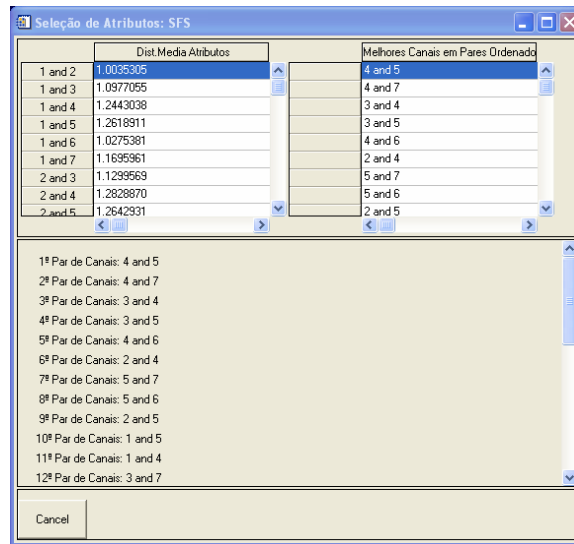
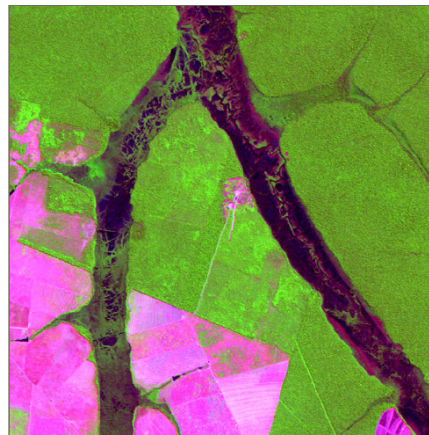
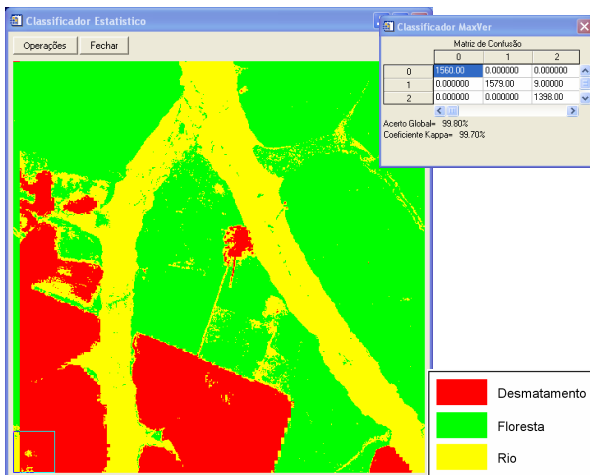


Figura 3. Interface de Seleção de Canais.

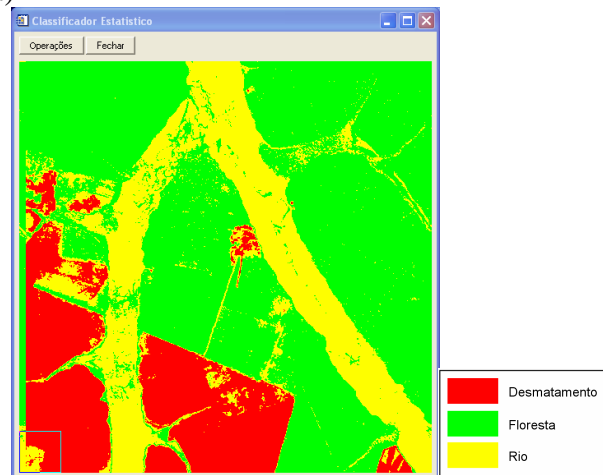
Com objetivo de validar o processo de seleção de canais, calculou-se o coeficiente Kappa para a classificação com o melhor par selecionado (canal 4 e 5) e para a classificação utilizando os 7 canais. O coeficiente kappa para a classificação com os 7 canais foi de 99,7% e para a classificação utilizando o canal 4 e 5 foi de 98,3%. A Figura 4 apresenta a imagem original e as imagens classificadas utilizando todos os canais e o par de canais selecionado.



(a)



(b)



(c)

Figura 4. (a) Imagem Original; (b) Classificação com 7 canais; (c) Classificação com o melhor par de canais (canal 4 e 5).

4. Conclusões e Recomendações

Esse trabalho apresentou métodos de seleção de canais (atributos) para a classificação de imagens. Através desse método, foi possível selecionar os pares de canais mais relevantes na separabilidade das classes. No processo de classificação, uma quantidade alta de canais pode não aumentar significativamente a precisão do classificador. Isso ocorre devido à redundância das informações.

Os resultados encontrados mostraram a viabilidade em se utilizar o método de seleção proposto, quando o objetivo é a redução da dimensionalidade sem que haja perda significativa no poder discriminatório entre as classes. Como não existe nenhuma relação determinística entre métodos de seleção de atributos padrão e erro de classificação, conclui-se que todas as estratégias de busca devem ser usadas para diminuir o número de avaliações na escolha da decisão final baseada no erro da classificação.

As interfaces gráficas implementadas no sistema, que acompanha as etapas do sistema de processamento de imagens, são interfaces intuitivas e fáceis de usar. No entanto, é de fundamental importância compreender as etapas do sistema de processamento de imagens para se obter uma melhor compreensão da seqüência das interfaces.

Uma das contribuições deste trabalho foi a pesquisa da influência do processo de seleção de canais nos sistemas de processamento de imagens, fundamental na obtenção de uma boa classificação. Pode-se, então, concluir que os métodos de seleção apresentaram os subconjuntos de pares de canais que melhor discrimina as classes de interesse, mas não garante que com esses subconjuntos o sistema obterá uma melhor classificação. Isso porque uma classificação que apresente uma boa acurácia, não depende apenas dos métodos de seleção, mas também do classificador utilizado.

Para complementar trabalho nesta área de processos de seleção de canais, existem outras pesquisas que poderiam dar continuidade e auxiliar futuros estudos:

- Implementar outros métodos de seleção de atributos, com o objetivo de selecionar os subconjuntos mais relevantes na separabilidade das classes, aplicando métodos estocásticos (determinados pelas leis da probabilidade; aleatórios) ou métodos determinísticos que evitem o efeito *nesting*.
- Aplicar os métodos de seleção de canais em imagens de sensores hiperespectrais, como Hyperion.
- Desenvolver métricas de avaliação e comparação com outros algoritmos de seleção de atributos e aplicar em outras áreas de estudos com dados oriundos de vários sensores.
- Agregar ao sistema medidas de avaliação da classificação através de análise estatística.

Espera-se que este trabalho possa ter contribuído para mostrar a importância da utilização de métodos de seleção de canais em sistemas de processamento de imagens.

5. Referências

Bhattacharyya, A.. On a measure of divergence between two statistical populations defined by their probability distributions. **Bulletin of the Calcutta Mathematics Society** 35, 99–110, 1943.

Bittencourt H.R., Clarke R.T., Feature Selection by Using Classification and Regression Trees (cart). **XX ISPRS Congress**, 12-23 July 2004, Istanbul, Turkey.

Boz, O. Feature Subset Selection by Using Sorted Feature Relevance. Proc. of The 2002 Intl. **Conf. on Machine Learning and Applications**, 2002.

Campos T.E. **Técnicas de Seleção de Características com Aplicações em Reconhecimento de Faces**. Dissertação Apresentada ao Instituto de Matemática e Estatística da Universidade de São Paulo. São Paulo, 2001.

Donnay, J.P. Use of remote sensing information in planning. In: Geographical Information and Planning, edited by J. Stillwell, S. Geertman & S. Openshaw, Berlin, Springer-Verlag, 1999, p. 242- 260.

Dutra, L.V. and Huber, R.. Feature Extraction and Selection for ERS 1/2 InSAR Classification: **Int. J. Remote Sensing**, 1999, vol. 20, n°. 5, 993-1016.

Jain. A. K., Robert P.W. Moa D., Moa. J. Statistical Pattern Recognition: A Review. **IEEE Trans. On Pattern Analysis And Machine Intelligence**, Vol. 22, N°.1, 2000.

Jain, A. K. and Zongker, D. Feature-selection: Evaluation, application, and small sample performance. **IEEE Trans. on Pattern Analysis and Machine Intelligence**, 19(2):153-158, 1997.

Kohn, A. F. Reconhecimento de padrões, uma abordagem estatística. PEE- Universidade de São Paulo, São Paulo, 1998. 205p.

Mao, J.; Mohiuddin , K.;A. Jain, K.; Parsimonious network design and feature selection through node pruning. **Proceedings**. 12th ICRP, p. 622-624, Jerusalem, 1994.

Narendra and K. Fukunaga. **A branch and bound algorithm for feature subset selection**. IEEE Transactions on Computers, 26(9):917-922, 1977.

Oliveira, J. A. **Classificação de regiões usando atributos de forma e seleção de atributos**. 2005. 104 p. (INPE-13746-TDI/1050). Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2005.

Richards, J. A. **Remote Sensing Digital Image Analysis: An Introduction**. New York, Inc: Springer-Verlag, v. 2 ,1993. 340 p.

Theodoridis, S. and Koutroumbas, K. **Pattern Recognition**. Academic Press, USA, 1st edition, 1999.