

CLAUDIONOR RIBEIRO DA SILVA

**USO DE ALGORITMOS GENÉTICOS COMO REDUTOR DE DIMENSIONALIDADE
NA CLASSIFICAÇÃO DE IMAGENS HIPERESPECTRAIS**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciências Geodésicas, Curso de Pós-Graduação em Ciências Geodésicas, Departamento de Geomática, Setor de Ciências da Terra, Universidade Federal do Paraná.

Orientador:

Prof. Dr. Jorge A. Silva Centeno

Co-Orientadora:

Profa. Dra. Selma R. A. Ribeiro

CURITIBA

2006

Quaisquer que sejam os dons que lhe tenham sido
concedidos eles podem ser ampliados e desenvolvidos
pela prática.

Rick Warren

*Dedico esse trabalho aos meus
insubstituíveis pais:*

Alexandre (in memoriam) e Juraci

AGRADECIMENTOS

À Deus a oportunidade de viver;

Ao Prof. Dr. Jorge A. Silva Centeno, a amizade, a orientação, a sua disponibilidade, ensinamentos e apoio fundamentais durante todo o processo de criação e elaboração desta dissertação. E especialmente a concessão e o crédito, nessa proposta desde o início.

À Prof^a. Dr^a. Selma R. A. Ribeiro, a amizade, a co-orientação, a sua disponibilidade, contribuição e paciência durante todo o processo de elaboração desta pesquisa.

À Mônica, a atenção e o carinho especial de uma literalmente grande secretária.

À todos os professores e colegas de curso, o ensino, as discussões, trocas de informações, e principalmente, a contribuição de todos pelo bom andamento do curso CPGCG (Curso de Pós-Graduação em Ciências Geodésicas).

À Universidade Federal do Paraná, em especial ao Departamento de Geomática, a autorização de uso das infra-estruturas: laboratórios, salas e equipamentos.

Às funcionárias que cuidam da limpeza.

À CAPES, o fomento à pesquisa e subsídio financeiro na forma de bolsas de estudo.

À toda minha família, o carinho e apoio indispensáveis ao espírito de paz, e em especial à minha mãe, o amor perene e inconfundível de mãe, fonte profusa de ânimo.

SUMÁRIO

LISTA DE TABELAS	vi
LISTA DE FIGURAS	vi
RESUMO	viii
ABSTRACT	ix
CAPITULO I - INTRODUÇÃO, REVISÃO BIBLIOGRÁFICA E MATERIAL	1
1. INTRODUÇÃO	1
1.1 OBJETIVOS.....	2
1.2 JUSTIFICATIVA.....	3
1.3 ESTRUTURA DA DISSERTAÇÃO.....	3
2. REVISÃO BIBLIOGRÁFICA	5
2.1 IMAGENS HIPERESPECTRAIS: APLICAÇÕES.....	5
2.2 CLASSIFICAÇÃO DE IMAGEM DIGITAL.....	7
2.3 REDUÇÃO DE DIMENSIONALIDADE.....	12
2.3.1 Metodos de Redução de Dimensionalidade.....	15
2.4 MEDIDAS DE SEPARABILIDADE.....	23
2.5 CRITÉRIO PARA CÁLCULO DA DIMENSÃO MÍNIMA.....	24
2.6 ALGORITMOS GENÉTICOS.....	26
2.6.1 Teoria da Evolução das Espécies.....	27
2.6.2 Seleção Natural e Seleção Artificial.....	27
2.6.3 Composição dos Algoritmos Genéticos.....	28
3. MATERIAL	30
CAPITULO II - METODOLOGIA PROPOSTA	32
4. MÉTODOS	32
4.1 ADAPTAÇÃO DOS ALGORITMOS GENÉTICOS.....	32
4.2 CODIFICAÇÃO DOS INDIVÍDUOS E DEFINIÇÃO DA POPULAÇÃO.....	32
4.3 TÉCNICAS DE SELEÇÃO UTILIZADAS.....	33
4.4 OPERADORES GENÉTICOS UTILIZADOS.....	35
4.5 PARÂMETROS UTILIZADOS NA FUNÇÃO DE APTIDÃO.....	36
4.6 CÁLCULO DA DIMENSÃO MÍNIMA.....	36

4.7	PROPOSTA DA FUNÇÃO DE APTIDÃO.....	37
4.7.1	Número de Bandas Presentes no Indivíduo.....	37
4.7.2	Separabilidade Entre as Classes.....	38
4.7.3	Correlação Entre as Bandas.....	38
	CAPITULO III – EXPERIMENTOS REALIZADOS.....	40
	5. EXPERIMENTOS.....	40
5.1	PRIMEIRO EXPERIMENTO: DEFINIÇÃO DOS PARÂMETROS QUE SERÃO USADOS NOS ALGORITMOS GENÉTICOS.....	41
5.2	SEGUNDO EXPERIMENTO: ANÁLISE DO COMPORTAMENTO DOS PARÂMETROS DEFINIDOS PARA A FUNÇÃO DE APTIDÃO.....	52
5.3	TERCEIRO EXPERIMENTO: VERIFICAÇÃO DA ACURÁCIA DO PRODUTO GERADO USANDO OS ALGORITMOS GENÉTICOS.....	56
5.4	QUARTO EXPERIMENTO: AVERIGUAÇÃO DO DESEMPENHO DOS ALGORITMOS GENÉTICOS EM RELAÇÃO AOS ALGORITMOS: SELEÇÃO SEQUENCIAL CRESCENTE E SELEÇÃO SEQUENCIAL DECRESCENTE.....	63
5.5	QUINTO EXPERIMENTO: ANÁLISE DO COMPORTAMENTO DOS ALGORITMOS GENÉTICOS FRENTE AO ALGORITMO <i>FEATURE EXTRACTION</i> : ANÁLISE POR COMPONENTES PRINCIPAIS.....	66
	6. CONCLUSÃO.....	69
	7. RECOMENDAÇÕES.....	72
	REFERÊNCIAS BIBLIOGRÁFICAS.....	73

LISTA DE TABELAS

TABELA 1 - Distribuição aproximada das bandas espectrais do sensor AVIRIS..	30
TABELA 2 - Tamanho das amostras de treinamento e verificação em <i>pixels</i>	31
TABELA 3 - Indivíduos com 30 bandas.....	42
TABELA 4 - Valores máximos de aptidão (indivíduo com 30 bandas).....	46
TABELA 5 - Tempo de processamento em segundos (1ª parte).....	47
TABELA 6 - Tempo de processamento em segundos (conjunto com ruídos).....	48
TABELA 7 – Valores máximos de aptidão (indivíduos com 50 bandas).....	48
TABELA 8 - Bandas selecionadas pelos Algoritmos Genéticos.....	50
TABELA 9 - Valores de acurácia dos mapas temáticos.....	57
TABELA 10 - Avaliação da acurácia dos mapas temáticos (2000 iterações).....	59
TABELA 11 - Valores do índice <i>kappa</i> da segunda etapa (4000 iterações).....	60
TABELA 12 - Módulo da diferença entre os índices <i>kappa</i> das duas etapas.....	60
TABELA 12 - Distribuição no espectro eletromagnético das bandas escolhidas..	61
TABELA 14 - Tempo de processamento demandado.....	63
TABELA 15 - Índices <i>kappa</i> referentes aos resultados dos três algoritmos.....	64
TABELA 16 - Valores do índice <i>kappa</i> referentes aos Componentes Principais..	67

LISTA DE FIGURAS

FIGURA 1 - Representação de uma imagem hiperespectral.....	05
FIGURA 2 - Efeito do fenômeno <i>Hughes</i>	12
FIGURA 3 - Representação do eixo dos componentes principais.....	17
FIGURA 4 - Exemplo hipotético dos eixos dos Componentes Canônicas.....	18
FIGURA 5 - Mapeamento de componente não-linear usando Redes Neurais.....	20
FIGURA 6 - Algoritmo Seleção Seqüencial Crescente (SFS).....	21
FIGURA 7 - Algoritmo Seleção Seqüencial Decrescente (SBS).....	22
FIGURA 8 - Representação de um cone e três possíveis estratos.....	24
FIGURA 9 - Funcionamento da técnica de seleção por <i>roleta</i>	28
FIGURA 10 - Escopo dos Algoritmos Genéticos.....	29
FIGURA 11 - Composição colorida falsa-cor (R=30, G=16, B=9).....	31
FIGURA 12 - Exemplo hipotético de indivíduos ou cromossomos.....	33
FIGURA 13 - Operadores de cruzamento ou <i>crossover</i> (a) e <i>mutação</i> (b).....	35

FIGURA 14 - Rotina para cálculo do número mínimo de bandas a reter.....	37
FIGURA 15 - Etapas seguidas na realização da metodologia.....	41
FIGURA 16 - Esquema da primeira parte do primeiro experimento.....	43
FIGURA 17 - Curva de convergência relativa à técnica de seleção <i>roleta</i>	43
FIGURA 18 - Convergência da curva de aptidão (1ª parte do 1º experimento)....	44
FIGURA 19 – Variação do tempo de processamento (1º experimento).....	47
FIGURA 20 - Esquema da segunda parte do primeiro experimento.....	49
FIGURA 21 - Convergência da curva de aptidão (2ª parte do 1º experimento)....	49
FIGURA 22 - Gráfico da correlação entre as bandas espectrais.....	51
FIGURA 23 - Convergência da função de aptidão.....	52
FIGURA 24 - Fluxograma das etapas seguidas no segundo experimento.....	53
FIGURA 25 - Curva de convergência dos parâmetros da função de aptidão.....	54
FIGURA 26 - Função para determinação de um ponto ótimo.....	55
FIGURA 27 – Comportamento da acurácia com o incremento de variáveis.....	58
FIGURA 28 - Etapas realizadas no terceiro experimento.....	59
FIGURA 29 - Esquema das etapas do quarto experimento.....	62
FIGURA 30 - Acurácia na classificação versus quantidade de bandas.....	64
FIGURA 31 - Mapas temáticos referentes ao grupo com 21 bandas.....	66
FIGURA 32 - Mapas temáticos referentes aos maiores valores de <i>kappa</i>	68

RESUMO

A análise de imagens hiperespectrais possibilita um estudo mais detalhado sobre os objetos na superfície da terra. Devido sua alta resolução espectral, a tarefa de análise desses dados defronta-se com o fenômeno de *Hughes*. Esse fenômeno ocorre devido ao número geralmente limitado de amostras frente a alta dimensionalidade dos dados hiperespectrais. Uma das possíveis soluções para esse problema é a redução da dimensionalidade. Essa técnica de redução de dimensão é dividida em duas abordagens: seleção de feições e extração de feições. A proposta desse trabalho foi o uso dos Algoritmos Genéticos como redutores de dimensionalidade de dados hiperespectrais, para fins de classificação. Foi realizado um estudo comparativo entre os algoritmos propostos e os tradicionais algoritmos seqüenciais: *Sequential Forward Selection* (SFS) e *Sequential Backward Selection* (SBS). Esses dois últimos algoritmos são conhecidos pela simplicidade tanto conceitual como computacional. Uma segunda comparação foi realizada entre algoritmos de seleção de feições e um algoritmo do tipo extração de feições (Análise por Componentes Principais). Nesse estudo foi analisado tanto a acurácia no processo de classificação como o tempo de processamento demandado por esses algoritmos. Após a redução da dimensionalidade, uma classificação foi executada. O classificador utilizado foi de máxima verossimilhança. Para julgamento da acurácia foram utilizados os índices *kappa*, acurácia global e a acurácia do ponto de vista do usuário. A metodologia proposta mostrou-se promissora, pois apresentou resultados, em termos de acurácia, superiores aos gerados pelos demais algoritmos estudados.

ABSTRACT

The detailed study of the objects in the earth surface makes possible by the images hyperspectral analysis. Due to high spectral resolution of data, the task of the analysis confronts with the well known Hughes' phenomenon. That phenomenon usually happens when the limited number of samples is compared to the dimensionality of the hyperspectral data. One of the possible solutions, for that problem, is the reduction of the dimensionality. This technique is divided in two approaches: feature selection and feature extraction. The proposal of that work was the use of the Genetic Algorithms as reducers of dimensionality of data hyperspectral, for classification ends. A comparative study was accomplished between the proposed algorithms and the traditional algorithms sequential: Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS). Those last two algorithms are known by the simplicity so much conceptual as computational. A second comparison was accomplished between algorithms of selection of features and an algorithm of the type extraction of features (Principal Components Analysis). In that study it was analyzed the accuracy as much as in the classification process as the time of processing disputed by those algorithms. After the reduction of the dimensionality, a classification was executed. The maximum likelihood classifier was used and also the kappa index, the global accuracy and the user's point of view accuracy were used for quality judgment. The proposed methodology was shown promising, because its results, in accuracy terms, was superior to the generated by the other studied algorithms.

CAPÍTULO I - INTRODUÇÃO, REVISÃO BIBLIOGRÁFICA E MATERIAL

1. INTRODUÇÃO

Sensores multiespectrais tradicionais, freqüentemente usados para observar a superfície da terra, possuem limitações tecnológicas no que tange ao número de bandas espectrais, isto é, permitem coletar informações em uma, ou no máximo duas dezenas de bandas. O aperfeiçoamento dos sensores possibilitou a coleta de informações em várias dezenas e até centenas de bandas espectrais. Essa mudança caracteriza a origem dos sensores hiperespectrais. Exemplo de sensores hiperespectrais é o AVIRIS (Airborne Visible/Infrared Imaging Spectrometer). Esse sensor produz 224 bandas espectrais na faixa de 0,4 a 2,5 μ m do espectro eletromagnético. Devido à característica hiperespectral, esses sensores possuem bandas estreitas e contínuas, e conseqüentemente, seus dados são altamente correlacionados.

A alta dimensionalidade e correlação dos dados hiperespectrais ocasionam possibilidades e desafios no processo de classificação de imagens digitais. Isto é, as técnicas de classificação utilizadas com sucesso nos dados multiespectrais, não apresentam a mesma eficiência nos dados hiperespectrais. Uma das causas dessa ineficiência é devida ao número geralmente reduzido de amostras, frente ao grande número de variáveis. O número pequeno de amostras fornece parâmetros incertos para os classificadores estatísticos e, portanto, o desempenho do classificador é comprometido. Outro problema associado à alta dimensionalidade é o custo computacional elevado.

A redução da dimensão é discutida, na literatura, em duas abordagens distintas: seleção de feições e extração de feições. No processo de seleção de feições o objetivo é selecionar um subgrupo de variáveis dentro do grupo original. Esse subgrupo selecionado deve ser ótimo para determinada finalidade. Por exemplo, para uma posterior classificação.

Na abordagem extração de feições, a finalidade é extrair um subgrupo de variáveis por meio de uma transformação das variáveis originais. Uma desvantagem

dessa abordagem é que o subconjunto de variáveis extraídas não tem um significado físico, devido à transformação realizada.

Nessa pesquisa será realizada uma análise dos resultados dos dois grupos: seleção de feições e extração de feições. No primeiro grupo, além dos Algoritmos Genéticos propostos, são estudadas as técnicas de redução de dimensão clássicas: seleção seqüencial crescente, seleção seqüencial decrescente. No grupo dos algoritmos de extração de feições é estudada apenas a Análise por Componentes Principais. Esta análise terá como propósito, além da identificação do efeito da redução da dimensão na acurácia da classificação, usando o classificador Máxima Verossimilhança, avaliar a potencialidade dos Algoritmos Genéticos, segundo as técnicas de seleção por *roleta*, *elitismo* e *torneio*, na redução de dimensionalidade. Adicionalmente, por se tratar de um parâmetro importante na classificação, o tempo de processamento também será verificado.

É utilizada nesse trabalho uma imagem do sensor AVIRIS, disponibilizada pela NASA, com 220 bandas. A proposta dessa pesquisa é a apresentação um algoritmo para a função de aptidão dos Algoritmos Genéticos que possibilita a seleção de subconjuntos de bandas, a partir da imagem original, que apresentem maior acurácia na classificação.

1.1 OBJETIVOS

O objetivo geral dessa pesquisa é apresentar uma técnica de redução de dimensionalidade de dados hiperespectrais, baseada em Algoritmos Genéticos, para fins de classificação.

Como objetivos específicos pretende-se:

- Analisar formas de adaptação da função de aptidão (*fitness*) dos Algoritmos Genéticos ao problema proposto, visando os melhores resultados na classificação.
- Avaliar o potencial do uso de Algoritmos Genéticos na eliminação de bandas com ruídos.
- Avaliar a eficiência dos Algoritmos Genéticos em relação aos algoritmos: Seleção Seqüencial Crescente e Seleção Seqüencial Decrescente.

- Analisar a qualidade dos mapas temáticos gerados a partir do uso dos Algoritmos Genéticos, utilizando como referência os produtos adquiridos a partir do uso do algoritmo de extração de feição, Análise por Componentes Principais.

1.2 JUSTIFICATIVA

A propriedade intrínseca dos algoritmos de seleção de feições foi a principal justificativa para a realização desse trabalho. A propriedade intrínseca mencionada refere-se à conservação da característica física dos dados, e conseqüentemente, à permissão de relação direta com os dados originais.

A proposta de uma metodologia inovadora no processamento digital de imagens, juntamente com o pequeno número de estudos realizados com os dados hiperespectrais no país, são um outro fator relevante dessa pesquisa.

A redução da dimensão, no contexto geral, é uma técnica que apresenta vantagens do ponto de vista econômico, pois contribui tanto para o armazenamento como para o processamento de dados de alta dimensionalidade. Isto é, após a redução da dimensão, os dados serão armazenados num menor espaço de memória, e o tempo exigido em quaisquer técnicas de processamento, será reduzido.

Apresentação de uma técnica que favorece a separação das bandas ruidosas daquelas sem ruídos.

1.3 ESTRUTURA DA DISSERTAÇÃO

O texto dessa dissertação foi dividido em três partes principais. A primeira é constituída da introdução e revisão bibliográfica (ou revisão de literatura) e material utilizado. Enquanto, a segunda parte apresenta a metodologia proposta e, por fim, a terceira parte mostra os experimentos, a conclusão e algumas recomendações.

A introdução é dividida em três seções. A primeira seção apresenta os objetivos: geral e específicos. Na segunda é apresentada a justificativa desse trabalho.

A revisão bibliográfica discorre sobre seis tópicos distintos: imagens hiperespectrais, classificação de imagens digitais, redução de dimensionalidade, medidas de separabilidade, critério para redução, e, finalmente, os Algoritmos Genéticos.

A metodologia proposta, abordada na segunda parte, é dividida em duas etapas. Na primeira etapa é apresentado o cálculo do número mínimo de bandas a serem preservadas; na segunda é apresentada a construção da função de aptidão.

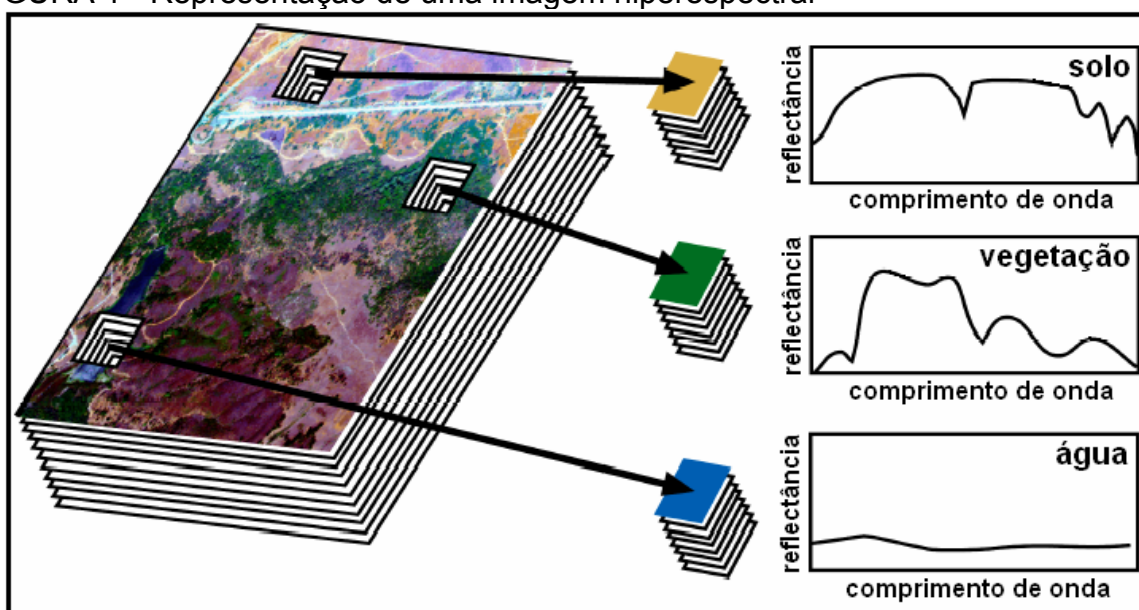
Na terceira parte os experimentos foram divididos em cinco etapas. Em cada uma dessas etapas são mostrados os resultados e discussão dos respectivos experimentos. Finalmente, ainda na terceira parte, foram apresentadas as duas últimas partes: a conclusiva e a de sugestões para trabalhos futuros.

2. REVISÃO BIBLIOGRÁFICA

2.1 IMAGENS HIPERESPECTRAIS: APLICAÇÕES

As imagens digitais hiperespectrais, como mostradas na FIGURA 1, são dezenas ou até centenas de matrizes (bidimensional) representativas da energia refletida pelos objetos presentes na superfície da terra. A assinatura espectral de um objeto é mais detalhada nas imagens hiperespectrais do que nas multiespectrais. As assinaturas espectrais, no nível de detalhe mostrado na FIGURA 1, eram obtidas apenas em laboratório, mas na atualidade, com o surgimento dos sensores hiperespectrais, tornou-se possível estimar essas assinaturas a partir dos dados fornecidos por esses sensores.

FIGURA 1 - Representação de uma imagem hiperespectral



FONTE: Modificado de SHIPPERT (2005).

As bandas espectrais componentes dessas imagens são subdivisões estreitas do espectro eletromagnético, geralmente com largura em torno de 10nm, o que caracteriza esse tipo de sensor como de alta resolução espectral. Além do grande volume de dados, uma outra característica dessas imagens é a alta correlação entre as bandas. A alta correlação é conseqüente do estreito intervalo espectral (10nm) representativo de cada banda e da continuidade dessas ao longo do espectro eletromagnético.

Devido sua informação espectral, as imagens hiperespectrais são essenciais em determinados estudos. Em estudos de separabilidade de classes, por exemplo, essas imagens, por fornecerem informações mais detalhadas sobre as feições a serem identificadas, proporcionam uma melhor discriminação de classes similares. Para solução de problemas de separabilidade de classes similares, as tradicionais imagens multiespectrais não são suficientes.

Infelizmente os dados hiperespectrais ainda são raridades no Brasil. O avanço rápido dessa tecnologia, e conseqüentemente a diminuição no custo, alimenta expectativas da criação de um acervo desse tipo de dados (ALIXANDRINI JÚNIOR, 2004).

A seguir serão apresentados alguns trabalhos que utilizaram como fonte de dados, as imagens hiperespectrais.

LUCCA et al. (2005) fizeram uso de dados do sensor HSS (Hyperspectral Scanner System, possuidor de 50 bandas espectrais ao longo do espectro refletido; foi utilizado no projeto do Sistema de Vigilância da Amazônia - SIVAM), para mapear as variações da temperatura associada com a descarga de água aquecida nas adjacências da Central Nuclear em Angra dos Reis. No contexto da metodologia proposta, concluíram que esses dados são adequados para tal fim, recomendando ainda sua utilização em diversas outras aplicações, tais como, detecção de derrame de óleo no mar, poluição, dispersão de sedimentos, monitoramento de ressurgências e frentes oceânicas.

Uma proposta para avaliar e monitorar os diferentes tipos de contaminação em ambientes mineiros, na Mina de São Domingos em Portugal, foi proposta por QUENTAL et al. (2005). Os resultados preliminares obtidos evidenciaram as potencialidades das imagens hiperespectrais no processo de avaliação da poluição associada à atividade mineira.

Buscando analisar o depósito supergênico de Níquel em Goiás, CARVALHO JÚNIOR et al. (2003), comprovaram que dados hiperespectrais representam uma nova abordagem com relação às técnicas de sensoriamento remoto para o uso geológico, isto é, estabelecem uma mudança epistemológica no processamento digital de imagens, apresentando como principal contribuição a possibilidade de avanços no estudo do *pixel* mistura (mistura espectral).

Na tentativa de identificar com precisão os diferentes componentes que formam uma cena, como solo e vegetação, D'AGOSTINI et al. (2001) apontaram as imagens hiperespectrais como uma nova concepção do sensoriamento remoto que possibilitaram sucesso em tal pesquisa.

2.2 CLASSIFICAÇÃO DE IMAGEM DIGITAL

O objetivo maior da classificação de imagens digitais é o de categorizar automaticamente todos os *pixels* de uma imagem a uma determinada classe ou tema de cobertura da terra (LILLESAND; KIEFER, 1987). A classificação de imagens digitais, segundo a participação do intérprete, é realizada de duas formas distintas: a classificação supervisionada e a classificação não supervisionada.

No processo de classificação supervisionada existe uma interação constante do analista com o sistema de análise de imagem. O analista, nesse caso, usa seu conhecimento prévio sobre a cena, ou seja, fornece informações sobre a “verdade de campo” e/ou informações do comportamento espectral dos alvos existentes nessa cena. Baseado nesse conhecimento prévio, o analista escolherá um conjunto de amostras representativas para cada alvo de interesse. A escolha das amostras de treinamento é de fundamental importância na classificação supervisionada, ou seja, elas devem ser selecionadas de forma que as classes de interesse sejam bem representadas. Essas amostras devem ser obrigatoriamente, homogêneas (NOVO, 1992). Outro fator importante é o tamanho dessas amostras, pois o aumento da dimensionalidade requer maior número de amostras (ZORTEA et al. 2005).

Pode-se caracterizar a classificação supervisionada por três passos:

- a) Seleção das amostras de treinamento para cada classe espectral. Para tanto, faz-se necessário uma fonte de dados de referência, como observações em campo, fotografias aéreas ou mapas topográficos, por exemplo.
- b) Complemento da fase de aprendizado do processo de classificação. Esse passo consiste basicamente em estimar parâmetros estatísticos populacionais para descrever as classes a partir dos dados de treinamento (puros e representativos) de cada classe.

c) O terceiro passo é a classificação propriamente dita, na qual cada *pixel* é classificado de acordo com a maior probabilidade relativa de pertencer a uma determinada classe.

Essa definição de classificação supervisionada é baseada na suposição de que as classes podem ser modeladas por distribuição de probabilidade, conseqüentemente são descritas pelos parâmetros dessa distribuição.

Segundo RICHARDS; JIA (1998), na classificação supervisionada é usual supor, no âmbito do sensoriamento remoto, que cada classe ou subclasse espectral pode ser descrita por uma distribuição de probabilidade no espaço multiespectral.

Cabe ressaltar que existem também outras técnicas de classificação supervisionada que não se baseiam nos parâmetros de distribuição probabilística, como exemplo, classificação por redes neurais artificiais (RICHARDS; JIA, 1998).

A outra forma de classificação, a não supervisionada, é independente do conhecimento *a priori* da classe espectral, isto é, as amostras não possuem rótulos conhecidos *a priori*. A classificação não supervisionada, segundo RICHARDS; JIA, (1998), é útil na fase inicial de análise para estimar as classes espectralmente separáveis e facilitar uma análise posterior detalhada, pelo método de classificação supervisionada.

Nesse trabalho empregou-se a classificação supervisionada e o algoritmo utilizado foi o classificador Máxima Verossimilhança. Esse classificador utiliza o vetor de médias e a matriz variância-covariância dos valores digitais, típicos de cada classe, para classificar um *pixel* desconhecido. No processo de classificação é assumido que a distribuição da nuvem de pontos (números digitais), formando a categoria de amostras de treinamento, é Gaussiana (distribuição normal). No caso da resposta espectral de alvos presente nas imagens de sensoriamento remoto, esta suposição é geralmente razoável. Assim sendo, a distribuição da resposta padrão de uma classe pode ser completamente descrita pelo vetor de média e pela matriz covariância; e a partir desses parâmetros torna-se possível o cálculo da probabilidade estatística de um *pixel* pertencer a uma determinada classe (RICHARDS; JIA, 1998). A probabilidade de pertinência é definida a partir da função densidade de probabilidade.

A teoria de embasamento do classificador Máxima Verossimilhança é um caso especial de uma teoria mais geral denominada de Bayesiana (ou Teorema de

Bayes). Essa é uma teoria mais rigorosa que envolve, entre outros fatores, a probabilidade *a priori* de cada classe. Ela ainda busca ponderar o custo do erro na classificação em cada classe. Apesar dessa teoria otimizar o resultado, ela é apenas teórica, pois na prática, em sensoriamento remoto, geralmente não se conhece a probabilidade *a priori*, nem o custo do erro de classificação. Para resolver tal problema, o classificador Máxima Verossimilhança assume igual probabilidade de ocorrência e igual custo do erro de classificação para todas as classes, (RICHARDS; JIA, 1998).

O principal empecilho do classificador Máxima Verossimilhança é o grande esforço computacional exigido durante o processo de classificação, Quanto maior o número de bandas espectrais e/ou de amostras, maior será o custo computacional. Por outro lado, um maior número de amostras proporciona maior desempenho do classificador. A qualidade dos resultados do classificador Máxima Verossimilhança decorre da dimensão da matriz variância-covariância e do vetor de médias usados para descrever cada classe. A redução de dimensionalidade é, portanto, um meio conveniente de otimização do classificador Máxima Verossimilhança.

Segundo RICHARDS; JIA (1998), A teoria do classificador Máxima Verossimilhança pode ser entendida como mostrada a seguir:

O elemento x , pertence à classe w_i se

$P(x \in w_i) > P(x \in w_j) \quad \forall \quad i \neq j$ ou equivalentemente, se

$$P(w_i | x) > P(w_j | x) \quad \text{onde} \quad P(w_i | x) = \frac{P(x | w_i)P(w_i)}{P(x)} \quad (1)$$

Admitindo a existência de uma distribuição Gaussiana dos elementos de cada classe tem-se para n-variáveis

$$P(x | w_i) = (2\pi)^{-n/2} |\Sigma_j|^{-1/2} \exp\left(-0.5(x - \mu_j)^t \Sigma_j^{-1} (x - \mu_j)\right) \quad (2)$$

onde, $|\Sigma_j|$ é o determinante da matriz de variância-covariância para a classe w_j ;

Σ_j^{-1} é a inversa da matriz de variância-covariância para a classe w_j ;

μ_j é o vetor médio da classe j .

Considerando um mesmo pixel, $P(x)$ é constante e pode ser desconsiderado em (1).

O termo $P(w_i)$ representa a probabilidade a priori de ocorrer classe w_i ; como ela é desconhecida, um valor constante é adotado. Esta simplificação, porém, tem como consequência o cálculo de um valor diferente daquele que seria o verdadeiro valor da probabilidade. Por esse motivo, seu resultado é chamado de função discriminante. Com isso a regra de classificação passa a ser:

$$G(x) = \ln(|\Sigma_j|) + (x - \mu_j)^t \Sigma_j^{-1} (x - \mu_j) \quad (3)$$

$$x \in w_j \text{ se } G_j(x) < G_k(x) \quad \forall j \neq k$$

A verificação da acurácia dos mapas temáticos resultantes do processo de classificação é realizada utilizando a matriz dos erros (ou matriz confusão). São extraídos dessa matriz: a estimativa global, a exatidão do ponto de vista do usuário e o denominado índice, coeficiente ou estatística *kappa*.

A estimativa global é a razão entre o total de pixels de teste classificados corretamente e o total de pixels utilizados como amostra de teste. Entretanto a exatidão do ponto de vista do usuário é calculada dividindo-se o valor dos pixels classificados corretamente em uma determinada classe pelo total de pixels classificados como pertencentes a essa classe (pixels classificados corretamente somado com os pixels classificados erroneamente como pertencente a uma dada classe). Por fim, o índice *kappa*, que é uma técnica de avaliação de dados discretos e multinomialmente distribuídos. Esse coeficiente é usualmente utilizado em classificação de imagens (MOREIRA, 2003). Uma vantagem desse índice sobre os dois apresentados anteriormente é fato de incluir no seu cálculo não somente os elementos da diagonal da matriz de erro, mas todos os elementos dessa matriz. Segundo MOREIRA (2003) um valor *kappa* dentro do intervalo de 0 a 0,2 é considerado ruim; no intervalo de 0,6 a 0,8 indica qualidade muito boa e no intervalo de 0,8 a 1 é considerado como excelente.

O índice *kappa* proposto por COHEN (1960) é obtido a partir da equação 4.

$$\kappa = \frac{N * \sum_{i=1}^M D(i,i) - \sum_{i=1}^M SL(i) * SC(i)}{N^2 - \sum_{i=1}^M SL(i) * SC(i)} \quad (4)$$

onde N é o número de bandas;

$D(i,i)$ é o elemento da diagonal da matriz de erro, na linha i ;

$SL(i)$ é o somatório da linha i ;

$SC(i)$ é o somatório da coluna i , e

M é o número de classes.

Entretanto, a obtenção de uma simples estimação pontual do valor de κ não proporciona nenhuma indicação da precisão dessa estimação. O conhecimento da variabilidade dos estimadores, e o emprego dessa informação, tanto na formulação de contrastes de hipóteses como na construção de intervalos de confiança são de fundamental importância no contexto da Estatística Inferencial (GALPARSORO; FERNÁNDEZ, 1999).

Para a formulação de contrastes de hipóteses, BISHOP; FIENBERG; HOLLAND (1977) propuseram a equação 5 para cálculo da variância assintótica quando suposto que κ seja diferente de zero.

$$\sigma^2(\kappa) = \frac{T_1(1-T_1)(1-T_2)^2 + 2(1-T_1)(1-T_2)(2T_1T_2 - T_3) + (1-T_1)^2(T_4 - 4T_2^2)}{(1-T_2)^4 n} \quad (5)$$

onde, $T_1 = \sum \pi_{ii}$, $T_2 = \sum \pi_i \pi_i$, $T_3 = \sum \pi_{ii}(\pi_i + \pi_i)$, $T_4 = \sum \sum \pi_{ij}(\pi_j + \pi_i)^2$;

π_{ii} são as probabilidades da diagonal da matriz erro;

π_i são as probabilidades da linha i da matriz erro;

n é o somatório das probabilidades somadas linha a linha ou coluna a coluna.

Contrastando a hipótese nula de que κ (kappa calculado) é igual a um valor dado κ_0 (kappa estipulado), frente a uma alternativa bilateral, o modelo

estatístico empregado, será: $\frac{|\kappa - \kappa_0|}{s(\kappa)}$, onde,

$s(\kappa)$ é a raiz quadrada de $s^2(\kappa)$;

$s^2(\kappa)$ é o estimador de $\sigma^2(\kappa)$.

O valor $s^2(\kappa)$ é obtido substituindo na equação 5 as probabilidades por proporções amostrais.

O outro enfoque, referente ao intervalo de confiança para o verdadeiro valor de κ , em sua particularidade, oferece além da melhor estimação, uma medida de erro que se pode cometer com a estimação de κ . Para tanto é necessário ter

simultaneamente, κ e o intervalo de confiança (GALPARSORO; FERNÁNDEZ, 1999).

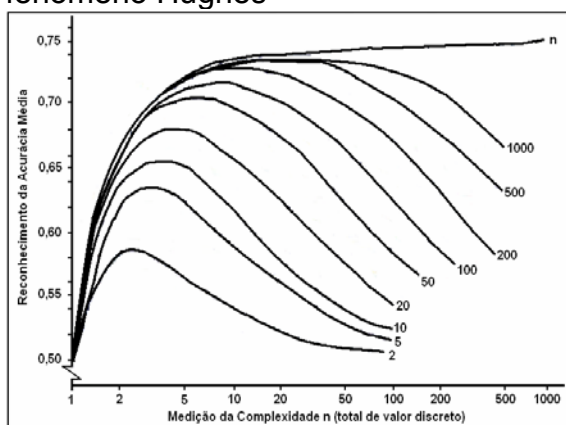
Um intervalo de confiança aproximado de $(1-\alpha)100\%$, construído pelo método padrão, é da forma: $\left[\kappa - Z_{\frac{1-\alpha}{2}} s(\kappa), \kappa + Z_{\frac{1-\alpha}{2}} s(\kappa) \right]$, onde $Z_{\frac{1-\alpha}{2}}$ é a percentagem de ordem $(1 - \frac{\alpha}{2}) * 100$ da distribuição normal padrão.

2.3 REDUÇÃO DE DIMENSIONALIDADE

O aumento de dimensionalidade teoricamente melhora a acurácia no processo de classificação (HSU; TSENG, 2005). Contudo trazem problemas como: a exigência de memória computacional para armazenamento desses dados e a diminuição no desempenho dos classificadores.

Esse segundo problema é o bem conhecido *fenômeno Hughes* (ZORTEA et al. 2005). O *fenômeno Hughes* ocorre devido ao número limitado de amostras frente a alta dimensionalidade. Isto é, para um número fixo de amostra, se for aumentada a dimensionalidade, num primeiro momento, a acurácia na classificação aumentará até um ponto máximo. A partir desse ponto, a acurácia diminuirá com a inclusão de novas variáveis, ou seja, com o aumento da dimensão. Aumentando o tamanho da amostra contorna-se esse declínio da acurácia, porém em trabalhos práticos geralmente o número de amostra é limitado. A FIGURA 2, elaborada por *Hughes* em 1968, ilustra esse fenômeno (SALEHI; ZOUJ, 2005).

FIGURA 2 – Efeito do fenômeno Hughes



FONTE: Modificada de (SALEHI; ZOUJ, 2005).

A redução na acurácia ocorre porque os parâmetros determinados para o classificador estatístico, a partir do pequeno número de amostras de treinamento, apresentam incertezas originadas da matriz variância-covariância (ZORTEA, et al. 2005).

Uma das possíveis e muito estudada solução para esse problema é a redução da dimensionalidade das feições que serão fornecidas ao classificador. A redução de dimensionalidade é uma representação do espaço original através de um subespaço de dimensões suficientemente reduzido (MARTINS JÚNIOR, 2004). O termo *dimensionalidade*, nesse contexto, refere-se ao espaço representado pelas variáveis (bandas) presentes nas imagens hiperespectrais.

Segundo SERPICO et al. (2002) a redução de dimensionalidade é realizada por duas abordagens: a primeira baseia-se na seleção do subgrupo das feições mais informativas no grupo original, de acordo com algum critério (seleção de feições). A segunda fundamenta-se na extração de um número limitado de feições, depois de realizada uma transformação ou combinação das variáveis do conjunto original (extração de feições).

A extração de feições é um passo opcional no processo de classificação, ou seja, é uma etapa utilizada no pré-processamento visando representar a imagem num formato mais apropriado para a classificação. De acordo com RAYMER et al. (2000) a extração de feições é o processo para derivação de novas feições a partir de um conjunto original, buscando tanto a redução no custo computacional como o aumento da eficiência dos classificadores estatísticos.

Segundo CONTRERAS (2004), os algoritmos de extração de feições devem ser projetados para preservar a informação de interesse num problema especial, como exemplo, na classificação. BEAUCHAMP (2004) enfatiza que esses algoritmos são modelos matemáticos e estatísticos aplicados sobre um conjunto de dados, no intuito de obter a melhor representação do dado original, eliminando a correlação entre bandas e reduzindo a dimensionalidade. Conseqüentemente, nessa abordagem, o conceito de banda pode ser perdido e a representação desses dados não pode ser correlacionada com bandas espectrais (BEAUCHAMP, 2004).

Os algoritmos de seleção de feições, apesar de eficientes, não apresentam resultados ótimos, por isso são denominados de sub-ótimos.

A principal proposta dos algoritmos de seleção de feições é reduzir o número de feições usado na classificação, mantendo uma acurácia aceitável (RAYMER, et al. 2000). De acordo com MARTINS JUNIOR (2004), esses algoritmos são divididos em três grupos principais: determinísticos com solução única, determinísticos com múltiplas soluções e estocásticos com múltiplas soluções.

Os algoritmos estocásticos geralmente retornam como solução, diversos conjuntos de características. Considerando duas execuções aplicadas aos mesmos dados de entrada, esses algoritmos retornam diferentes conjuntos solução. Os outros algoritmos, devido a característica intrínseca determinística, apresentam sempre o mesmo conjunto resposta, onde esse conjunto pode ser formado por uma única ou várias soluções, de acordo com o algoritmo utilizado (MARTINS JUNIOR, 2004).

A seguir serão apresentados alguns trabalhos em que os autores propuseram ou aplicaram os redutores de dimensionalidade.

Segundo RICHARDS; JIA (1998) o custo computacional, no processo de classificação de imagens, cresce rapidamente com o número de bandas associadas ao *pixel* de interesse, ou seja, com o aumento de dimensionalidade. Tal fato sugere a necessidade de reduzir a dimensionalidade, valorizando os sensores que possuem maior poder de discriminação espectral (CAMPOS, 2000). O processo de redução de dimensionalidade, nesse caso, apresenta um papel crucial (CADIMA, 2001).

De acordo com CAMPOS (2000) existem dois motivos primordiais para que seja realizada a redução de dimensionalidade: esforço computacional e precisão do classificador. Isto é, depois de reduzida a dimensionalidade para um espaço otimizado, o classificador terá um melhor desempenho, e conseqüentemente, demandará menos memória. Baseado ainda nos estudos desse autor, há uma alerta em relação ao teorema do *patinho feito* de Watanabe, que mostra a capacidade de fazer com que dois padrões arbitrários se tornem similares, a partir de um grande número de feições correlacionadas. Porém a redução de dimensionalidade descomunal faz com que o classificador venha a perder o poder de discriminação.

Em PIZARRO; FERNANDES (2003), que buscam a determinação dos membros de referência (valores digitais dos *pixels* representativos de cada uma das classes em estudo), encontra-se a aplicação de dois métodos de redução de dimensionalidade: Análise por Componentes Principais e a Transformada *Wavelet*

Discreta (Daubechies). Segundo esses autores, a Análise por Componentes Principais é uma das técnicas mais difundida no processamento de imagens hiperespectrais. Eles descrevem que a Transformada *Wavelet* Discreta é uma técnica que atua no domínio espectral de cada *pixel* das imagens hiperespectrais.

ZORTEA et al. (2005) propuseram três diferentes algoritmos: Segmentação Comprimento Constante (SCC), Segmentação Comprimento Variável com Otimização Central (SCV-OC) e Segmentação Comprimento Variável com Otimização Total (SCV-OT). Esses algoritmos consistem basicamente da divisão da curva de resposta espectral em segmentos contínuos, variável central e variável total, respectivamente. O objetivo é reduzir a dimensionalidade de imagens hiperespectrais (AVIRIS) através da substituição dos números digitais por um número reduzido de parâmetros estatísticos.

GALVÃO et al. (2005) trabalharam com imagens do sensor HYPERION, sensor a bordo do satélite Earth Observing-1 (EO-1). Esse sensor é capacitado para coletar informações, sobre a superfície da Terra, em 242 bandas espectrais. O trabalho desses autores visou discriminar cinco variedades de cana-de-açúcar no Estado de São Paulo. Para tanto, utilizaram um método que maximiza a distância de *Mahalanobis* entre os grupos de variedades mais similares, adequando como redutor de dimensionalidade, a análise discriminante. Concluíram finalmente, que essas imagens são adequadas no processo de discriminação de variedades de cana-de-açúcar.

2.3.1 Métodos de Redução de Dimensionalidade

Existem na literatura abordagens sobre vários métodos de redução de dimensionalidade. Serão apresentados aqui, de forma breve, alguns desses métodos que se julga mais importantes, ou pelos menos, mais utilizados.

➤ Análise por Componentes Principais

No contexto de redução de dimensão, a Análise por Componentes Principais surge como uma das técnicas (utilizada para reduzir ou remover a redundância existente entre dados) mais difundida em diversas áreas do conhecimento (PLA, 1986). A Análise por Componentes Principais pode ser usada *a priori* tanto no

processo de interpretação visual, como no processo de classificação automatizada. Uma das maiores vantagens dessa técnica, no processo de classificação, é a de propiciar um melhor desempenho aos classificadores, tendo em vista a ocorrência de redução de dimensionalidade. Depois de realizada a Análise por Componentes Principais espera-se que os dados originais pertencentes ao espaço n -dimensional possam ser representados num espaço p -dimensional, onde $n \gg p$ (LILLESAND; KIEFER, 1987).

Uma matriz M composta de DN_i variáveis observadas, descrevendo o comportamento de i indivíduos, pode ser analisada por Componentes Principais. Essa técnica caracteriza-se por uma rotação dos eixos dos dados originais para um novo sistema (espaço). Nesse processo a correlação entre as sucessivas combinações lineares é eliminada.

Os autovetores da matriz de variância-covariância (ou de correlações) representam os coeficientes que definem cada combinação linear. O autovalor pertinente a cada autovetor indica a variância da componente principal correspondente (CADIMA, 2001).

O comportamento dessa técnica, no espaço bidimensional, pode ser melhor interpretado pelas equações (6) e (7), e exemplificado pela FIGURA 3.

$$DN_{cp1} = a_{11} * DN_A + a_{12} * DN_B \quad (6)$$

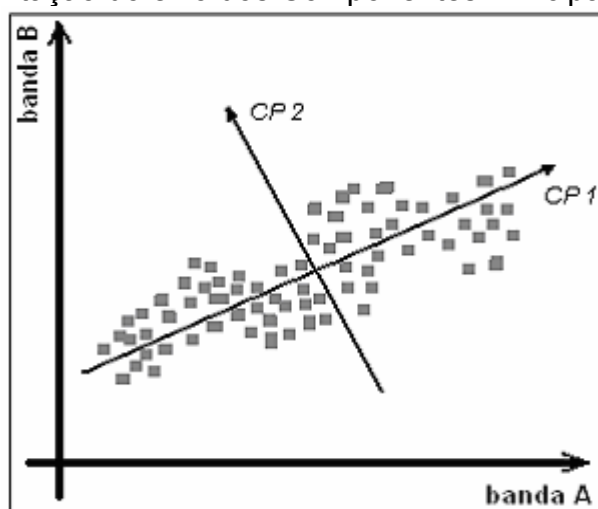
$$DN_{cp2} = a_{21} * DN_A + a_{22} * DN_B \quad (7)$$

onde, DN_{cp1} , DN_{cp2} são os números digitais do novo sistema (Componentes Principais);

DN_A , DN_B são os números digitais do sistema antigo, e

a_{11} , a_{12} , a_{21} , a_{22} são os coeficientes de transformação dos sistemas (LILLESAND; KIEFER, 1987).

FIGURA 3 - Representação do eixo dos Componentes Principais



FONTE: Modificado de LILLESAND; KIEFER (1987).

Analisando a FIGURA 3 é notória a distribuição da maior variância ao longo do primeiro componente principal (CP_1). Percebe-se também que o segundo componente principal CP_2 abrange, de forma perpendicular à CP_1 , a segunda maior variância. O raciocínio é análogo para o espaço n-dimensional, isto é, a i-ésima componente principal (CP_i) possuirá a i-ésima maior variância. Uma discussão mais detalhada sobre Análise por Componentes Principais pode ser encontrada em: PLA, 1986; CROSTA, 1993; JOHNSON e WICHERN, 1998; RICHARDS; JIA, 1998.

➤ Transformada *Wavelet*

Segundo MOIGNE (2005) este método é aplicado sobre o domínio espectral de cada *pixel*. Após realizada a decomposição *Wavelet* sobre a assinatura de cada *pixel*, os coeficientes resultantes dessa decomposição apresentarão uma redução de dimensionalidade dos dados, e estes poderão ser usados em aplicações, tais como na classificação supervisionada.

A ordem de complexidade produzida pelo método de redução baseado em *Wavelet* é de (MN) , onde N é o número de bandas e M é o número de *pixels* no domínio espacial, enquanto que na análise por componentes principais essa ordem é de $((MN^2) + N^3)$. Isso mostra a eficiência computacional da *Wavelet* sobre a Análise por Componentes Principais.

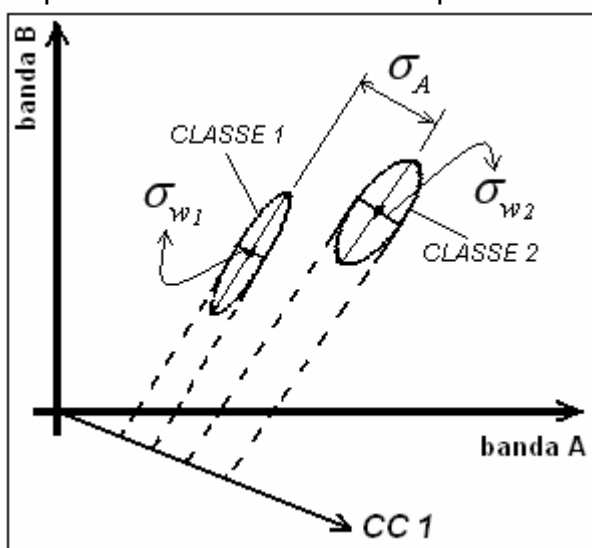
MOIGNE (2005) apresenta resultados indicando esse método como promissor. A justificativa é que esse método foi mais eficiente que a Análise por

Componentes Principais, tanto do ponto de vista de eficiência computacional como de exatidão total na classificação.

➤ Análise por Componentes Canônicas

A Análise de Componentes Canônicas é uma técnica apropriada para separabilidade, quando se têm informações *a priori* sobre as feições que serão separadas (LILLESAND; KIEFER, 1987). A Análise de Componentes Canônicas busca posicionar um sistema de eixos de forma que as variâncias: interna e entre as classes, sejam minimizadas e maximizadas, respectivamente (RICHARDS; JIA 1998). Essa propriedade é a diferença básica entre a Análise por Componentes Canônicas e a Análise por Componentes Principais, pois essa última não se preocupa com a variância interna das classes. A FIGURA 4 mostra um exemplo hipotético do funcionamento da Análise por Componentes Canônicas, num espaço bi-dimensional.

FIGURA 4 – Exemplo hipotético dos eixos dos Componentes Canônicas



FONTE: Adaptada de RICHARDS; JIA, (1998).

Segundo JOHNSON; WICHERN (1998), a Análise de Componentes Canônicas é uma técnica estatística que busca identificar e quantificar a associação entre dois grupos de variáveis, tendo por objetivo a determinação das combinações lineares c_1x e c_2y (onde x e y são as variáveis e c_1 , c_2 são os coeficientes das combinações lineares) com maior correlação possível, a fim de discernir o relacionamento entre os dois grupos de variáveis.

A FIGURA 4 apresenta o sistema de eixos representativo da Análise de Componentes Canônicas em relação ao sistema original dos dados. Percebe-se que a primeira componente de correlação (CC 1), possui a maior percentagem de informação sobre variância mínima e máxima separabilidade. Para o espaço N-dimensional o raciocínio é análogo ao apresentado para Análise por Componentes Principais, isto é, a i-ésima Componente Canônica recebe o i-ésimo maior valor referente à capacidade de separabilidade de classes.

A Correlação Canônica pode ser calculada a partir da equação 8:

$$\rho^* = \text{corr}(U, V) \quad (8)$$

onde $U = c_1 x$ e $V = c_2 y$. Nestas duas últimas igualdades,

x e y são variáveis;

$c_1 = e_i \Sigma_i^{-1/2}$; $c_2 = f_i \Sigma_j^{-1/2}$, donde

e_i é o autovetor correspondente ao i-ésimo autovalor de $\Sigma_i^{-1/2} \Sigma_{ij}^{-1} \Sigma_{ji} \Sigma_i^{-1/2}$, e

f_i é o autovetor correspondente ao i-ésimo autovalor de $\Sigma_j^{-1/2} \Sigma_{ji}^{-1} \Sigma_{ij} \Sigma_j^{-1/2}$.

Ressalta-se que os vetores de autovalores, correspondentes aos dois casos, podem ter dimensões diferentes. Finalmente,

Σ_i , Σ_j são as matrizes variância-covariância de x e y , e

Σ_{ji} é a matriz de covariância cruzada (JOHNSON; WICHERN, 1998).

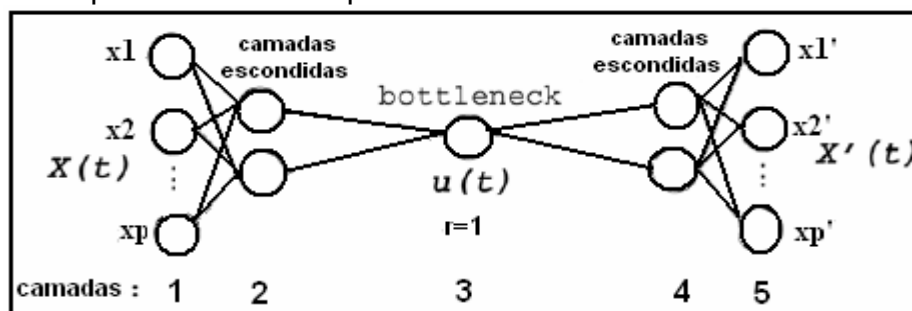
➤ Redes Neurais Artificiais

Os métodos de redução de dimensionalidade tratados acima são de característica linear, todavia, Redes Neurais Artificiais são redutores de dimensão não linear, ou seja, são algoritmos estocásticos que levam em consideração vários parâmetros de entrada (informações), fazendo uma busca aleatória no espaço n-dimensional.

BOTELHO et al. (2005) propuseram um método de redução de dimensão através da Análise por Componentes Principais Não-Lineares, utilizando Redes Neurais Artificiais, denominado de C-NLPCA (Cascaded nonlinear principal component analysis). Como a Análise por Componentes Principais é restrita a um julgamento simplesmente linear dos dados, o método proposto tem a vantagem de atuar de forma não linear, ou seja, usando multi-camadas em Redes Neurais

Artificiais. A FIGURA 5 mostra a idéia de como essa técnica atua sobre os dados, buscando reduzir a dimensão.

FIGURA 5 – Mapeamento de componente não-linear usando Redes Neurais



FONTE: Adaptado de BOTELHO et al. (2005)

Na execução da NLPCA, a Rede Neural Artificial usa três camadas de neurônios denominadas de camadas escondidas, que se posicionam entre as camadas de entrada e de saída. Estas camadas escondidas têm funções não-lineares de ativação funcionando entre a camada de entrada e o neurônio *bottleneck* (garganta ou pescoço de garrafa), e entre a camada de saída e o neurônio *bottleneck* como mostra a FIGURA 5. A função da Rede Neural, neste caso, é modelar uma composição dessas funções. Das cinco camadas da NLPCA, as camadas de entrada e saída (1 e 5) tem p nós, a camada 3 tem r nós, sendo $r < p$. O esperado é que os nós de saída reproduzam os sinais de entrada apresentados à rede. Os nós das camadas 2 e 4 devem ter funções não-lineares de ativação, e os nós das camadas 1, 3 e 5 usam funções lineares de ativação (BOTELHO et al., 2005).

Baseado ainda nas afirmações desses autores, é confirmado que a rede de NLPCA permite a compressão e redução dos dados, porque as entradas p -dimensional devem passar através da camada r -dimensional do *bottleneck* antes de reproduzir as entradas. Como $r < p$, a informação contida na camada *bottleneck* é então reduzida em relação àquela contida na camada de entrada.

➤ Seleção Seqüencial Crescente (Sequential Forward Selection – SFS)

Essa metodologia adota a estratégia *bottom-up*, isto é, trata-se de um algoritmo que, baseado num critério, seleciona uma variável a cada passo. Esse processo é repetido até que seja preenchido um subconjunto de interesse. Segundo

BON (2005) esse algoritmo pode ser escrito como segue: sendo X o vetor das variáveis (ou bandas); d o número total de variáveis; $J()$ como uma função critério que aponta as melhores variáveis no conjunto original; e $Y_0 = \emptyset$ um conjunto de variáveis, o processo iterativo pode, então, ser iniciado. O algoritmo seqüencial crescente inicia o processo iterativo com Y_0 , e, após k passos, o novo conjunto Y_k possuirá k variáveis, sendo, portanto, uma variável selecionada em cada passo iterativo. Sendo assim, Y_k e $B = (X - Y_k)$ têm tamanhos iguais a k e $d - k$, respectivamente. As variáveis B_j do conjunto de variáveis restantes são ordenadas de tal forma que: $J(Y_k \cup \{B_1\}) \geq J(Y_k \cup \{B_2\}) \geq \dots \geq J(Y_k \cup \{B_{d-k}\})$.

Por fim, seleciona-se a variável B_1 e junta-a ao conjunto Y_k para formar Y_{k+1} . Sendo $Y_{k+1} = Y_k \cup \{B_1\}$. A FIGURA 6 apresenta o escopo desse algoritmo.

A função critério $J()$ foi elaborada utilizando a divergência transformada.

FIGURA 6 – Algoritmo Seleção Seqüencial Crescente

```

ALGORITMO SFS(X,p)

Entradas
  X      - Conjunto completo de variáveis.
  p ∈ ℕ  - Número de variáveis a selecionar

Saídas
  Y      - Conjunto de p variáveis selecionadas de X.

Algoritmo
  Y₀ ← ∅ - Conjunto inicial de variáveis escolhidas
  L₀ ← X - Conjunto inicial de variáveis disponíveis

  para k variando de 0 até (p-1)
    seja B: J(Yₖ U B) = máx {J(Yₖ U B)}
    Yₖ₊₁ ← Yₖ U B
    Lₖ₊₁ ← Lₖ - B
  fim
  Y ← Yₚ - Último conjunto de variáveis escolhidas

```

FONTE: Modificado de BON, (2001).

De acordo com SERPICO et al. (2002) esse método de busca tem uma baixa exigência computacional, porém apresenta uma desvantagem: uma feição selecionada não pode mais ser eliminada numa próxima iteração. Nos Algoritmos Genéticos, a permuta de uma banda já selecionada, por outra, é permitida.

➤ Seleção Seqüencial Decrescente (Sequential Backward Selection – SBS)

Essa metodologia adota a estratégia *top-down*, isto é, trata-se de um algoritmo que, baseado num critério, seleciona uma variável que será eliminada, em cada passo. Esse processo é repetido até que o subconjunto restante seja do tamanho de interesse.

Segundo BON (2001) esse algoritmo pode ser escrito como segue: considera-se X como o vetor das variáveis (ou bandas); d como o número total de variáveis; e $J()$ como uma função critério que aponta as melhores variáveis do conjunto para serem eliminadas. Sendo assim o algoritmo começa com o conjunto original das variáveis, $\tilde{Y}_0 = X$. Após k iterações são descartadas k variáveis e o conjunto original \tilde{Y}_0 é reduzido ao conjunto \tilde{Y}_k . Nesse caso, \tilde{Y}_k possui tamanho igual a $d - k$. Posteriormente ordenam-se as variáveis B_j do conjunto de variáveis disponíveis \tilde{Y}_k , de tal forma que: $J(\tilde{Y}_k - \{B_1\}) \geq J(\tilde{Y}_k - \{B_2\}) \geq \dots \geq J(\tilde{Y}_k - \{B_{d-k}\})$.

Por fim, seleciona-se B_1 e elimina-o de \tilde{Y}_k para formar \tilde{Y}_{k+1} . Sendo $\tilde{Y}_{k+1} = \tilde{Y}_k - \{B_1\}$. A FIGURA 7 apresenta o escopo desse algoritmo.

FIGURA 7 – Algoritmo Seleção Seqüencial Decrescente

```

ALGORITMO SBS(X,p)
Entradas
  X      - Conjunto completo de variáveis.
  p ∈ ℕ  - Número de variáveis a descartar
Saídas
  Y      - Conjunto de p variáveis selecionadas de X.
Algoritmo
   $\tilde{Y}_0 \leftarrow X$  - Conjunto inicial de variáveis escolhidas
  para k variando de 0 até (d-p-1)
    seja B:  $J(\tilde{Y}_k - B) = \max \{J(\tilde{Y}_k - B)\}$ 
     $\tilde{Y}_{k+1} \leftarrow \tilde{Y}_k - B$ 
  fim
   $Y \leftarrow \tilde{Y}_{d-p}$  - Último conjunto de variáveis escolhidas
  
```

FONTE: Modificado de BON, (2001).

A função critério $J()$, para esse algoritmo, foi definida utilizando a distância *Jeffries Matusita*.

2.4 MEDIDAS DE SEPARABILIDADE

Vários métodos de separabilidade entre classes são propostos na literatura, como: distância de *Mahalanobis*, distância de *Bhattacharyya*, separabilidade estatística, divergência, divergência transformada. Por se tratar de um método simplista, fundamentado em critérios quantitativos e de baixa exigência computacional (CHUVIECO, 1990), e ainda, por apresentar bons resultados (SWAIN; DAVIS, 1978), utilizar-se-á nesse trabalho a divergência transformada.

A divergência transformada, segundo CHUVIECO (1990) é um aperfeiçoamento da divergência, isto é, é uma forma de obter a divergência dentro de uma determinada escala.

Admitindo-se que os dados em estudo, são normalmente distribuídos $P(x/w_i) = N(U_i, \Sigma_i)$ (para imagens digitais essa aproximação é razoável), a divergência pode ser determinada em função do vetor das médias U_i e da matriz de covariâncias Σ_i , de acordo com a equação 9.

$$D_{1,2} = \frac{1}{2} \text{tr}[(\Sigma_1 - \Sigma_2)(\Sigma_2^{-1} - \Sigma_1^{-1})] + \frac{1}{2} \text{tr}[(\Sigma_1^{-1} - \Sigma_2^{-1})(U_1 - U_2)(U_1 - U_2)^t] \quad (9)$$

onde, $D_{1,2}$ é a medida da divergência entre as classes 1 e 2;

tr é o traço da matriz resultante do produto entre colchetes;

Σ_1 e Σ_2 são as matrizes de covariâncias das amostras nas classes 1 e 2, respectivamente, e

U_1 e U_2 são os vetores das médias das amostras nas classes 1 e 2, respectivamente.

A equação 10 mostra como é determinada a divergência transformada a partir da divergência.

$$DT_{1,2} = c[1 - e^{-\frac{D_{1,2}}{8}}] \quad (10)$$

onde, c é uma constante introduzida para proporcionar uma escala na medida da divergência transformada.

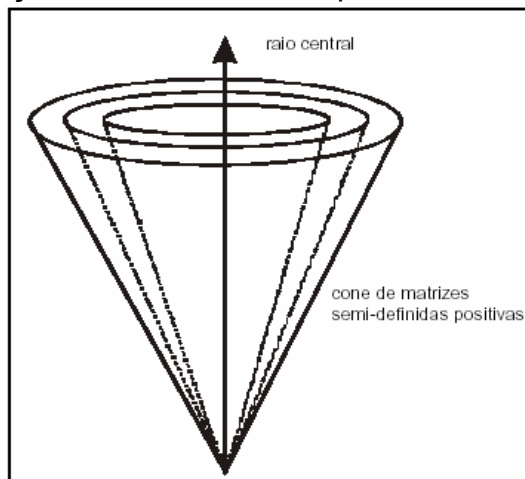
Existe uma discussão sobre qual o melhor valor para c . De acordo com CHUVIECO (1990), alguns autores recomendam $c=2000$, enquanto outros preferem

$c=100$. Esses últimos apóiam na justificativa de poderem trabalhar com os valores em porcentagem. Em RICHARDS; JIA (1998) o valor de c é igual a 2.

2.5 CRITÉRIO PARA CÁLCULO DA DIMENSÃO MÍNIMA

No processo de redução de dimensionalidade, a escolha do tamanho da dimensão reduzida tem sido definida de forma particular, onde cada analista impõe seu próprio critério. Tendo em vista esse problema, utilizar-se-á nesse trabalho o método de escolha de fatores proposto por CADIMA (2001). Utilizando a estatística e a álgebra matricial, esse método busca apresentar através da estrutura geométrica de um cone de matrizes semi-definidas positivas (FIGURA 8), o número de fatores (ou variáveis) a ser selecionado após a redução de dimensionalidade.

FIGURA 8 - Representação de um cone e três possíveis estratos



FONTE: Adaptado de CADIMA (2001).

No intuito de definir esse critério, CADIMA (2001) utilizou a teoria matemática que define a geometria de um cone e as características de uma matriz semi-definida positiva, baseando exclusivamente nas equações 11 e 12, apresentadas abaixo.

$$\cos(\alpha) = \frac{\langle kv \rangle}{\|k\| * \|v\|} \quad (11)$$

$$x' * A * x \geq 0, \forall x \in R^n / \{0\} \quad (12)$$

onde, $\langle kv \rangle$ é o traço da matriz gerada a partir do produto interno matricial de k e v ; $\|k\|$ é a norma da matriz k ;

x é um vetor qualquer; e

A é uma matriz simétrica de forma quadrática não negativa.

Dado um espaço de dimensão D , a equação 11 pode ser usada para definir o ângulo (α) entre duas matrizes (k e v) quaisquer que pertençam ao espaço D . Baseado nessa afirmação e na teoria matemática da estrutura do cone, CADIMA (2001) utilizou essa equação para definir o ângulo formado entre o eixo central (definido pela matriz identidade – I) e um outro eixo qualquer definido pelas matrizes (V) cujos autovalores (λ) pertençam ao espaço dos números reais positivos R_0^+ , e tenha dimensão igual ao espaço de I . A equação 12, por sua vez, caracteriza uma matriz simétrica cujas formas quadráticas são não negativas (semi-definida positiva). Exemplos dessas matrizes são as de covariâncias ou de correlações.

CADIMA (2001) substituiu na equação 11 a matriz k por I , e a matriz v por V . Nesse trabalho, V é a matriz semi-definida positiva das covariâncias da imagem hiperespectral utilizada.

Segundo CADIMA (2001) a pseudo-característica de uma matriz semi-definida positiva V , corresponde à menor característica das matrizes semi-definida positiva que se encontram no mesmo estrato do cone onde se localiza a matriz V . A localização desta matriz sugere a possibilidade de redução de dimensionalidade até à sua pseudo-característica. As vantagens desse método, baseado na pseudo-característica da matriz das covariâncias (Σ) ou de correlações (ρ) dos dados sob análise, são apresentadas a seguir:

- É um critério objetivo e não individual;
- Envolve cálculo simples (traços de Σ e de Σ^2);
- Utiliza a totalidade da informação disponível (variâncias em Σ);
- É de natureza geométrica e não depende da natureza dos dados estudados;
- Não exige qualquer hipótese distribucional ou modelo subjacente.

Vale enfatizar que a pseudo-característica representa a dimensionalidade mínima admissível, indicando que a matriz Σ está numa região do cone fechada para matrizes de características inferiores. Assim, podem ser consideradas quaisquer matrizes de características (dimensionalidades) superiores.

2.6 ALGORITMOS GENÉTICOS

A partir da década de 30, estudiosos da área biológica e informática começaram a se interessar pelos Algoritmos Genéticos, todavia só na década de 70 é que houve maiores avanços nessa linha de pesquisa. JOHN HOLLAND (1975), citado por GOLDBERG (1989) como o pioneiro no assunto buscou elaborar algoritmos que descrevessem o comportamento da natureza no que tange à evolução das espécies e aos processos genéticos dos organismos biológicos. O princípio de funcionamento desses algoritmos baseia-se na teoria Darwiniana de seleção natural e da genética de Mendel. A idéia principal enfocada por JOHN HOLLAND (1975), e outros estudiosos, era resolver problemas reais, em especial, problemas complexos de otimização, através de sistemas inspirados na natureza, simulando os processos naturais.

GOLDBERG (1989), um dos precursores dos Algoritmos Genéticos, descreveu-os como sendo uma pesquisa baseada no mecanismo de seleção e genética natural objetivando a otimização. Segundo esse autor, os Algoritmos Genéticos superam os outros métodos tradicionais de otimização encontrados na literatura por terem as seguintes vantagens:

- a) Trabalham com a codificação dos parâmetros e não com os dados reais;
- b) Fazem busca numa população e não num único ponto;
- c) Usam a informação de aptidão e não outro conhecimento auxiliar; e
- d) Usam regra de transição probabilística e não determinística.

LUTTON; MARTINEZ (2004) caracterizaram os Algoritmos Genéticos como uma técnica estocástica, relativamente lenta, mas de grande eficiência no processo de busca em espaço dimensional elevado.

Os Algoritmos Genéticos, a Programação Genética e a Programação Evolucionária são modelos trabalhados na área de Computação Evolucionária. As Redes Neurais Artificiais, os Sistemas Fuzzy e a Computação Evolucionária, por sua vez, fazem parte da denominada Inteligência Computacional. A Inteligência Computacional é caracterizada por manipulação numérica (e não simbólica) do conhecimento, adaptabilidade e tolerância a informações imprecisas (LOPES, 1999).

2.6.1 Teoria da Evolução das Espécies

Foram muitas as teorias que tentavam explicar a origem da vida e a evolução das espécies. Aristóteles (filósofo grego) propôs a teoria da abiogênese que perdurou até o século XIX. Essa idéia julgava que os seres inanimados, sob uma força capaz de comandar diversas reações, passavam a ter vida. Nesse século o cientista francês Luis Pasteur derrubou tal teoria, provando que a vida surgia de materiais nutritivos expostos a microorganismos vivos.

O biólogo francês Lamarck, no início do século XIX, defendeu a teoria do transformismo, onde julgava que na luta pelo alimento, os seres vivos desenvolviam fisicamente atendendo as suas necessidades alimentares, e transmitiam essas características para seus descendentes. Lamarck julgou que o crescimento das populações seria em progressão geométrica (PAULINO, 1995).

Em 1859 o naturalista inglês Charles R. Darwin, influenciado pela teoria de Lamarck, propôs a teoria da evolução das espécies. A teoria Darwiniana afirma que a evolução ocorre devido à seleção natural: “Não há exceção à regra de que cada ser orgânico aumenta naturalmente a uma média tão alta que, se não destruído, a terra logo estará povoada com a progênie de um único par” (BOLTON, 1944).

Após a realização do trabalho de Mendel (descoberta da genética), a teoria Darwiniana ganhou um significado maior, donde foi possível explicar as mudanças genéticas dos indivíduos e conseqüentemente a evolução da espécie.

2.6.2 Seleção Natural e Seleção Artificial

A principal diferença conceitual entre a seleção natural (processo que ocorre naturalmente, sem a ação do homem), e a seleção artificial (implementada nos programas), é que a seleção natural não é propriamente uma seleção, ou seja, não existem regras para que ela aconteça, enquanto a seleção artificial é dotada de regras impostas com objetivos específicos.

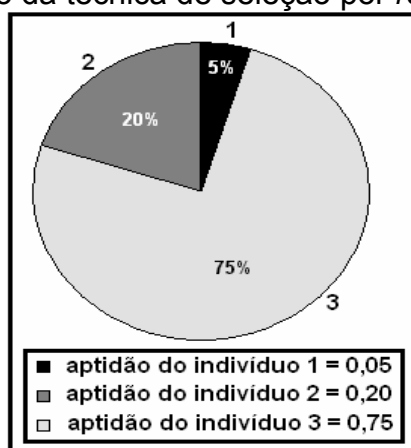
A técnica de seleção utilizada pelos Algoritmos Genéticos é a seleção artificial, porém esses algoritmos baseiam-se nos princípios de seleção natural e sobrevivência dos indivíduos mais aptos, e também nos processos genéticos dos

organismos vivos, descritos respectivamente por Charles Darwin (1809 - 1882) em *Origem das Espécies* e por Gregor J. Mendel (1822 – 1884).

No decorrer deste texto, quando se fizer menção ao termo seleção, referir-se-á à seleção artificial. Entre as técnicas de seleção mais utilizadas podem ser citadas: *roleta*, *torneio* e *elitismo* (GOLDBERG, 1989).

A seleção por *roleta* é uma técnica que consiste da escolha aleatória de um elemento. Cada elemento tem a probabilidade de ser selecionado, proporcional ao seu valor de aptidão. Para uma população com 3 indivíduos, como ilustrado na FIGURA 9, a probabilidade, por exemplo, do indivíduo 2 ser selecionado é de 0,20.

FIGURA 9 – Funcionamento da técnica de seleção por *roleta*



Existem várias formas de aplicar a técnica de seleção por *torneio*, um exemplo seria a seleção aleatória, a priori, de uma dupla de indivíduos, e posteriormente, selecionar um desses dois indivíduos, de acordo com um critério. Esse critério pode ser, por exemplo, o maior valor de aptidão. Assim, o indivíduo que possuir maior aptidão será selecionado.

Finalmente, a seleção *elitismo* é a preservação de um grupo dos indivíduos mais aptos, pertence à população atual, para a nova população. Os demais indivíduos que completarão a nova população serão selecionados por outra técnica como a *roleta* ou *torneio*.

2.6.3 Composição dos Algoritmos Genéticos

O esquema básico do escopo dos Algoritmos Genéticos é composto pelas seguintes componentes:

- a) codificação das variáveis de interesse no problema (cromossomo);
- b) configurações da população inicial;
- c) proposta da função de aptidão;
- d) definição dos Operadores Genéticos que serão utilizados no processo de reprodução (*cruzamento* e *mutação*);
- e) definição dos valores dos parâmetros que serão usados (tamanho da população e probabilidades associadas aos operadores genéticos).

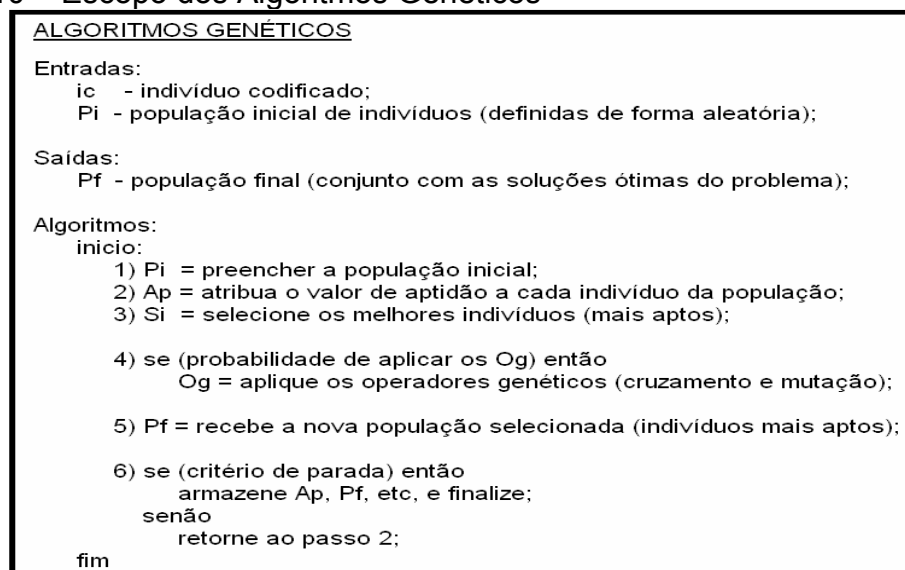
Os elementos comuns nos Algoritmos Genéticos simples, segundo MITCHELL (1997), são: a população de cromossomos, a seleção de acordo com a aptidão, o *cruzamento* para produzir nova prole, e a *mutação* aleatória da nova geração. Existe ainda um quarto elemento que é a inversão, mas por ser raramente usado, não será discutido nem aplicado nesse trabalho.

Os parâmetros (ou constantes) que definem o funcionamento dos Algoritmos Genéticos simples, propostos por GOLDBERG (1989), são:

- Tamanho máximo que a população poderá assumir;
- Tamanho máximo do indivíduo (ou cromossomo);
- Vetores que representarão os indivíduos da população;
- O valor de aptidão;
- Vetores que representarão os indivíduos da nova população.

A FIGURA 10 ilustra o escopo básico de funcionamento dos Algoritmos Genéticos.

FIGURA 10 – Escopo dos Algoritmos Genéticos



3. MATERIAL

Para o desenvolvimento dessa pesquisa foram utilizados os seguintes recursos:

- Hardware:
 - a. Computador: HD 40 Gb, RAM 256, 2.2 Sempron;
- Software:
 - a. Ambiente Matlab 5.3 pertencente à UFPR;
 - b. Software Multispec (free);
 - c. Imagem Hiperespectral disponibilizada pela NASA para fins de pesquisa;

3.1 DESCRIÇÃO DA IMAGEM HIPERESPECTRAL

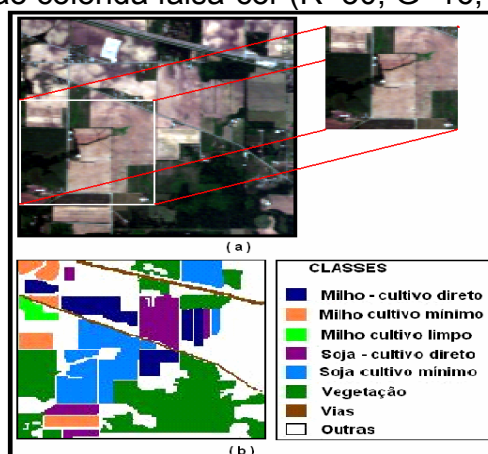
Essa imagem hiperespectral foi obtida a partir do sensor AVIRIS (aerotransportado). É uma imagem composta de 220 bandas, distribuídas ao longo do espectro eletromagnético como apresenta a TABELA 1.

TABELA 1 – Distribuição aproximada das bandas espectrais do sensor AVIRIS

	Azul	Verde	Vermelho	Infravermelho Próximo	Infravermelho Médio
Faixa Espectral (μm)	0,45 - 0,50	0,50 - 0,60	0,60 – 0,72	0,72 – 1,30	1,30 – 2,50
Bandas	1 a 5	6 a 13	16 a 25	26 a 86	87 a 220

A imagem tem resolução espacial de 20m (a resolução espacial desse sensor é função da altura do vôo) e representa uma área agrícola localizada na Califórnia/EUA, em 1994. A FIGURA 11(a) mostra uma composição colorida (falsacor) formada a partir das bandas 9, 16 e 30. Essa figura apresenta em destaque o recorte (70x70 *pixels*) que foi utilizado neste trabalho. O mapa de "verdade de campo", também fornecido pela NASA, está ilustrado na FIGURA 11(b).

FIGURA 11 – Composição colorida falsa-cor (R=30, G=16, B=9)



A justificativa da realização do recorte (70x70 *pixels*) sobre a imagem original é que um grande número de experimentos seria executado, portanto o objetivo foi ganhar tempo no processamento desses testes. Nesse recorte encontram-se as classes: milho, soja, vegetação nativa e soja cultivo direto. As demais feições presentes no recorte foram denominadas de classe "outros". A TABELA 2 mostra o tamanho das amostras utilizadas em todos os experimentos desse trabalho.

TABELA 2 – Tamanho das amostras de treinamento e verificação em *pixels*

classes	treinamento	verificação
vegetação	121	200
cultivo direto	75	97
milho	100	106
soja	194	200
outros	113	128

As classes de cultivo de milho e soja presentes nessa imagem são altamente semelhantes entre si. Em dados multiespectrais é impossível a separabilidade dessas classes, porém em dados hiperespectrais, devido ao grande número de bandas ao longo do espectro, torna-se possível a separação dessas classes. Baseado nessa premissa os Algoritmos Genéticos selecionam as bandas que apresentam a melhor separabilidade das classes.

CAPÍTULO II - METODOLOGIA PROPOSTA

4. MÉTODOS

4.1 ADAPTAÇÃO DOS ALGORITMOS GENÉTICOS

Foram utilizados neste trabalho os Algoritmos Genéticos Simples propostos por GOLDBERG (1989). Como o próprio nome profere, este é um simples algoritmo composto pelas componentes básicas apresentadas na seção 2.6.3, porém adaptável ao problema proposto. A seguir serão apresentadas as etapas necessárias para o ajuste desses algoritmos ao problema em questão:

- a) codificação dos indivíduos e definição da população;
- b) operadores genéticos utilizados;
- c) técnicas de seleção;
- d) descrição da função de aptidão;
- e) medida de similaridade entre classes através da divergência transformada; e
- f) medida de similaridade entre bandas através da matriz de correlação.

4.2 CODIFICAÇÃO DOS INDIVÍDUOS E DEFINIÇÃO DA POPULAÇÃO

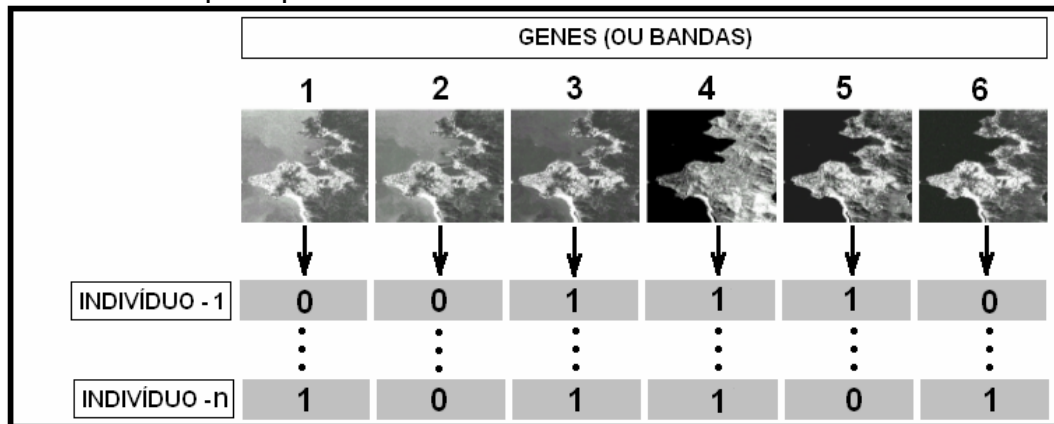
Segundo MITCHELL (1997) o uso extensivo da codificação binária é devido, acima de tudo, ao fato histórico sobre a origem dos Algoritmos Genéticos. Assim, por já ser consagrada na literatura e por adaptar-se bem ao problema em questão, esse tipo de codificação será utilizado neste trabalho.

A codificação binária, no contexto desse trabalho, representa o conjunto de bandas ótimo para a classificação. Para isso, uma cadeia de b elementos binários é usada para representar o total de b bandas. Cada elemento pode assumir dois valores: 0 (zero) ou 1 (um), onde o 0 indica que a banda correspondente não será considerada na classificação, enquanto que o valor 1 indica que a banda é útil para a classificação.

Um exemplo hipotético de indivíduo, já codificado, está ilustrado na FIGURA 12. As imagens apresentadas nessa figura são referentes às bandas 1, 2, 3, 4, 5 e 7

do sensor TM do Landsat. Pode-se observar na FIGURA 12 que apenas as bandas 3, 4 e 5 estão presentes no indivíduo 1, enquanto que as demais bandas, não pertencem. Da mesma forma, no indivíduo n , as bandas 1, 3, 4 e 6 estão contidas, e as outras duas restantes, estão ausentes. As bandas ausentes não farão parte do processo de classificação.

FIGURA 12 - Exemplo hipotético de indivíduos ou cromossomos



Após realizada a codificação e a determinação do tamanho dos indivíduos, a população será definida como um grupo de n indivíduos (FIGURA 12). A população inicial é criada de forma aleatória. Para cada gene no indivíduo escolhe-se aleatoriamente o número 1 (um) ou 0 (zero). Esse processo é repetido até que todos os genes de todos os indivíduos da população sejam preenchidos. O tamanho da população é definido pelo analista de acordo com seu interesse particular. As demais populações do processo evolutivo serão criadas iterativamente, sem a participação do analista.

4.3 TÉCNICAS DE SELEÇÃO UTILIZADAS

Foram utilizadas nesse trabalho, três das técnicas de seleção indicadas por GOLDBERG (1989), para uso nos Algoritmos Genéticos simples. Essas técnicas são: roleta, torneio e elitismo.

A técnica roleta, segundo MICHALEWICZ (1999), pode ser implementada da seguinte forma:

Considere uma população com i indivíduos:

a) calcular a aptidão F de cada indivíduo ($F(IND_i)$);

b) fazer o somatório S dos valores de aptidão de todos os indivíduos

$$S = \sum_{i=1}^{Tpop} F(IND_i), \text{ onde } Tpop \text{ é o tamanho da população;}$$

c) calcular a probabilidade de seleção P_i de cada indivíduo ($P_i = \frac{F(IND_i)}{S}$), e

d) determinar a probabilidade cumulativa q_i para cada indivíduo ($q_i = \sum_{j=1}^i P_j$).

A seleção propriamente dita, de um indivíduo i , será processada como segue:

a) gerar um número aleatório r entre 0 e 1;

b) se $r < q_1$ então selecionar o primeiro indivíduo, senão, selecionar o i -ésimo indivíduo ($IND_i (2 \leq i \leq Tpop)$), tal que $q_{i-1} < r \leq q_i$.

Existem várias formas de aplicar a técnica de seleção por *torneio*, uma delas, segundo MITCHELL (1997), é:

a) escolher aleatoriamente dois indivíduos na população;

b) definir um parâmetro l (geralmente $l = 0,75$) entre 0 e 1;

c) gerar um número aleatório ρ entre esse mesmo intervalo (0,1); e

d) se $\rho < l$, o melhor indivíduo (de maior aptidão), entre os dois selecionados no item a), é escolhido, caso contrário, o pior indivíduo (de menor aptidão) é selecionado. Esse processo é repetido $Tpop$ vezes, onde $Tpop$ é o tamanho da população.

Das três técnicas, a seleção por *elitismo* é a mais simples. A técnica por elitismo pode ser processada da seguinte forma:

a) organizar, em ordem decrescente, os valores de aptidão de todos os indivíduos, num vetor;

b) escolher um grupo com os x primeiros valores (elite), desse vetor, onde $x < Tpop$.

c) selecionar os indivíduos correspondentes a esses valores de aptidão;

d) completar a população selecionando os indivíduos por outra técnica, como exemplo, por roleta ou torneio.

4.4 OPERADORES GENÉTICOS UTILIZADOS

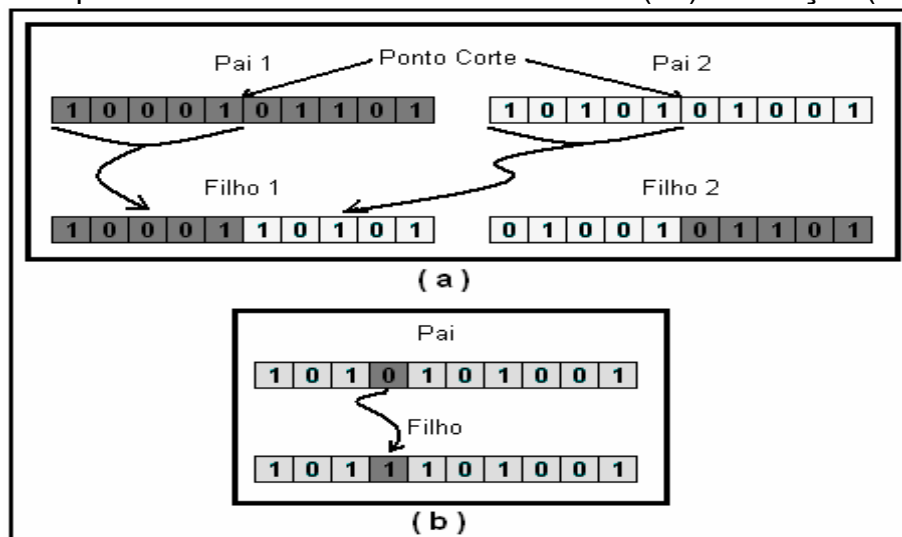
Os Operadores Genéticos, segundo GOLDBERG (1989), são a *reprodução*, *mutação* e o *cruzamento*. Esses operadores têm por função estimularem a evolução da espécie.

A *reprodução* acontece quando um determinado indivíduo é copiado para a próxima população. Para tanto, esse indivíduo deve ser apto, isto é, deve ter um valor de aptidão elevado o bastante para que ele seja selecionado dentre os outros indivíduos da população.

O *cruzamento* atua de forma que dois ou mais cromossomos pais trocam material genético entre si gerando suas proles. A *mutação* atua alterando o código genético, ou seja, um ou mais genes são modificados no cromossomo pai, escolhido aleatoriamente, gerando um filho mutante. A escolha dos cromossomos para a aplicação desses operadores é realizada de forma aleatória. A FIGURA 13 mostra com maior detalhe esses operadores.

Há uma discussão sobre a consideração da seleção como um operador, alguns autores como MITCHELL (1997) assegura ser um operador, enquanto outros, como GOLDBERG (1989) e LOPES (1999), preferem avaliá-la como uma técnica.

FIGURA 13 - Operadores de cruzamento ou *crossover* (a) e *mutação* (b)



Serão utilizados neste trabalho os três operadores propostos por GOLDBERG (1989), isto é, os operadores de *reprodução*, *mutação* e *cruzamento*. A operação de *reprodução* depende fielmente da função objetivo, ou seja, quanto

maior o valor de aptidão de um indivíduo, maior será a sua probabilidade de *reprodução*. Serão aplicados também a *mutação* e o *cruzamento simples*. O funcionamento desse último operador baseia-se na escolha de um único *ponto de corte* (ponto entre dois *locus* quaisquer), de forma aleatória, em dois cromossomos candidatos a serem pais, e posteriormente, executa a troca de genes entre esses dois pais, como mostrado na FIGURA 13. Além desses operadores serão aplicadas três técnicas de seleção: *roleta*, *torneio* e *elitismo*.

4.5 PARÂMETROS UTILIZADOS NA FUNÇÃO DE APTIDÃO

O potencial de cada indivíduo em solucionar o problema é calculado de acordo com a função de aptidão. Nesse trabalho, essa função foi desenvolvida compreendendo:

- a) o número de bandas presentes no indivíduo;
- b) a medida de similaridade por meio da divergência transformada, e
- c) a matriz de correlação referente às bandas da imagem utilizada.

O parâmetro do item (a) informará à função de aptidão o quão reduzido está o indivíduo. Enquanto que o item (b) fornecerá a separabilidade existente entre as classes de interesse, em cada banda presente no indivíduo. Por fim, o parâmetro (c) indicará a correlação existente entre as bandas presentes no indivíduo.

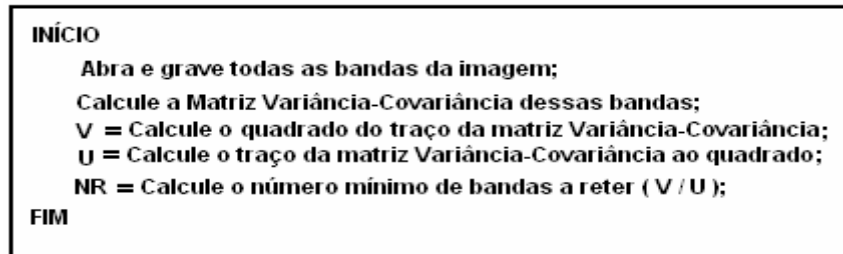
4.6 CÁLCULO DA DIMENSÃO MÍNIMA

Nesta fase buscou-se encontrar o número mínimo de bandas capaz de representar o conjunto original sem grandes prejuízos de informação. Posteriormente, esse valor foi usado na função de aptidão com o intuito de indicar o número mínimo de bandas que um indivíduo pode ter, sendo ainda considerado como apto. Isto é, o indivíduo que contém um número de bandas igual ou menor que esse valor, fornecerá uma parcela de peso pequena, ou nula, à função de aptidão.

Como citado na seção 2.5, CADIMA (2001) apresentou um método para determinar de forma automática o número mínimo de variáveis que pode substituir o espaço original sem comprometer a qualidade da informação. No caso desse trabalho, as variáveis são as bandas da imagem hiperespectral. Baseado nesse

método elaborou-se uma rotina no ambiente Matlab, como mostrado na FIGURA 14 abaixo.

FIGURA 14 - Rotina para cálculo do número mínimo de bandas a reter



A partir das equações 11 e 12, para as 220 bandas da imagem hiperespectral utilizada nesse trabalho, obteve-se o valor de $NR=1,516$. Assim, o número mínimo de bandas necessário para representar o conjunto original foi de 2 bandas ($NR \cong 2$). Isso implica que os indivíduos devem possuir três ou mais bandas.

Esse mesmo procedimento foi aplicado no conjunto com 190 bandas que também será usado nos experimentos. O valor encontrado foi quase igual aquele encontrado para as 220 bandas. Portanto, para esse conjunto de 190 bandas utilizou-se, também, 2 bandas como valor mínimo.

4.7 PROPOSTA DA FUNÇÃO DE APTIDÃO

A função de aptidão foi elaborada utilizando os parâmetros: número de bandas presentes no indivíduo; divergência transformada para medir a separabilidade entre as classes, e por fim, a correlação entre as bandas.

4.7.1 Número de Bandas Presentes no Indivíduo

A parcela fornecida à função de aptidão, pelo número de bandas escolhidas (p_{NR}), foi determinada com base no número mínimo admissível, calculado na seção 4.6. O cálculo dessa parcela foi processado de acordo com a equação 13.

$$p_{NR} = \begin{cases} 1 - \frac{Bp}{Bt} \rightarrow se \rightarrow Bp > NR \\ 0 \rightarrow se \rightarrow Bp \leq NR \end{cases} \quad (13)$$

onde, B_p é o número de bandas presentes no indivíduo; B_t é o número total de bandas (tamanho do cromossomo); NR é o número mínimo admissível de bandas.

4.7.2 Separabilidade Entre as Classes

De acordo com a equação 10, o cálculo da divergência é efetuado entre apenas duas classes. Entretanto, nesse trabalho são estudadas quatro classes. Assim, realizou-se uma combinação das quatro classes, duas a duas, e posteriormente, efetuou-se o cálculo da divergência entre todos os pares de classes gerados. Após o cálculo das divergências, escolheu-se o menor valor entre elas, e em seguida, calculou-se a divergência transformada. A aptidão de cada indivíduo, segundo essa medida de similaridade, foi computada pelo menor valor do grupo das divergências transformadas relativas ao referido conjunto de bandas. A equação 14 mostra como foi calculada essa parcela da divergência transformada fornecida à função de aptidão, para cada indivíduo. A divisão por 2000 foi realizada para normalizar o valor de p_DT entre 0 e 1.

$$p_DT_j = \frac{\text{mínimo}(DT_i)}{2000} \quad (14)$$

onde p_DT_j é a parcela de contribuição da divergência transformada à função de aptidão, para o indivíduo j ; DT_i é a divergência transformada calculada para banda i .

4.7.3 Correlação Entre as Bandas

Uma menor correlação entre duas variáveis (bandas), possibilita uma maior separabilidade das feições (classes) existentes nessas variáveis. Sendo assim, a parcela fornecida à função aptidão pela correlação entre as bandas foi estabelecida como a diferença entre a unidade (1) e o maior coeficiente de intercorrelação, em módulo, obtido a partir da matriz de correlação das bandas existentes no indivíduo. Quanto maior o valor resultante dessa diferença, maior será a contribuição da matriz de correlação para a função de aptidão. A parcela da função de aptidão fornecida pela correlação é calculada pela equação 15.

$$p_ρ = 1 - \left| \text{máximo}(A_{i,j}) \right| \quad (\forall j > i) \quad (15)$$

onde, p_{ρ} é a parcela fornecida pela correlação à função de aptidão; $A_{i,j}$ é o elemento da posição (i, j) na matriz de correlação.

O valor de aptidão para cada indivíduo foi então definido como uma média dos valores encontrados pelo cálculo da divergência transformada, da correlação entre bandas e do número de bandas presentes no indivíduo, conforme apresentado na equação 16.

$$Aptidão = \frac{(p_{\rho} + p_{DT} + p_{NR})}{3} \quad (16)$$

CAPÍTULO III - EXPERIMENTOS REALIZADOS

5. EXPERIMENTOS

Os experimentos foram divididos em cinco etapas distintas. No primeiro experimento objetivou-se tanto estimar os parâmetros necessários aos Algoritmos Genéticos, levando em consideração o tempo de processamento, como verificar a capacidade desses algoritmos em eliminar bandas com ruídos.

O segundo experimento objetivou analisar a variação dos três componentes internos da função de aptidão (número de bandas presentes no indivíduo, correlação entre bandas e a separabilidade entre as classes - divergência transformada).

O terceiro experimento foi desenvolvido para verificar o potencial dos Algoritmos Genéticos como redutores de dimensionalidade. Essa verificação foi realizada através da acurácia da classificação. A acurácia, por sua vez, foi determinada a partir dos mapas temáticos obtidos a partir das bandas selecionadas pelos Algoritmos Genéticos.

Uma comparação entre os resultados dos Algoritmos Genéticos e dos algoritmos: seleção seqüencial crescente e seleção seqüencial decrescente foi executada no quarto experimento, objetivando, portanto, avaliar o tempo de processamento e a acurácia da classificação.

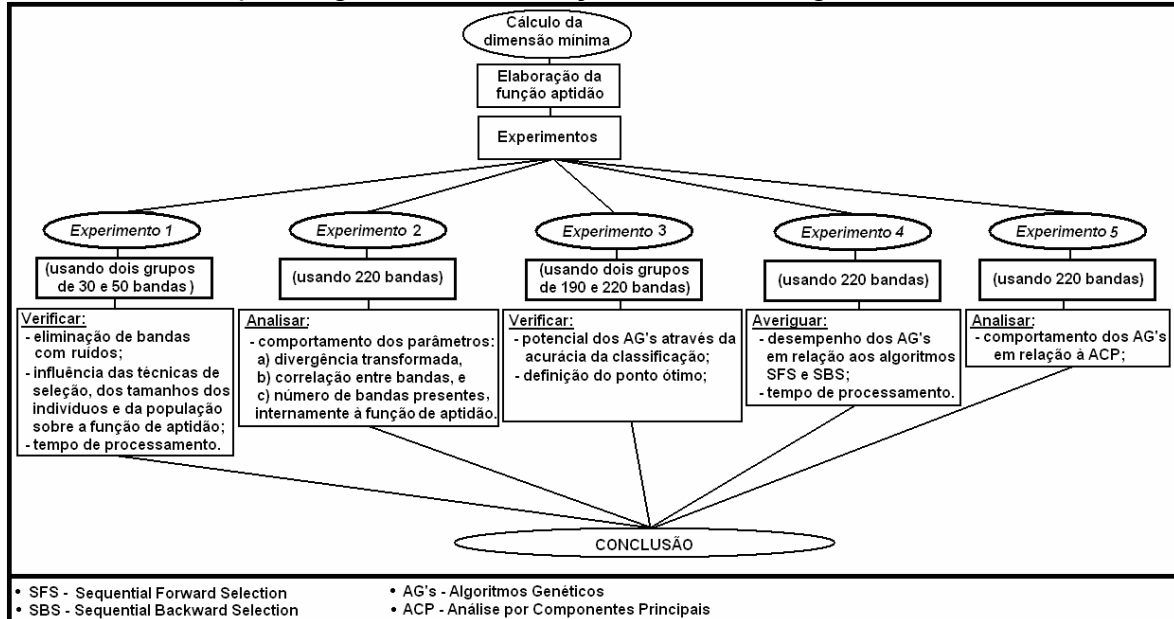
Finalmente, no quinto experimento, fez-se uma comparação, em termos de acurácia, dos resultados obtidos a partir dos Algoritmos Genéticos e da Análise por Componentes Principais. Ressalta-se que nesse último experimento o objetivo foi comparar os resultados dos Algoritmos Genéticos com os derivados de uma técnica já consagrada na literatura, Análise por Componentes Principais. Contudo, sabe-se que se trata de duas técnicas distintas: seleção de feições e extração de feições, respectivamente.

Considerou-se para todos os experimentos, as probabilidades de *mutação* e de *cruzamento*, constantes. Os valores dessas probabilidades foram os valores utilizados por MITCHELL (1997): 0,700 para a probabilidade de *cruzamento* e por

GOLDBERG (1989): 0,05 (inverso do tamanho da população utilizada $\frac{1}{20}$) para a *mutação*.

A FIGURA 15 apresenta um fluxograma desses experimentos.

FIGURA 15 – Etapas seguidas na realização da metodologia



5.1 PRIMEIRO EXPERIMENTO: DEFINIÇÃO DOS PARÂMETROS QUE SERÃO USADOS NOS ALGORITMOS GENÉTICOS

O primeiro experimento foi realizado em duas partes: na primeira parte foi analisada a variação do tempo de processamento em relação aos parâmetros: tamanho da população, número de gerações (iterações), tamanho do indivíduo e técnica de seleção. Na segunda parte foi verificada a capacidade dos Algoritmos Genéticos em eliminar bandas ruidosas.

Os parâmetros iniciais utilizados nesse experimento, tais como: tamanho dos indivíduos e da população e as quantidades de iterações (gerações) foram definidos de forma aleatória. Isto é, os melhores parâmetros para os Algoritmos Genéticos foram definidos de forma empírica, por meio desse primeiro experimento.

Para realização desse experimento, utilizou-se dois subgrupos da imagem original: um com 30 e outro com 50 bandas espectrais. A justificativa para o uso desses subgrupos é a redução do tempo de processamento e o fato de poder

selecionar as bandas que facilitem, posteriormente, a interpretação dos resultados provindos da técnica proposta. Essas bandas permitem ainda, melhor inferência na análise e validação dos resultados.

5.1.1 Análise do Tempo de Processamento em Relação aos Parâmetros Usados nos Algoritmos Genéticos

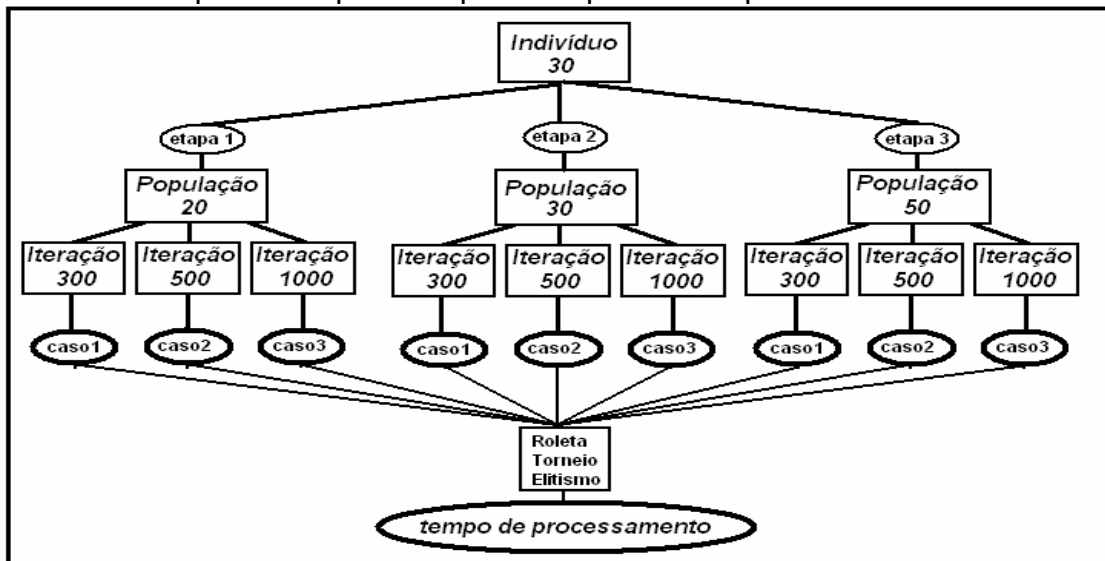
O indivíduo utilizado nessa primeira parte foi composto de 30 bandas, todas livres de ruídos, porém correlacionadas entre si. A TABELA 3 apresenta as bandas presentes nesse indivíduo.

TABELA 3 - Indivíduo com 30 bandas

azul	verde	vermelho	infravermelho próximo	infravermelh o médio
3	9	17	30	88
4	10	18	31	102
5	11	19	34	114
	12	20	35	115
	13	21	39	129
			40	130
			68	174
			69	175
				205

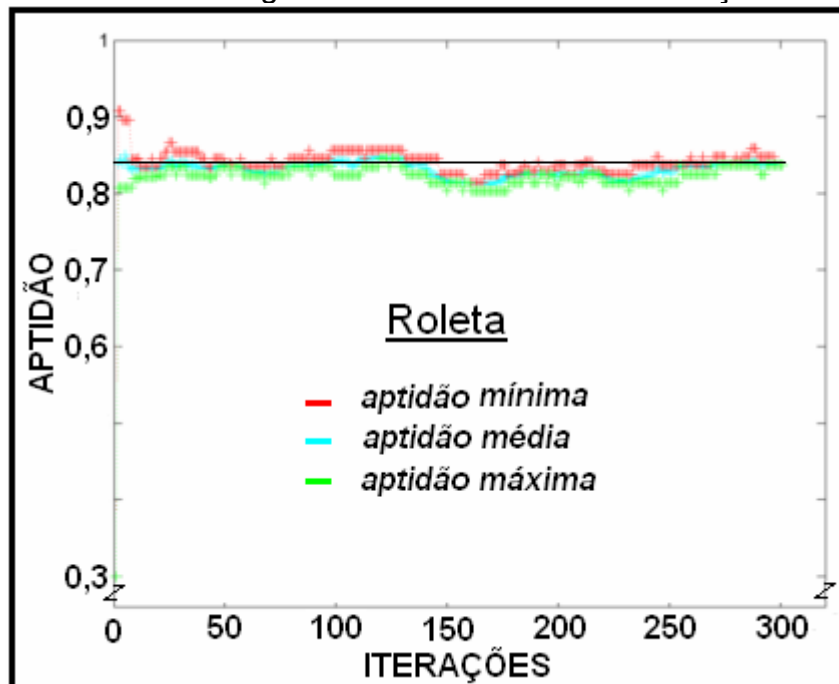
Os testes foram realizados em três etapas distintas. Na primeira etapa testou-se os Algoritmos Genéticos variando o tamanho da população em 20, 30 e 50 indivíduos. Posteriormente, para cada população, variou-se o número das gerações (iterações) em 300, 500 e 1000. Finalmente, em cada um desses grupos (casos) de gerações, alterou-se a técnica de seleção em: seleção por *torneio*, *roleta* e *elitismo*. A FIGURA 16 evidencia melhor essa primeira parte através de um esquema gráfico.

FIGURA 16 - Esquema da primeira parte do primeiro experimento



Na primeira etapa (etapa 1), percebeu-se que a curva de convergência da função de aptidão, relativa à técnica de seleção *roleta*, apresentou um declínio no decorrer das gerações, como exemplificado na FIGURA 17. Esse fenômeno foi observado nos três casos (nos três grupos de iterações: 300, 500 e 1000) dessa etapa. A ocorrência desse fenômeno deve-se, entre outros fatores, à característica intrínseca dessa técnica de seleção.

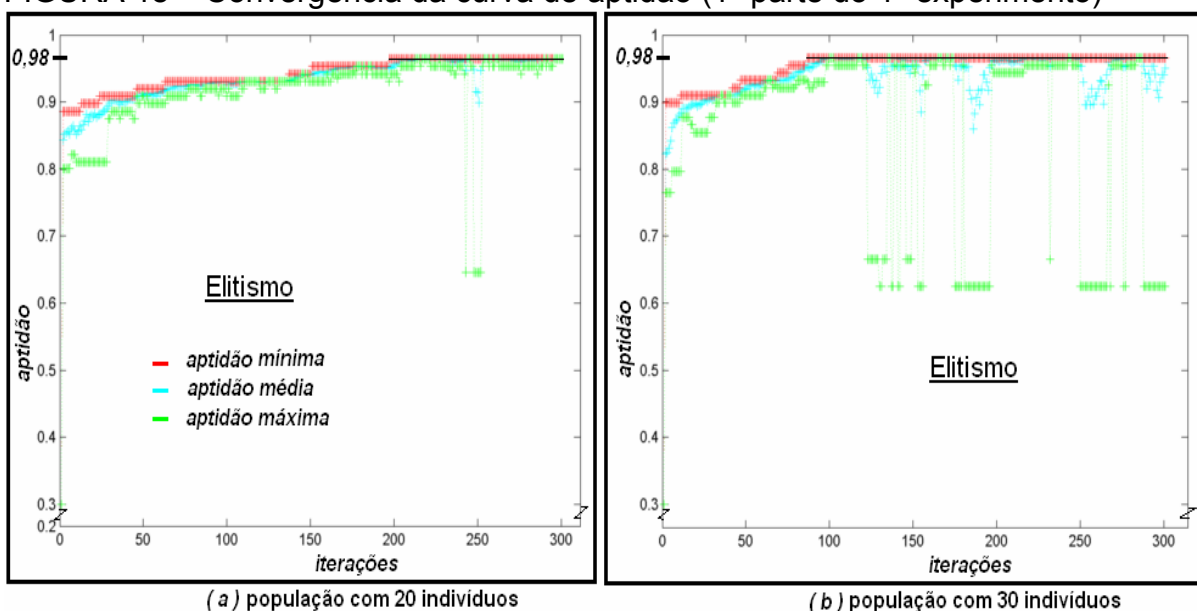
FIGURA 17 – Curva de convergência relativa à técnica de seleção *roleta*



A técnica de seleção *roleta* possui a característica de convergência prematura, isto é, ao atingir um ponto ótimo (podendo esse ponto ser um ótimo local), a curva de convergência estabiliza. A estabilização da curva de convergência e a possibilidade de seleção de indivíduos ruins (não aptos) são fatores que proporcionam esse declínio.

Analisando a etapa 2 verificou-se que ao aumentar a população, de 20 para 30 indivíduos, a curva de convergência, referente à técnica de seleção *elitismo*, estabilizou num mesmo valor de aptidão (0.9800), como ilustrado na FIGURA 18. O aumento da população proporcionou uma convergência mais rápida, isto é, na população com 20 indivíduos (FIGURA 18 a) a convergência ocorreu após as 200 iterações, contudo, na população com 30 indivíduos (FIGURA 18 b), a convergência se deu em aproximadamente 100 iterações. Isso leva a conclusão de que o aumento da população proporciona uma maior probabilidade do grupo de "elite" (grupo selecionado pela técnica *elitismo* a cada iteração) possuir maior aptidão mais rapidamente, e conseqüentemente, uma inclinação da curva de convergência mais acelerada.

FIGURA 18 – Convergência da curva de aptidão (1ª parte do 1º experimento)



O crescimento da população de 20 para 50 indivíduos não influenciou na curva de convergência. Isto é, os valores de aptidão, após estabilização da curva de convergência, para o grupo de 500 iterações, permaneceram constantes (0,9800) nas três etapas.

Analisando os resultados da técnica de seleção *torneio* verificou-se que ao aumentar a população de 20 para 30 e de 20 para 50 indivíduos, mantendo o grupo de iteração constante (500), a curva de convergência sofreu uma variação muito pequena, de 0.9680 para 0.9700. Portanto, para o grupo de 500 iterações, pode-se considerar que houve uma estabilização dessa curva num valor constante de aptidão igual a 0.9700.

Na população de 20 indivíduos com a técnica de seleção *torneio*, o início da estabilização da curva de convergência ocorreu após as 150 iterações, nos três casos. Porém, para a população de 30 indivíduos, a estabilização ocorreu em aproximadamente 250 iterações. O aumento do tamanho da população implica num maior número de indivíduos a serem selecionados. Portanto, como a técnica de seleção por *torneio* possibilita a escolha do melhor ou do pior indivíduo, a ocorrência do pior indivíduo pode aumentar, tendo em vista que o número de indivíduos aumentou. Um maior número de indivíduos pouco aptos numa população acarreta, obviamente, um retardo no início da estabilização da curva de convergência.

Na segunda etapa (etapa 2), a técnica de seleção *roleta* falhou novamente. Isto é, com a variação dos grupos de iterações em: 300, 500 e 1000, a curva de convergência, em cada um dos três casos, apresentou um declínio.

Os resultados obtidos a partir da etapa 3 (população com 50 indivíduos) permitiram verificar que a variação do número de iterações, não influenciou de forma considerável no processo de convergência da curva de aptidão. A variação da aptidão máxima de uma população final para outra foi muito pequena, de 0,980 para 0,984. Observou-se ainda, nessa etapa, que igualmente às outras duas etapas, a convergência referente à técnica de seleção *roleta*, falhou, ao provocar um declínio na curva de convergência.

A TABELA 4 apresenta os valores máximos de aptidão obtidos a partir das três etapas.

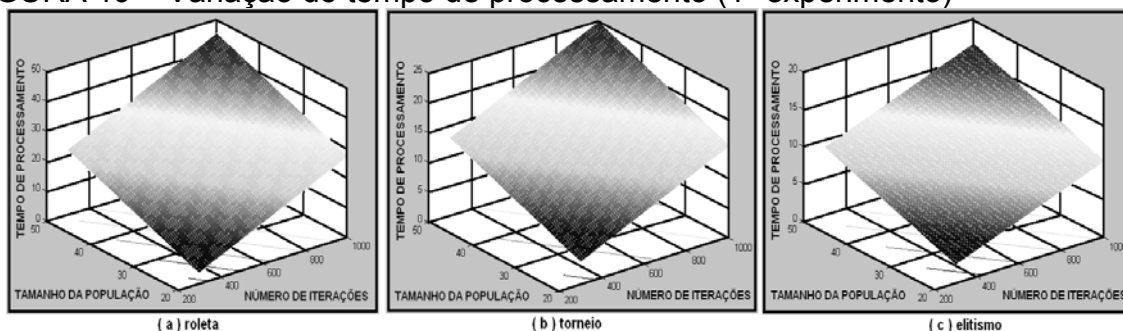
TABELA 4 – Valores máximos de aptidão (indivíduo com 30 bandas)

	ITERAÇÕES	ROLETA	TORNEIO	ELITISMO
POPULAÇÃO 20	300	0,8962	0,9558	0,9642
	500	0,8635	0,9633	0,9610
	1000	0,8880	0,9582	0,9658
POPULAÇÃO 30	300	0,8809	0,9517	0,9658
	500	0,9087	0,9597	0,9611
	1000	0,8903	0,9610	0,9642
POPULAÇÃO 50	300	0,8981	0,9633	0,9598
	500	0,8643	0,9570	0,9658
	1000	0,8864	0,9658	0,9620

Analisando os dados apresentados na TABELA 4, Verificou-se que a técnica de seleção por elitismo proporcionou maiores valores de aptidão em 6 dos 9 casos apresentados. A técnica de seleção por torneio superou a técnica de seleção por elitismo em 3 casos: no grupo de 500 iterações e população com vinte indivíduos, e nos grupos de 300 e 1000 iterações, ambos com a população de 50 indivíduos. A técnica de seleção por roleta apresentou valores inferiores, em relação às duas outras técnicas, em todos os casos.

Uma análise de regressão linear múltipla foi realizada no intuito de verificar a variação do tempo de processamento em função das variações do tamanho da população e do número de iterações. A análise por regressão foi realizada sobre os resultados obtidos nas três etapas da primeira parte desse primeiro experimento, isto é, foi realizada sobre os resultados de cada uma das três técnicas de seleção: *torneio*, *roleta* e *elitismo*. A FIGURA 19 ilustra os resultados obtidos a partir dessa análise.

FIGURA 19 – Variação do tempo de processamento (1º experimento)



Os planos ilustrados nos três gráficos da FIGURA 19 permitiram concluir que o aumento do tempo de processamento é mais acentuado com a ampliação da população do que com aumento do número de iterações.

Apesar de apresentarem um comportamento semelhante, os três gráficos mostraram a ocorrência de um menor tempo de processamento para a técnica de seleção *elitismo*, e um maior tempo para a técnica de seleção *roleta*, conforme apresentado na TABELA 5. Nessa tabela estão representados todos os resultados advindos das três técnicas de seleção. Pode ser verificada, na última coluna dessa tabela, uma maior eficiência da técnica de seleção *elitismo*.

TABELA 5 – Tempo de processamento em segundos (1ª parte)

	ITERAÇÕES	ROLETA	TORNEIO	ELITISMO
POPULAÇÃO 20	300	8,079	5,25	4,016
	500	12,844	7,656	5,266
	1000	19,797	13,156	8,281
POPULAÇÃO 30	300	10,235	8,343	4,812
	500	16,703	10,438	6,938
	1000	32,938	17,61	12,406
POPULAÇÃO 50	300	14,844	10,031	7,609
	500	28,063	15,797	10,687
	1000	51,828	28,078	18,813

Com o aumento do indivíduo para 50 bandas o tempo de processamento, como mostrado na TABELA 6, apresentou comportamento similar àquele sucedido

com o indivíduo de 30 bandas. Isto é, foi demandado um maior tempo de processamento no uso da técnica de seleção *roleta* e um menor ao usar a técnica de seleção *elitismo*.

TABELA 6 – Tempo de processamento em segundos (conjunto com ruídos)

	ROLETA	TORNEIO	ELITISMO
300 ITERAÇÕES	20,2350	12,2500	8,3120
500 ITERAÇÕES	36,7970	17,1410	13,3430
1000 ITERAÇÕES	92,7970	28,5000	17,8440

O tempo de processamento demandado em cada técnica de seleção cresceu consideravelmente com o aumento do tamanho do indivíduo e do número de iterações. Por exemplo, para a variação de 500 para 1000 iterações e técnica de seleção *roleta*, a variação do tempo para o indivíduo de 30 bandas foi de 16,235, enquanto para o indivíduo de 50 bandas, o tempo foi de 56,000 segundos.

A TABELA 7 apresenta os valores de aptidão referentes ao indivíduo de 50 bandas.

TABELA 7 – Valores máximos de aptidão (indivíduo com 50 bandas)

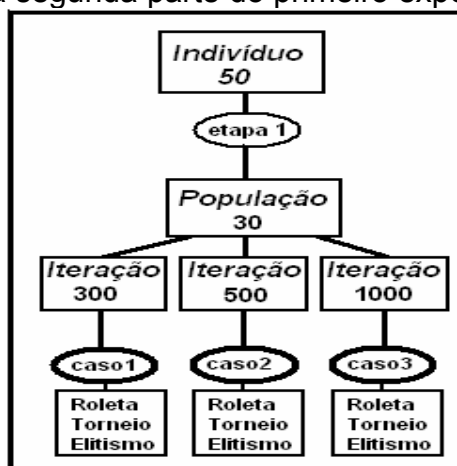
	ITERAÇÕES	ROLETA	TORNEIO	ELITISMO
POPULAÇÃO 30	300	0.8691	0.9525	0.9522
	500	0.8589	0.9710	0.9784
	1000	0.8324	0.9799	0.9688

Nas tabelas 4 e 7 pôde ser observado que ao variar o tamanho do indivíduo, de 30 para 50 bandas, a técnica de seleção por torneio apresentou maiores valores de aptidão, nos três casos: 300, 500 e 1000 iterações. A técnica de seleção por roleta, se comparada com as outras duas técnicas, continuou apresentando menores valores de aptidão.

5.1.2 Capacidade dos Algoritmos Genéticos em Eliminar Bandas Ruidosas

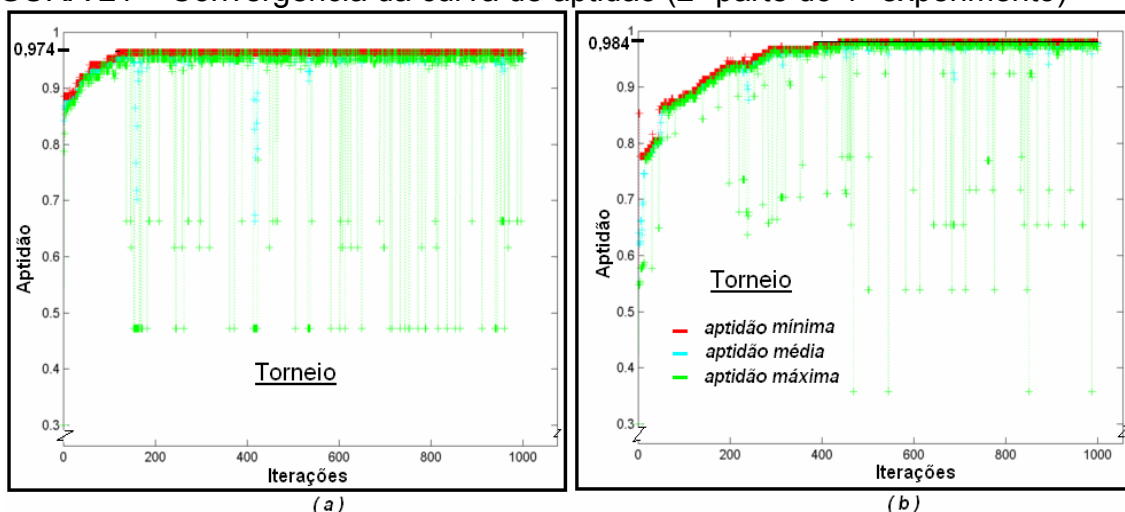
Com o objetivo de avaliar a eliminação de bandas com ruídos, usando os Algoritmos Genéticos, utilizou-se o indivíduo com 50 bandas, o qual possui 10 bandas com ruídos. A FIGURA 20 evidencia o procedimento adotado para realização desta segunda parte do primeiro experimento.

FIGURA 20 - Esquema da segunda parte do primeiro experimento



O aumento do indivíduo provocou uma variação na inclinação da curva de convergência. Pode-se verificar a alteração no comportamento dessa curva na FIGURA 21. O gráfico apresentado na FIGURA 21 (a) corresponde à curva de convergência da função de aptidão para a população com 30 indivíduos sem ruídos. O gráfico da FIGURA 21 (b) representa o comportamento da aptidão para a população com 50 indivíduos, com ruídos. Ambos os casos referem-se ao grupo de 1000 iterações.

FIGURA 21 – Convergência da curva de aptidão (2ª parte do 1º experimento)



O aumento do indivíduo de 30 para 50 bandas, cresce a probabilidade de selecionar um grupo de maior aptidão, pois um maior número de bandas aumenta as possibilidades de combinação entre essas bandas.

Depois de realizadas as iterações (300, 500 e 1000) em cada um dos casos, as bandas selecionadas como ótimas, para uma posterior classificação, foram

aquelas apresentadas na TABELA 8. Devido ao grande número de bandas escolhidas pela técnica de seleção *roleta*, elas não foram apresentadas nessa tabela. Contudo, foi realizado um estudo sobre os resultados dessa técnica (*roleta*), onde pôde ser observado que além de repetir as falhas ocorridas na primeira parte desse experimento, ocorreu ainda a indesejada seleção de bandas ruidosas, nos três casos (300, 500 e 1000 iterações). Os valores em negrito, mostrados na TABELA 8, referem-se às bandas ruidosas que não foram eliminadas pelos Algoritmos Genéticos.

TABELA 8 – Bandas selecionadas pelos Algoritmos Genéticos

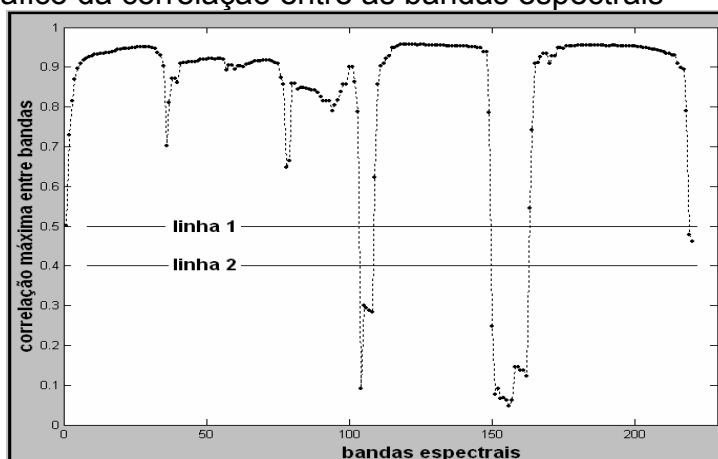
TORNEIO	ELITISMO
2-3-64-115- 151 -175-176	28-30-42-44-86-115-116
28-86-115-205	36- 153 -177
137- 154 -177	38-102-207

Devido à falha ocorrida na convergência da função de aptidão, em todas as etapas, preferiu-se por eliminar a técnica de seleção *roleta* dos próximos experimentos.

Outra mudança foi introduzida na definição da função de aptidão objetivando corrigir a preservação de bandas com ruído. Essa mudança foi realizada com a inserção de um fator, na função de aptidão, que mede o índice de ruído nas bandas. A definição desse fator foi realizada por meio de estudos sobre a correlação entre as bandas. Uma segunda opção, de uso desse fator, foi verificada inserindo-o no processo de criação da população inicial. Isto é, os indivíduos da população inicial foram condicionados à não apresentação de bandas ruidosas. Essa segunda opção, além de apresentar resultados compatíveis com a primeira, exigiu menos tempo de processamento. Portanto, a segunda opção foi utilizada nessa pesquisa.

Sendo C a matriz de correlação entre as 220 bandas da imagem hiperespectral utilizada, e A a matriz absoluta de C , extraiu-se da matriz A , o vetor m , correspondente aos valores máximos de suas colunas. Posteriormente plotou-se os valores contidos no vetor m , como ilustrado na FIGURA 22. Observando essa figura, verificou-se que existem oscilações bruscas (picos e vales) no gráfico ilustrado. Analisando esse gráfico mais detalhadamente observou-se que as regiões dos vales, ou seja, as regiões de baixos valores máximos de correlação são exatamente as regiões correspondentes às bandas com ruídos.

FIGURA 22 – Gráfico da correlação entre as bandas espectrais



Na FIGURA 22 foram traçadas duas linhas horizontais (linha 1 e 2) para exemplificar o procedimento adotado na determinação do fator de avaliação do ruído nas bandas. Essas linhas caracterizam o ponto abaixo do qual todas as bandas presentes são ruidosas, ou seja, é o limiar que define quais bandas serão eliminadas por serem altamente ruidosas. Cada analista poderá escolher uma linha (ou valor), como critério ou limite de decisão. Especificamente nesse trabalho, o critério adotado foi o representado pela linha 1. A adoção desse critério foi baseada numa análise visual sobre a imagem original, isto é, verificou-se que as bandas que têm correlação menor ou igual a 0,5 são totalmente ruidosas.

Essa informação foi incluída nos Algoritmos genéticos tornando possível caracterizar uma banda como ruidosa ou livre de ruídos. Assim, se uma banda possui alto índice de ruído e estiver presente num determinado indivíduo, ele será classificado com inábil ou não apto, pelos Algoritmos Genéticos.

Após realizadas estas mudanças nos Algoritmos Genéticos, novos testes foram realizados e verificou-se que os problemas apontados foram corrigidos. Isto é, as bandas ruidosas foram totalmente excluídas dos indivíduos. A FIGURA 23 mostra com detalhes a convergência da função de aptidão, destacando o momento em que os indivíduos se tornam livres das bandas com ruídos. Isso pode ser observado no instante em que a curva sofre uma inclinação acentuada, após as 500 iterações. Percebe-se nesse momento que os indivíduos passam a receber uma ponderação ou valor de aptidão muito maior, passando de um valor de aproximadamente 0,500 para um valor maior que 0,800. A partir das 1000 iterações a curva de convergência

começa a estabilizar. Portanto pode-se afirmar, com maior certeza, que a partir desse ponto os indivíduos estão totalmente livre de bandas com ruídos.

FIGURA 23 - Convergência da função de aptidão



Devido a grande dimensão da matriz de correlação, ela não será apresentada nesse trabalho.

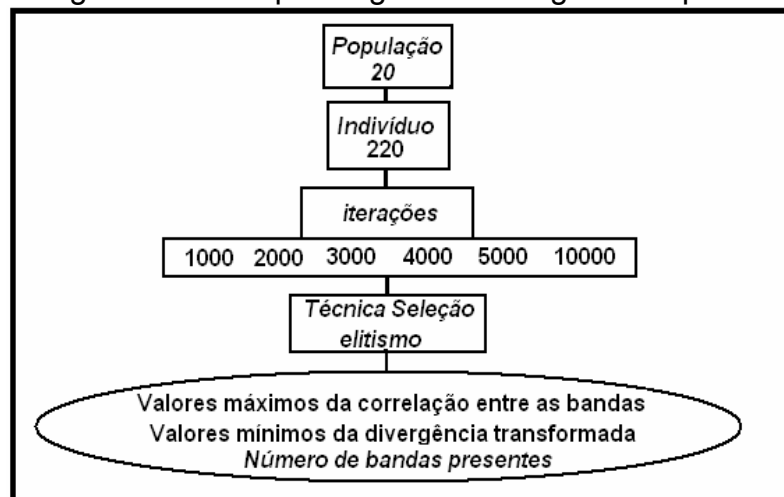
5.2 SEGUNDO EXPERIMENTO: ANÁLISE DO COMPORTAMENTO DOS PARÂMETROS DEFINIDOS PARA A FUNÇÃO DE APTIDÃO

Nesse experimento buscou-se analisar a função de aptidão através do comportamento de seus parâmetros internos, apresentados na seção 4.7. Os parâmetros relacionados com a divergência transformada (separabilidade entre as classes) e com a correlação entre as bandas foram determinados a partir de seus valores mínimos e máximos, respectivamente.

Para a realização desse experimento utilizou-se: o indivíduo com 220 bandas e a técnica de seleção *elitismo*. O número de iterações foi variado de: 1000, 2000, 3000, 4000, 5000 e 10000. A opção pelo indivíduo com 220 bandas deve-se ao fato de que esse indivíduo possui todas as bandas da imagem estudada, e ainda, por possuir bandas tanto com ruídos como livre de ruídos. Como a técnica de seleção não tem influencia direta na análise realizada nesse experimento, poderia ser escolhida qualquer uma das duas: *torneio* ou *elitismo*. As quantidades de iterações foram definidas de forma que permitissem uma análise pormenorizada das variações dos parâmetros da função de aptidão. Foram realizados outros testes com iterações menores que 1000, porém, em todos eles, o número de bandas selecionadas foi muito grande, isto é, ocorreu uma pequena redução da dimensão. A

população foi definida com 20 indivíduos, baseando-se nos resultados do primeiro experimento, ou seja, objetivando tanto um menor tempo de processamento como menor número de iterações. A FIGURA 24 apresenta o fluxograma seguido nesse experimento.

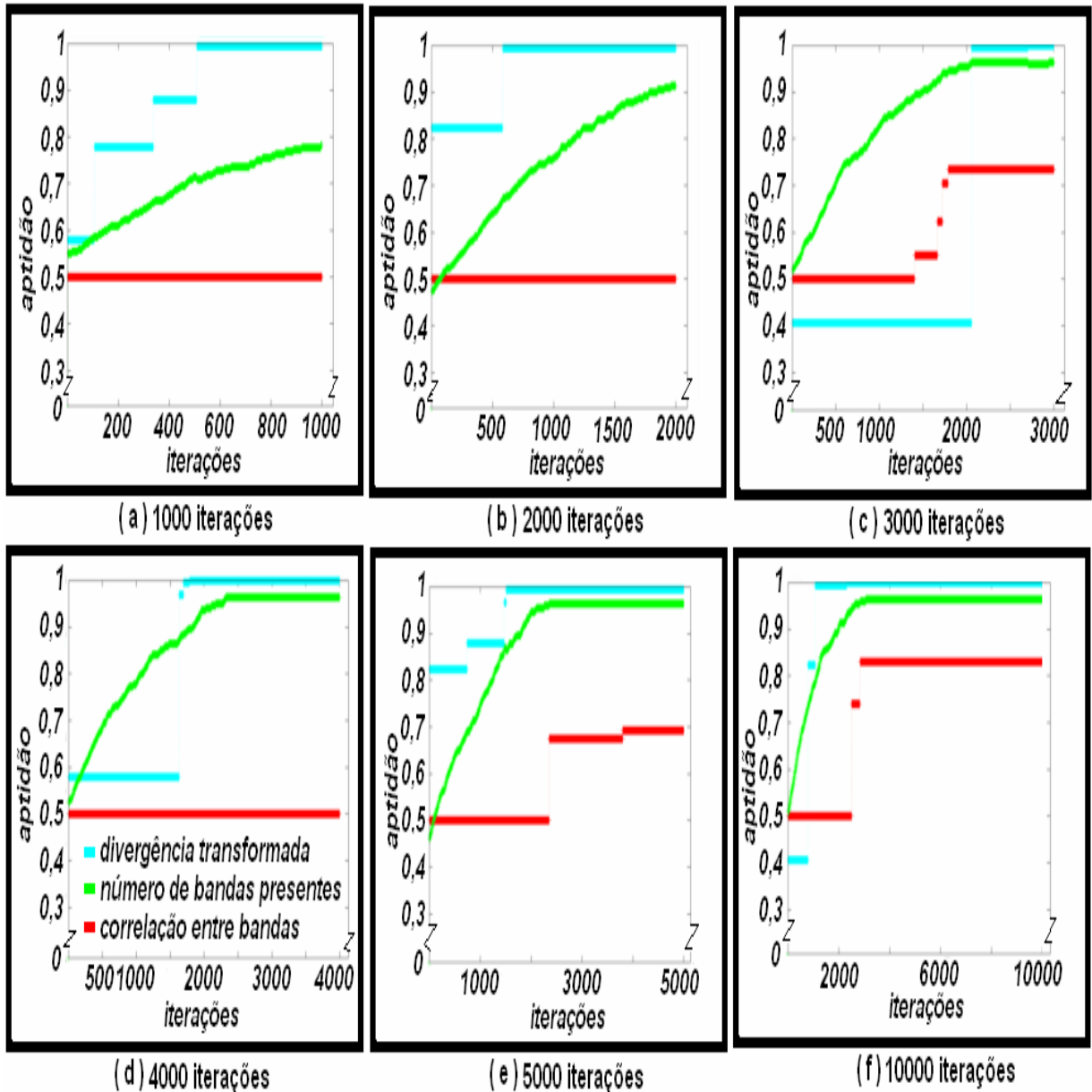
FIGURA 24 - Fluxograma das etapas seguidas no segundo experimento



As curvas relativas aos três parâmetros, derivadas desse experimento, foram plotadas e apresentadas na FIGURA 25.

A partir dos resultados obtidos nesse experimento verificou-se que a eliminação das bandas com ruído ocorreu em todos os grupos de iterações (1000, 2000, 3000, 4000, 5000 e 10000). Examinando as curvas apresentadas na FIGURA 25, referentes aos seis grupos de iterações, percebeu-se que houve um aumento acentuado nas curvas do parâmetro relativo à divergência transformada. A evolução dessa curva contribui consideravelmente para o desenvolvimento do processo evolutivo, isto é, os indivíduos tornam-se mais aptos, e conseqüentemente, com maior valor de aptidão.

FIGURA 25 - Curvas de convergência dos parâmetros da função de aptidão



O parâmetro relativo à divergência transformada foi quem mais contribuiu, nas últimas iterações, para a função de aptidão. Todavia, no início do processo evolutivo, nos grupos de iterações: 3000, 4000 e 10000, esse parâmetro foi superado por pelo menos um dos outros dois parâmetros. Outro fenômeno ocorrido com o parâmetro relacionado à divergência transformada foi o alcance do ponto ótimo em todos os experimentos, isto é, as bandas selecionadas tiveram divergência transformada igual a 2000, e, portanto, p_{DT} igual a 1, como discutido na seção 4.7.2.

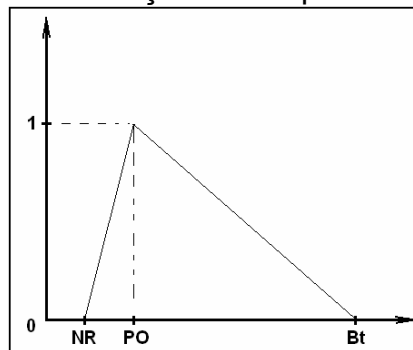
A partir do grupo de 4000 iterações, percebeu-se que a curva relacionada ao número de bandas presentes no indivíduo, permanece estabilizada. Contudo, até o grupo de 2000 iterações, existiu uma variação de crescimento na inclinação dessa curva. Esse fenômeno confirma a eliminação de bandas dos indivíduos (redução da dimensão) até um ponto designado como ótimo. A equação 17 mostra a função linear utilizada para definição desse ponto ótimo. Por exemplo, se usar $PO = 10$ o número de bandas considerado como ótimo será de 10 bandas, se $PO = 15$, esse número será de 15 bandas, e assim sucessivamente. A FIGURA 26 mostra o gráfico que define o ponto ótimo.

$$0 \rightarrow se \rightarrow Bp \leq NR$$

$$p_{NR} = \begin{cases} \frac{(Bp - NR)}{(PO - NR)} \rightarrow se \rightarrow NR < Bp < PO \\ -\frac{(Bp - Bt)}{(Bt - PO)} \rightarrow se \rightarrow Bp \geq PO \end{cases} \quad (17)$$

Onde PO é o ponto ótimo determinado pelo usuário.

FIGURA 26 – Função para determinação de um ponto ótimo



Outra observação importante foi o crescimento da curva de convergência do parâmetro relacionado à correlação entre as bandas. Nos experimentos com 1000 e 2000 iterações, a parcela fornecida à função de aptidão permaneceu constante, em 0,5. Entretanto, a partir de 3000 iterações, é verificada uma elevação brusca e de grandeza considerável na curva referente a esse parâmetro.

A partir da elevação dessa curva pode-se concluir que os indivíduos desses grupos tiveram maior valor de aptidão uma vez que os outros parâmetros também cresceram.

Numa análise geral sobre os gráficos da FIGURA 25, observou-se que grupos inferiores a 2000 iterações não são suficientes para um melhor aproveitamento do parâmetro relacionado à correlação entre as bandas. Contudo, um aumento exagerado do número de iterações torna-se dispendioso, computacionalmente. Pôde-se observar ainda a ocorrência de um caso em que esse parâmetro não contribuiu com a função de aptidão, como sucedido no grupo de 4000 iterações. Esse fato reforça a discussão acerca do cuidado a ser tomado com a variação do número de iterações, pois o grupo de 4000 iterações é um número relativamente grande. Isso desperta também a necessidade de repetição do experimento, sob as mesmas configurações dos Algoritmos Genéticos, para verificação da ocorrência do mesmo evento, uma vez que se trata de algoritmos estocásticos e de um processo evolutivo de natureza aleatória.

5.3 TERCEIRO EXPERIMENTO: VERIFICAÇÃO DA ACURÁCIA DO PRODUTO GERADO USANDO OS ALGORITMOS GENÉTICOS

No intuito de avaliar a variação da acurácia da classificação, ao alterar o número de bandas, foram definidos 16 diferentes grupos de bandas a serem selecionados pelos Algoritmos Genéticos. Cada grupo difere do seu subsequente em 3 bandas, sendo que o primeiro grupo tem 3, e o último, 48 bandas.

Utilizou-se nesse experimento o indivíduo de 220 bandas, a população de 20 indivíduos e a técnica de seleção por *elitismo*. Depois de submetida essa configuração aos Algoritmos Genéticos, condicionada à quantidade de bandas a serem selecionadas, obteve-se os 16 grupos de bandas desejados. Para a obtenção de cada grupo foi realizado um conjunto de 4000 iterações.

Para um melhor entendimento, esse experimento foi dividido em duas partes. Na primeira parte é definido o grupo de bandas que apresenta maior acurácia, e são apresentados ainda, os resultados com discussão acerca da classificação das bandas selecionadas pelos Algoritmos Genéticos. Na segunda parte é apresentada uma discussão sobre as bandas selecionadas pelos Algoritmos Genéticos, isto é, sobre a distribuição dessas bandas ao longo do espectro eletromagnético.

5.3.1 Análise da Acurácia da Classificação das Bandas Seleccionadas pelos Algoritmos Genéticos

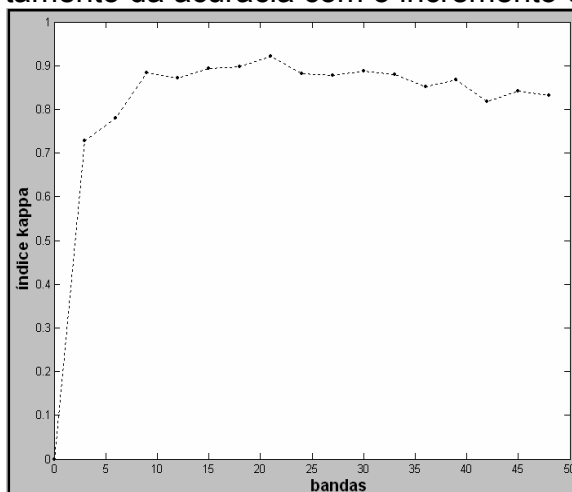
Utilizando os grupos de bandas seleccionadas, processou-se a classificação pelo método máxima verossimilhança. A acurácia dos mapas temáticos resultantes foi analisada através do índice *kappa*. Antes da análise sobre os valores desses índices (*kappa*), realizou-se uma inferência estatística, a partir da qual foi constatado que todos os valores *kappa*, mostrados na TABELA 9, estão dentro de um intervalo de confiança de 95%.

TABELA 9 – Valores de acurácia dos mapas temáticos

Grupos	Número de Bandas	Acurácia (<i>kappa</i>)
1	3	0,7290
2	6	0,7791
3	9	0,8837
4	12	0,8710
5	15	0,8931
6	18	0,8973
7	21	0,9218
8	24	0,8805
9	27	0,8776
10	30	0,8871
11	33	0,8789
12	36	0,8524
13	39	0,8678
14	42	0,8185
15	45	0,8411
16	48	0,8312

Os valores de acurácia mostrados na TABELA 9 apresentaram um crescimento desde o grupo de 3 bandas até o grupo de 21 bandas. A partir desse grupo (21 bandas) os valores de acurácia começaram a decrescer. Esse fenômeno já foi discutido anteriormente, é o bem conhecido fenômeno de *Hughes*. Portanto, para esses grupos de bandas, pode-se dizer que o de 21 bandas é o ponto ótimo da curva da acurácia. A FIGURA 27 ilustra o gráfico dos grupos versus valores de *kappa*.

FIGURA 27 – Comportamento da acurácia com o incremento de variáveis



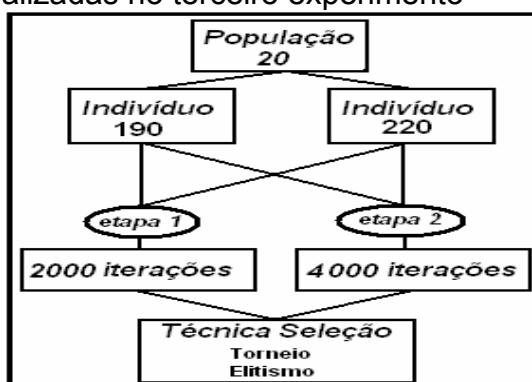
De posse da informação do grupo ótimo, iniciou-se a segunda parte do terceiro experimento. Dois indivíduos, um com 190 e outro com 220 bandas, foram utilizados. O conjunto de 220 bandas inclui bandas ruidosas. O outro conjunto é composto de 190 bandas sem ruídos.

A partir desses dois indivíduos aplicou-se os Algoritmos Genéticos da seguinte forma: para cada um dos indivíduos de 190 e 220 bandas, manteve-se a população de tamanho 20, realizando seguidamente duas etapas: a primeira refere-se à aplicação de 2000 iterações, enquanto que na segunda foram realizadas 4000 iterações. Por fim, ainda foram aplicadas as duas técnicas de seleção: *torneio* e *elitismo*, em cada uma das etapas.

A escolha dos grupos de 2000 e 4000 iterações foi baseada nos resultados do experimento anterior. Tanto na etapa 1 como na etapa 2 foram selecionados, pelos Algoritmos Genéticos, grupos de vinte e uma bandas como ponto ótimo.

A FIGURA 28 ilustra o esquema seguido na realização das etapas desenvolvidas nesse experimento. Em ambas as etapas foram consideradas constantes as probabilidades de mutação e *cruzamento*, com valores iguais aos utilizados nos experimentos anteriores.

FIGURA 28 - Etapas realizadas no terceiro experimento



A acurácia dos mapas temáticos gerados nessa etapa foi avaliada através do índice *kappa*, da acurácia global e da acurácia do ponto de vista do usuário. A TABELA 10 apresenta os resultados dessa avaliação.

TABELA 10 - Avaliação da acurácia dos mapas temáticos (2000 iterações)

	índice kappa		acurácia global		acurácia usuário	
	torneio	elitismo	torneio	elitismo	torneio	elitismo
190 bandas	0,8743	0,9028	0,9018	0,9242	0,9025	0,9234
220 bandas	0,8833	0,8641	0,9093	0,8943	0,9105	0,8961

A TABELA 10 mostra que ao aumentar o indivíduo, a acurácia referente à técnica de seleção *elitismo* cresceu, contudo, a acurácia relativa à técnica de seleção *torneio*, sofreu um leve declínio (0,009). Com o aumento do indivíduo, ou seja, com a inserção de novas bandas, que nesse caso são em maioria, bandas ruidosas e a conservação do número de iterações constante (2000), a técnica de seleção *elitismo* mostrou-se mais eficiente que a técnica *torneio* na localização do ponto ótimo, isto é, a curva referente ao parâmetro, números de bandas presentes no indivíduo, atingiu seu máximo em um menor número de gerações. A característica do *elitismo* em preservar um grupo dos indivíduos mais aptos pode ter sido o fator causador dessa maior eficiência.

O indivíduo com 190 bandas apresentou maior acurácia da classificação sobre o indivíduo de 220 bandas, na técnica de seleção *elitismo*, como mostrado na segunda coluna da TABELA 10.

Na segunda etapa foi verificada a acurácia dos mapas temáticos. Foi realizada uma inferência estatística sobre os valores *kappa* para verificação de sua confiabilidade.

A TABELA 11 apresenta os valores dos índices *kappa*, acurácia global e acurácia do ponto de vista do usuário, calculados sobre todos os mapas temáticos resultantes dos Algoritmos Genéticos, na etapa 2.

TABELA 11 – Valores do índice *kappa* da segunda etapa (4000 iterações)

	índice <i>kappa</i>		acurácia global		acurácia usuário	
	torneio	elitismo	torneio	elitismo	torneio	elitismo
190 bandas	0,8875	0,8957	0,9125	0,9189	0,9114	0,9201
220 bandas	0,8822	0,8876	0,9082	0,9125	0,9056	0,9091

Diferentemente da etapa 1, a acurácia, referente à técnica de seleção torneio, diminuiu em 0,005 com o aumento do tamanho do indivíduo, ou seja, permaneceu praticamente constante.

A técnica de seleção por *elitismo* apresentou maior acurácia que a técnica de seleção por torneio, em ambos os casos, com os indivíduos de 190 e 220 bandas.

Na passagem da etapa 1 para a etapa 2, como mostrado na TABELA 12, a acurácia referente ao indivíduo de 190 bandas na técnica de seleção *elitismo* e ao indivíduo de 220 bandas na técnica de seleção *torneio* permaneceu praticamente constante, com diferenças de 0,0011 e 0,0071, respectivamente. Todavia, nos outros dois casos houve maior crescimento da acurácia. Esse fenômeno mostrou que dependendo da técnica de seleção e do tamanho do indivíduo é preciso dosar o número de iterações a ser realizado, pois um número regulado de iterações, além de melhorar o tempo de processamento, produz acurácia, no mapa final, compatível àquelas advindas de maiores iterações.

TABELA 12 – Módulo da diferença entre os índices *kappa* das duas etapas

	TORNEIO	ELITISMO
Grupo 190 bandas	0,0132	0,0071
Grupo 220 bandas	0,0011	0,0235

Obviamente que o fenômeno discutido no parágrafo anterior é na realidade uma tendência, pois, por se tratar de algoritmos estocásticos, poderão ocorrer casos contrários, isto é, um grande número de iterações poderá proporcionar maior valor de acurácia onde todos os parâmetros da função de aptidão estejam contribuindo com seus valores ótimos, porém, de acordo com os experimentos anteriores, o tempo de processamento será sempre maior.

5.3.2 Verificação das Bandas Seleccionadas pelos Algoritmos Genéticos

As bandas seleccionadas pelos Algoritmos Genéticos apresentaram distribuição ao longo do espectro eletromagnético como mostrado a TABELA 13.

Sabe-se do sensoriamento remoto que a vegetação apresenta alta reflectância na região do infravermelho próximo, e relativamente, baixa reflectância no infravermelho médio. Contudo, a feição solo apresenta alta reflectância no infravermelho médio. O fato das classes utilizadas nesse estudo serem de culturas de milho e soja recém plantadas, ou ainda, sem cultivo, a maior parte da energia refletida refere-se ao solo. A classe de vegetação nativa, porém, cobre todo o solo. Nesse contexto, a melhor região para separabilidade dessas classes seria no infravermelho médio. Entretanto, devido a grande semelhança espectral das classes de culturas de milho e soja, a separabilidade só é possível em pequenas faixas descontínuas ao longo de todo o espectro eletromagnético, isto é, no visível, infravermelho próximo e médio.

TABELA 13 – Distribuição no espectro eletromagnético das bandas escolhidas

		2000 iterações			4000 iterações		
		Visível	IVP	IVM	Visível	IVP	IVM
190 bandas	<i>elitismo</i>	5	4	12	2	11	8
	<i>torneio</i>	0	10	11	5	7	9
220 bandas	<i>elitismo</i>	3	4	14	6	3	12
	<i>torneio</i>	4	7	10	3	3	15

IVP - Infravermelho Próximo
IVM - Infravermelho Médio

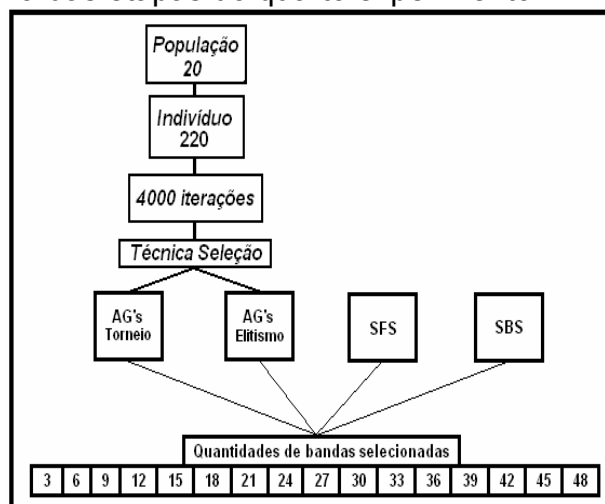
Devido a forma de busca aleatória no seu processo de atuação, os Algoritmos Genéticos apresentaram como aptos para a solução do problema, dado que existe essa descontinuidade das possíveis soluções.

5.4 QUARTO EXPERIMENTO: AVERIGUAÇÃO DO DESEMPENHO DOS ALGORITMOS GENÉTICOS EM RELAÇÃO AOS ALGORITMOS: SELEÇÃO SEQÜENCIAL CRESCENTE E SELEÇÃO SEQÜENCIAL DECRESCENTE

O quarto experimento foi estruturado conforme o fluxograma apresentado na FIGURA 29. Nesse experimento buscou-se apresentar uma comparação dos resultados obtidos a partir dos Algoritmos Genéticos, com aqueles advindos dos dois

algoritmos tradicionais. Esses últimos são do tipo seleção de feições: Seleção Seqüencial Crescente e Seleção Seqüencial Decrescente.

FIGURA 29 - Esquema das etapas do quarto experimento



Nesse experimento optou-se também por conservar a população com 20 indivíduos e usar um grupo de 4000 iterações. Posteriormente foram aplicadas as duas técnicas de seleção: *torneio* e *elitismo*. Finalmente, os Algoritmos Genéticos, juntamente com os algoritmos Seleção Seqüencial Crescente e Seleção Seqüencial Decrescente foram aplicados em 16 grupos de bandas, como ilustra a FIGURA 29.

A TABELA 14 mostra o tempo exigido pelos Algoritmos Genéticos referente às duas técnicas de seleção, e ainda, o tempo demandado pelos algoritmos Seleção Seqüencial Crescente e Seleção Seqüencial Decrescente. A partir dessa tabela conclui-se que os algoritmos Seleção Seqüencial Crescente e Seleção Seqüencial Decrescente são mais eficientes em termos de esforço computacional que os Algoritmos Genéticos. Esse curto tempo de processamento demandado é devido à simplicidade, tanto estrutural como de operação, desses dois algoritmos seqüenciais.

TABELA 14 - Tempo de processamento demandado

Grupos	Número de Bandas	segundos		minutos	
		SFS	SBS	Torneio	Elitismo
1	3	1,187	1,609	3,428	2,495
2	6	1,203	1,611	3,566	2,746
3	9	1,203	1,626	3,619	2,844
4	12	1,219	1,632	3,673	2,866
5	15	1,219	1,641	3,579	2,778
6	18	1,25	1,645	3,508	2,739
7	21	1,25	1,65	3,606	3,017
8	24	1,235	1,652	3,579	2,972
9	27	1,281	1,654	3,741	2,941
10	30	1,281	1,656	3,678	3,013
11	33	1,282	1,657	3,78	3,029
12	36	1,282	1,66	4,078	3,158
13	39	1,285	1,672	3,828	2,977
14	42	1,285	1,684	3,924	3,228
15	45	1,297	1,696	3,981	3,268
16	48	1,297	1,703	4,311	3,377

Analisando a TABELA 14 verificou-se novamente que o tempo exigido pela técnica de seleção por *torneio* foi superior ao demandado pela técnica de seleção por *elitismo*. Isso permite concluir, a superioridade da técnica *elitismo* em relação à técnica de seleção por *torneio*, no que tange à exigência computacional.

Os Algoritmos Genéticos mostraram-se menos eficientes que os dois algoritmos seqüenciais, no que tange ao tempo de processamento. Entretanto, demonstraram-se mais eficientes em termos de acurácia da classificação. Isto é, para todos os 16 grupos de bandas selecionadas (de 3 a 48), os Algoritmos Genéticos comprovaram ser mais eficientes ao apresentar resultados que geraram mapas temáticos com índice *kappa*, acurácia do ponto de vista do usuário e estimativa (ou exatidão) global, sempre superior aos índices de avaliação da acurácia apresentados pelos resultados dos algoritmos seqüenciais (SFS e SBS).

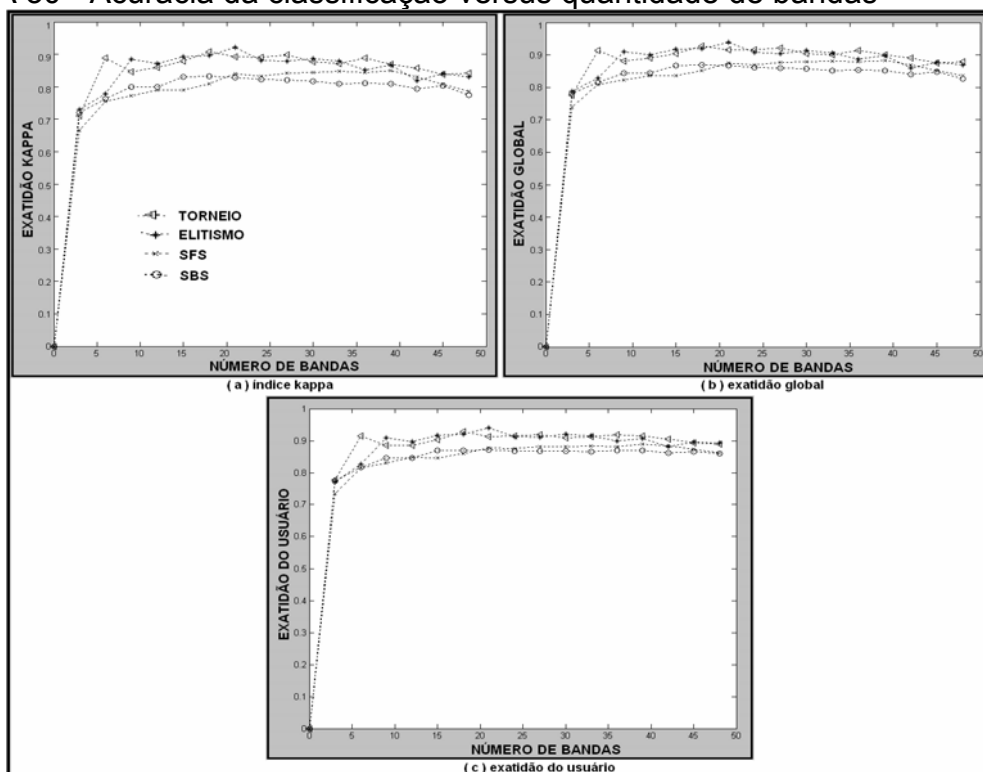
Os valores dos índices *kappa*, calculados para cada mapa temático resultante do processo de classificação dos resultados de cada um dos três algoritmos, estão representados na TABELA 15. Essa tabela evidencia nas primeiras linhas, o que foi comentado na seção 2.3, de que a redução exagerada provoca uma diminuição na eficiência do classificador.

TABELA 15 - Índices *kappa* referentes aos resultados dos três algoritmos

Grupos	Número de Bandas	Índice <i>kappa</i>			
		SFS	SBS	Torneio	Elitismo
1	3	0,6657	0,7196	0,7111	0,729
2	6	0,7543	0,7629	0,8882	0,7791
3	9	0,7726	0,8003	0,8464	0,8837
4	12	0,7903	0,8002	0,8586	0,871
5	15	0,7905	0,8307	0,8778	0,8931
6	18	0,8094	0,8317	0,9067	0,8973
7	21	0,8396	0,8292	0,8917	0,9218
8	24	0,8327	0,8232	0,8902	0,8805
9	27	0,8411	0,8205	0,8986	0,8776
10	30	0,8436	0,8178	0,8766	0,8871
11	33	0,8475	0,8087	0,8718	0,8789
12	36	0,8432	0,8116	0,8888	0,8524
13	39	0,8497	0,8091	0,8703	0,8678
14	42	0,8312	0,7937	0,858	0,8185
15	45	0,8069	0,8026	0,8384	0,8411
16	48	0,7857	0,7739	0,8428	0,8312

A FIGURA 30 mostra os gráficos dos índices de acurácia da classificação resultantes dos três algoritmos.

FIGURA 30 - Acurácia da classificação versus quantidade de bandas



A FIGURA 30(a), referente ao índice *kappa*, mostra com maior clareza a superioridade dos Algoritmos Genéticos sobre os demais. Nos outros dois índices, acurácia global (FIGURA 30b) e acurácia do ponto de vista do usuário (FIGURA 30c), os valores obtidos a partir dos Algoritmos Genéticos, apesar de um menor destaque, também foram superiores.

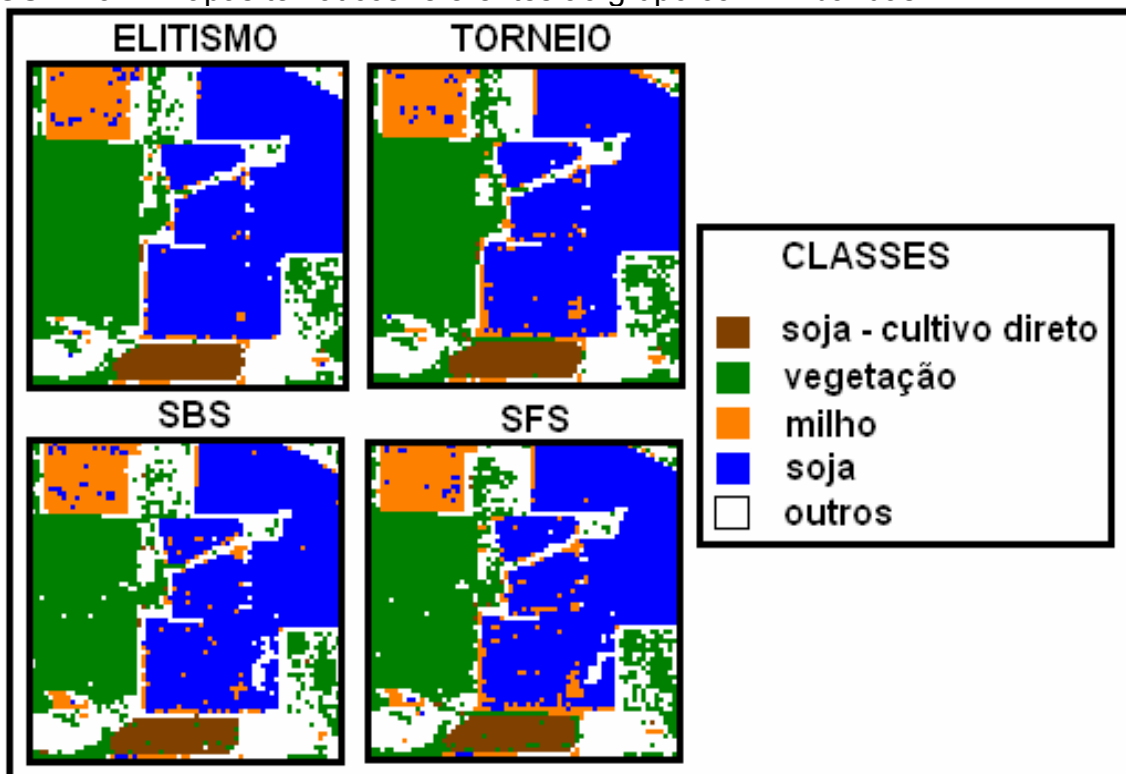
Percebeu-se, nos três gráficos, que todos os algoritmos tendem ao mesmo valor de acurácia com aumento exagerado de bandas. Porém, nesse ponto ocorre uma redução da acurácia, causada pelo efeito do *fenômeno Hughes*, apresentado na seção 2.3.

Analisando os resultados das duas técnicas de seleção dos Algoritmos Genéticos percebeu-se que elas apresentaram um comportamento semelhante entre si. Somente nos subgrupos inferiores a dez bandas é que houve uma maior discrepância entre elas, sendo a técnica *torneio* superior à técnica *elitismo* no subgrupo de 6 bandas. Mas, fazendo uma análise geral sobre toda a curva não se pode afirmar a superioridade de uma técnica sobre a outra.

Nas curvas referentes aos Algoritmos Genéticos observou-se que o ponto de máxima acurácia ocorreu no grupo de 21 bandas. Nesse mesmo ponto, nas curvas dos algoritmos, Seleção Seqüencial Crescente e Seleção Seqüencial Decrescente, verificou-se que houve uma inversão entre eles. Isto é, de zero até aproximadamente vinte bandas, o algoritmo Seleção Seqüencial Decrescente foi superior, todavia, para grupos maiores que vinte bandas o algoritmo Seleção Seqüencial Crescente superou o Seleção Seqüencial Decrescente.

Para uma visualização dos resultados desse experimento, os mapas temáticos, referentes aos resultados do subgrupo de 21 bandas foram ilustrados na FIGURA 31.

FIGURA 31 - Mapas temáticos referentes ao grupo com 21 bandas



Observando a TABELA 15 verifica-se que o mapa temático originado da técnica de seleção por *elitismo*, se comparado com o mapa advindo da seleção *torneio*, apresenta maior índice *kappa*. Examinando os mapas dessas duas técnicas pode-se verificar uma maior confusão, entre as classes: *milho/soja* e *vegetação/outros*, no mapa da seleção por *torneio*. Isso justifica o maior índice *kappa* do mapa temático advindo da seleção por *elitismo*. Esse fenômeno é repetido nos mapas originados dos algoritmos Seleção Seqüencial Crescente e Seleção Seqüencial Decrescente, sendo o primeiro superior ao segundo. Adiciona-se ainda, nessa última comparação, uma maior confusão na classe *cultivo direto*.

5.5 QUINTO EXPERIMENTO: ANÁLISE DO COMPORTAMENTO DOS ALGORITMOS GENÉTICOS FRENTE AO ALGORITMO DO TIPO EXTRAÇÃO DE FEIÇÕES: ANÁLISE POR COMPONENTES PRINCIPAIS

Finalmente, o quinto experimento foi realizado buscando analisar a acurácia dos mapas temáticos resultantes dos Algoritmos Genéticos e da Análise por Componentes Principais.

Como mencionado na seção 5, o objetivo da comparação entre as duas técnicas não leva em consideração o comportamento da última técnica (Análise por Componentes Principais), senão, somente, os resultados obtidos a partir do seu uso. Uma comparação mais detalhada entre essas duas técnicas não seria apropriada, pois trata-se de duas técnicas distintas: uma do tipo seleção de feições e a outra do tipo extração de feições.

Na TABELA 16 estão presentes os valores dos índices *kappa*, acurácia global e acurácia do ponto de vista do usuário, referentes aos mapas temáticos resultantes da Análise por Componentes Principais aplicada sobre os grupos de 190 e 220 bandas. Na última coluna da TABELA 16 estão as porcentagens referentes à informação original representada pelos subconjuntos de 21 componentes, utilizados na classificação.

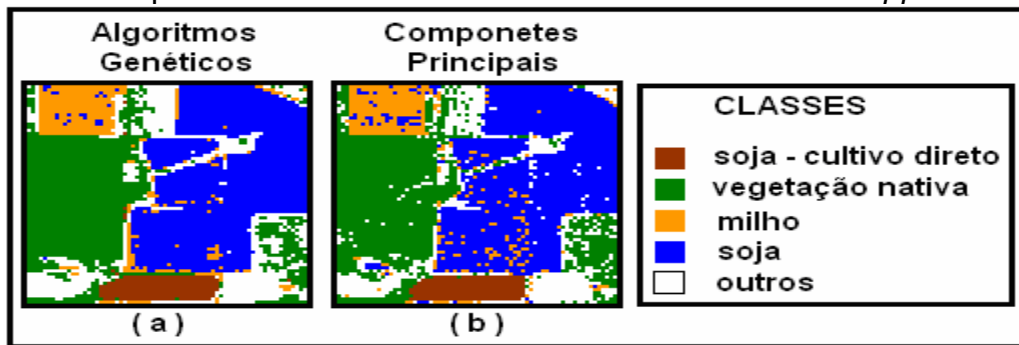
TABELA 16 – Valores do índice *kappa* referentes aos Componentes Principais

	índice kappa	erro global	erro usuário	% (21 bandas)
ACP 190	0,8565	0,8890	0,9010	99,87
ACP 220	0,8748	0,9029	0,9101	99,85

Analisando as tabelas 10, 11 e 16, constatou-se que, para os dois casos aplicados nos Algoritmos Genéticos, a acurácia dos mapas temáticos advindos desses algoritmos foi superior à acurácia dos mapas derivados da Análise por Componentes Principais. Essa superioridade apresenta os Algoritmos Genéticos como promissores no processo de redução de dimensionalidade, pois essa tem ainda sobre aquela técnica, a vantagem de não modificar os dados originais.

A seguir, na FIGURA 32, serão ilustrados os mapas temáticos referentes aos maiores índices *kappa* (0,9218 e 0,8748) gerados a partir dos resultados dos Algoritmos Genéticos e da Análise por Componentes Principais, respectivamente.

FIGURA 32 - Mapas temáticos referentes aos maiores valores de *kappa*



É perceptível uma maior confusão entre os pares de classes: milho/soja e vegetação/outros, no mapa temático ilustrado na FIGURA 32(b) do que a confusão ocorrida no mapa ilustrado na FIGURA 32(a). Isso, obviamente, justifica o menor valor de *kappa* para o mapa resultante da análise por componentes principais.

6. CONCLUSÃO

A eficiência dos Algoritmos Genéticos na redução de dimensionalidade, para fins de classificação, permite afirmar que a metodologia proposta é promissora. Dois parâmetros podem ser levados em consideração nessa avaliação: a eliminação de bandas com ruídos de forma semi-automática e a superioridade da acurácia na classificação sobre as demais técnicas de redução de dimensionalidade estudadas nesse trabalho.

O procedimento de correção gradativa da função de aptidão ao longo dos experimentos foi ideal para o aperfeiçoamento dos Algoritmos Genéticos ao problema proposto. Verificou-se, inicialmente, que esses algoritmos não foram eficientes na eliminação de bandas com ruídos. Contudo, após acrescentado a esses algoritmos um novo fator (ou critério) relacionado com a correlação entre as bandas, constatou-se que eles tornaram hábeis na eliminação de tais bandas.

Os resultados dessa pesquisa comprovaram a evolução da função de aptidão, e conseqüentemente, dos Algoritmos Genéticos, ao apresentarem maior acurácia em relação aos algoritmos Seleção Seqüencial Crescente e Seleção Seqüencial Decrescente.

Nos resultados referentes aos parâmetros da função de aptidão, verificou-se que a influência do parâmetro relacionado ao número de bandas presentes no indivíduo foi crescente ao longo das iterações, até aproximadamente, o grupo de 2000 iterações. A partir desse grupo de 4000 iterações, a curva desse parâmetro estabilizou-se. Essa estabilização implica na contribuição estável (valor constante) do parâmetro para a evolução da espécie. Portanto, a participação dos outros parâmetros torna-se importante no processo de evolução se comparado ao parâmetro relacionado com número de bandas presentes.

Os parâmetros referentes à divergência transformada mostraram-se relevantes no processo de evolução, pois, em todos os casos testados, esses parâmetros apresentaram maior contribuição à função de aptidão. Porém, como ocorreu com o parâmetro relacionado ao número de bandas, o parâmetro referente à divergência transformada também estabilizou-se. Por conseguinte, a partir das 3000 iterações apenas o parâmetro relacionado à correlação entre as bandas contribui para a evolução da espécie.

O uso dos valores mínimos e máximos para os parâmetros da divergência transformada e correlação entre as bandas, respectivamente, foi uma forma encontrada para evitar que a função de aptidão fosse influenciada por situações às vezes indesejadas. Por exemplo, o uso da média dos valores desses parâmetros em bandas onde uma única classe oferece boa separabilidade provoca uma superestimação, e conseqüentemente, afeta a função de aptidão. A média, em casos como o exemplificado, pode mascarar seus resultados por ser influenciada por um valor muito alto ou muito baixo, tanto de separabilidade como de correlação.

A técnica de seleção *elitismo* destacou-se em todos os experimentos realizados, pois apresentou menor tempo de processamento e uma boa acurácia em relação aos outros métodos estudados nessa pesquisa. Fazendo uma análise geral acerca do tempo de processamento, verificou-se que esse fator (tempo de processamento) ainda pode ser melhorado. Existe essa necessidade de reduzir o tempo de processamento, pois a discrepância entre o tempo demandado pelos algoritmos Seleção Seqüencial Crescente e Seleção Seqüencial Decrescente, em relação ao exigido pelos Algoritmos Genéticos, é muito grande. Acredita-se que o ambiente Matlab 5.3, utilizado nessa pesquisa, contribui em parte para esse crescimento do tempo de processamento. Não obstante, o fator, custo computacional, pode ser amenizado se utilizadas outras linguagens de programação, como por exemplo C++.

Uma comparação, em termos de acurácia, entre os resultados advindos dos Algoritmos Genéticos e aqueles obtidos a partir dos algoritmos Seleção Seqüencial Crescente e Seleção Seqüencial Decrescente, possibilitou uma visão otimista sobre os Algoritmos Genéticos. Isto é, nos resultados provenientes da classificação de todos os subgrupos criados nesse experimento, os Algoritmos Genéticos apresentaram maiores valores de acurácia, ou seja, apesar do menor esforço computacional exigido pelos algoritmos Seleção Seqüencial Crescente e Seleção Seqüencial Decrescente, os resultados de acurácia da classificação provenientes desses algoritmos foram inferiores aos oriundos dos Algoritmos Genéticos.

A verificação do comportamento dos Algoritmos Genéticos frente ao algoritmo de extração de feição, Análise por Componentes Principais, também mostrou que estes algoritmos são promissores no processo de redução de dimensionalidade para fins de classificação. Isto é, nos dois testes realizados, com

190 e 220 bandas, os resultados advindos dos Algoritmos Genéticos foram superiores, aos resultados oriundos da Análise por Componentes Principais, em termos de acurácia do produto gerado.

Apesar de tratar de duas técnicas distintas, uma do tipo seleção de feições e outra extração de feições, ainda assim, considera-se importantes esses resultados por dois motivos: primeiro porque a técnica de extração de feições (Análise por Componentes Principais) modifica os dados originais, perdendo o significado físico, enquanto que a técnica seleção de feições (Algoritmos Genéticos) mantém a originalidade dos dados. Segundo, porque partindo da premissa de que a Análise por Componentes Principais é uma técnica comprovadamente eficiente na extração de feições, esses resultados demonstraram que a metodologia proposta é realmente promissora.

O método proposto, após aperfeiçoamentos realizados ao longo dos experimentos, mostrou-se hábil no processo de eliminação de bandas ruidosas. Mostrou-se também apto para selecionar os melhores grupos de bandas que apresentem maior acurácia na classificação. Dessa premissa, considerou-se os Algoritmos Genéticos como eficientes no processo de redução de dimensionalidade para fins de classificação.

7. RECOMENDAÇÕES

- Avaliar outras medidas de similaridades dentro da função de aptidão. Por exemplo, distância de *Bhattacharyya*, *Jeffries Matusita*;
- Avaliar outras medidas de similaridades nos algoritmos Seleção Seqüencial Crescente e Seleção Seqüencial Decrescente. Foi utilizada nesse trabalho apenas a medida da divergência transformada;
- Avaliar outras técnicas de seleção existente nos Algoritmos Genéticos;
- Variar tamanho das amostras de treinamento e de teste. Acredita-se que esse fator pode influenciar os resultados dos Algoritmos Genéticos, por se ter parâmetros estatísticos ainda mais confiáveis para cálculo da medida de similaridade;
- Aplicar essa metodologia em outras imagens que tenham um conjunto diferente de classes.

REFERÊNCIAS BIBLIOGRÁFICAS

ALIXANDRINI JÚNIOR, M. J. **O Potencial das Imagens Hiperespectrais**. In: Congresso Brasileiro de Cadastro Técnico Multifinalitário, 2004, Florianópolis. COBRAC 2004 Florianópolis: UFSC, 2004. p. 1-6.

BEAUCHAMP, W. E. L. **Parallelization of Hyperspectral Imaging Classification and Dimensionality Reduction Algorithms**. A thesis for the degree of MASTER OF SCIENCE IN COMPUTER ENGINEERING. Mayaguez Campus, University of Puerto Rico, 2004. p.67.

BISHOP, Y. M. M.; FIENBERG, S. E.; HOLLAND, P. W. **Discrete Multivariate Analysis: Theory and Practice**. Cambridge, Massachusetts: MIT Press; 1977.

BOLTON, S. K. **Famous Men of Science** (traduzido por J. C. Ribeiro Penna). 19ª edição, Editora: Universitária. São Paulo/SP, 222 pg, 1944.

BON, F. J. C. **Selección y Extracción de Características** Disponível em: http://www-etsi2.ugr.es/depar/ccia/rf/www/tema5_00-01_www/node16.html. Acesso em: 19/12/2005.

BOTELHO, S. C.; BEM, R. de; FIGUEIREDO, M.; LAUTENSCHLGER, W.; CENTENO, T.; MATA, M. **C-NLPCA: Extracting Non-Linear Principal Components of Image Datasets**. XII Simposio Brasileiro de Sensoriamento Remoto. Anais do SBSR. Goiânia, abril/2005. p.3495-3502.

CADIMA, J. F. C. L. **Redução de Dimensionalidade Através Duma Análise em Componentes Principais: Um Critério para o Número de Componentes Principais a Reter**. Revista de Estatística (INE), Portugal, 1o. quadrimestre 2001, p. 37-49. 2001.

CAMPOS, T. E. **Técnicas de Seleção de Atributos e de Classificação para Reconhecimento de Faces**. Projeto de Pesquisa – Exame de Qualificação, Departamento de Ciência da Computação – IME/USP. São Paulo, 2000.

CARVALHO JÚNIOR, O. A.; CARVALHO, A. P. F. de; MENESES, P. R.; GUIMARÃES, R. F.; MARTINS, É. S. **Análise de Imagens Hiperespectrais pelo Método Multiple Endmember Spectral Mixture Analysis (MESMA) em Depósito Supergênico de Níquel**. Revista Brasileira de Geociências 33(1): p.63-74, março/2003.

CENTENO, J. A. S. **Sensoriamento Remoto e Processamento de Imagens Digitais**. Departamento de Geomática, Editora do Curso de Pós-Graduação em Ciências Geodésicas - UFPR. Curitiba/PR 2004.

CHUVIECO, E. **Fundamentos de Teledetección Espacial**. Editora: Ediciones Rialp/AS. Madrid, 1990. 41-48 p.

COHEN, J. A **Coefficient of Agreement of Nominal Scales, Educational and Psychological Measurement.** 1960. p.37-46.

CONTRERAS, S. M. **A Comparison of Resolution Enhancement Methods as Pre-Processing for Classification of Hyperspectral Images.** Thesis for the degree of MASTER OF SCIENCE IN COMPUTER ENGINEERING. Department of Electrical and Computer Engineering. University of Puerto Rico, 2004. p.73.

CROSTA, A. P. **Processamento Digital de Imagens de Sensoriamento Remoto.** Editora IG/UNICAMP. Campinas/SP 1993. 170p.

D'AGOSTINI, A. M.; CARVALHO JÚNIOR, O. A.; CARVALHO, A. P. F.; GUIMARÃES, R. F. **Análise de Mistura Espectral em Imagens Hiperespectrais (Aviris) no Parque Nacional da Chapada dos Veadeiros, GO.** Congresso Brasileiro de Cartografia, Porto Alegre (RS). CD-ROM do XX Congresso Brasileiro de Cartografia. Porto Alegre: FENEA/ISPRS/SIBSIG/SBC, 2001. vol 1. p.1-10.

GALPARSORO, I. L. U.; FERNÁNDEZ, S. P. **Medidas de Concordancia: el Índice Kappa.** Unidad de Epidemiología Clínica y Bioestadística - Cad Aten Primaria. A Coruña – Espana, 1999. p.169-171. Disponível em: <http://www.fisterra.com/mbe/investigal/kappa/kappa.htm>.

GALVÃO, L. S.; FORMAGGIO, A. R.; TISOT, D. A. **Discriminação de Variedades de Cana-de-Açúcar com Dados Hiperespectrais do Sensor HYPERION/EO1.** Revista Brasileira de Cartografia, cidade, (ISSN 0560-4613), N° 57/01, p.7-14, abril 2005.

GOLDBERG, D.E. **Genetic Algorithms in Search, Optimization and Machine Learning.** Editora: Addison-Wesley. New York 1989.

HSU, P.; TSENG, Y. **Feature Extraction for Hyper Spectral Image.** Disponível em: <http://www.gisdevelopment.net/aars/acrs/1999/ts10/ts10405.shtml>. Acesso em: 31/08/2005.

JOHNSON, R.; WICHERN, D. – **Applied Multivariate Statistical Analysis.** Editora: Prentice Hall. New Jersey, 1998, p.816.

KAEWPIJIT, S.; MOIGNE, J. L.; GHAZAWI, T. E. **Automatic Reduction of Hyperspectral Imagery Using Wavelet Spectral Analysis.** IEEE Transactions on Geoscience and Remote Sensing, Vol.41, No.4, p.863- 871, 2003.

KAEWPIJIT, S.; MOIGNE, J. L.; GHAZAWI, T. E. **A Wavelet-based PCA Reduction for Hyperspectral Imagery.** Disponível em: <http://www.esto.nasa.gov/conferences>. Acesso em: 31/08/05.

LILLESAND, T. M.; KIEFER, R. W. **Remote Sensing and Image Interpretation.** 2ª edição, Editora: John Wiley; Sons. New York 1987. p.(655 – 661; 668 – 689 e 675 - 678)

LOPES, H. S. **Algoritmos Genéticos em Projetos de Engenharia: Aplicações e Perspectivas Futuras**. Simpósio Brasileiro de Automação Inteligente, 4., 1999, São Paulo. *Anais do SBAI'99*. São Paulo. 1999. p.64-74.

LOPES, F. M. **Algoritmo Genético Restrito por Listas Tabu no Contexto de Mineração de Dados**. Dissertação de Mestrado, Curso de Pós-Graduação em Informática - UFPR. Curitiba, 2001. p.70.

LUCCA, E. V. D.; BANDEIRA, J. V.; LORENZZETTI, J. A.; MOREIRA, R. C.; CASTRO, R. M.; SALIM, L. H.; ZALOTI JÚNIOR, O. D.; ESPOSITO, E. S. C. **Uso de Sensor Hiperespectral Aerotransportado no Monitoramento da Pluma Termal Oceânica Decorrente da Descarga de Refrigeração da Central Nuclear de Angra dos Reis**. Revista Brasileira de Cartografia Nº 57/01, 2005. (ISSN 1808-0936). p.48-55.

LUTTON, E.; MARTINEZ, P. - **A Genetic Algorithm for the Detection of 2D Geometric Primitives in Images**; 1051-4651/94 IEEE. INRIA - Rocquencourt, B.P. 105, 78153 LE CHESNAY Cedex, France 1994.

MARTINS Jr., D. C. **Redução de Dimensionalidade Utilizando Entropia Condicional Média Aplicada a Problemas de Bioinformática e de Processamento de Imagens**. Dissertação de Mestrado, Ciência da Computação - Instituto de Matemática e Estatística da Universidade de São Paulo. São Paulo, 2004. p.114.

MICHALEWICZ, Z. **Genetic Algorithms + Data Structures = Evolution Programs**. Editora: Springer (ISBN 3-540-60676-9). 3rd rev. and extended ed. New York, 1999.

MITCHELL, M. **An Introduction to Genetic Algorithms**. Cambridge: Mit Press. p.1997. 207

MOIGNE, J. L. **Multi-Sensor Image Registration, Fusion and Dimension Reduction**. Disponível em: <http://www.esto.nasa.gov/conferences>. Acesso em: 31/08/05.

MOREIRA, M. A. **Fundamentos de Sensoriamento Remoto e Metodologia de Aplicação**. 2^a edição, Editora: Universidade Federal de Viçosa – UFV. Viçosa/MG p.2003. 307

NOVO, E. M. L. M. **Sensoriamento Remoto – Princípios e Aplicações**. 2^a edição, Editora: Edgard Blucher Ltda. São Paulo/SP 1992. p.(1-5, 205-265).

PAULINO, W. R. **Biologia Atual (Genética-Evolução-Ecologia)**. 7^a edição, Editora: Ática S.A. São Paulo/SP 1995. p. (12-30, 88-118)

PIZARRO, M. A.; FERNANDES D. **Aplicação da medida de Divergência na Determinação de Membros de Referência em Imagens Hiperespectrais**. Simpósio Brasileiro de Sensoriamento Remoto, 11, 2003, Belo Horizonte. Anais XI SBSR, INPE, p.1083-1090.

PLA, L. E. – **Analisis Multivariado: Metodo de Componentes Principales**. Editora: Eva V. Chesneau. Universidad Nacional Experimental Francisco de Miranda Coro. Falcón, Venezuela p.29-38 1986.

QUENTAL, L.; ABREU, M. M.; OLIVEIRA, V.; SOUSA, P.; BATISTA, M. J.; BRITO, G.; VAIRINHO, M.; SOUSA, J.; MARTINS, L. **Imagens Hiperespectrais para Avaliação e Monitorização Ambiental em Áreas Mineiras: Resultados Preliminares do Projecto Mineo na Mina de São Domingos, Alentejo**. Disponível em: http://www.brgm.fr/mineo/Publications/CPGM_Quental.pdf. Acesso em: 31/08/2005.

RAYMER, M. L.; PUNCH, W. F.; GOODMAN, E. D.; KUHN, L. A.; JAIN, A. K. **Dimensionality Reduction Using Genetic Algorithms**. IEEE Transactions on Evolutionary Computation, vol. 4, N°. 2, July 2000.

RICHARDS, J. A.; JIA, X. **Remote Sensing Digital Image Analysis – an Introduction**. 3ª edição, Editora: Springer. Australian, p.363, 1998.

SALEHI, B.; ZOUJ M.J.V. **Wavelet-Based Reduction of Hyperspectral Imagery**. Faculty of Geodesy and Geomatics Engineering, K.N.Toosi University of Technology, Tehran Iran. Disponível em: <http://www.gisdevelopment.net/technology/ip/mi04013pf.htm>. Acesso em: 31/08/2005.

SERPICO, S. B.; D'INCA, M.; MELGANI, F.; MOSER, G. **A comparison of feature reduction techniques for classification of hyperspectral remote-sensing data**. Image and Signal Processing of Remote Sensing VIII, Sebastiano B. Serpico, Editor, In: Proceedings of SPIE, v. 4885, 2002.

SHIPPERT, P. **Introduction to Hyperspectral Image Analysis**. Disponível em: www.researchsystems.com. Acesso em: 28/11/2005.

SWAIN, P. H.; DAVIS, M. S. **Remote Sensing the Quantitative Approach**. Editora: McGraw-Hill. New York, p.166-174, 1978.

UNICAMP - <http://www.mundogeo.com.br/resumos/10-EOBrasil%20BETO%20Palestra.htm>. Acesso em: 31/08/2005.

ZORTEA, M.; HAERTEL, V.; CLARKE, R. T. **Investigação Sobre Métodos para Redução da Dimensionalidade dos Dados em Imagens Hiperespectrais**. Revista Brasileira de Cartografia N° 57/01 (ISSN 0560-4613), p.36-42, abril/2005.