

ANÁLISE DISCRIMINANTE ASSOCIADA À ORIENTAÇÃO A OBJETO PARA O MAPEAMENTO DA COBERTURA VEGETAL EM UMA IMAGEM DE ALTA RESOLUÇÃO

Alessandro Ribeiro Campos*

Resumo

As imagens de satélite derivadas de sensores com melhores resoluções temporais, espaciais, espectrais e radiométricas são cada vez mais utilizadas para diversas finalidades, como o mapeamento da cobertura vegetal, tendo em vista a potencial melhora na precisão desses mapeamentos. Por outro lado, a maior disponibilidade e utilização desses sistemas sensores impõem novos desafios para o processamento dessas imagens para a extração de informações, sobretudo no que se refere à classificação da cobertura vegetal, principalmente pela disponibilidade de um grande volume de dados e suas implicações no campo do custo computacional e das técnicas de processamento. A partir disso, o principal objetivo deste artigo foi investigar se a análise discriminante associada à orientação a objeto pode classificar corretamente os tipos de cobertura vegetal em uma imagem de alta resolução espacial. Para tanto, conciliou o ambiente computacional dos sistemas de informação geográfica e do pacote estatístico para execução do processo. Os resultados indicaram que houve uma correspondência satisfatória entre as classes previstas pela probabilidade condicional de pertencimento da análise discriminante para os objetos nos tipos de cobertura vegetal.

Palavras-chave: Classificação, análise discriminante, orientação a objeto

Abstract

Satellite imagery derived from sensors with better resolutions temporal, spatial, spectral and radiometric are increasingly used for various purposes, such as mapping of vegetation, with a view to potential improvement in the accuracy of these maps. Moreover, the increased availability and use of these sensor systems, poses new challenges for processing these images to extract information, especially with regard to the classification of vegetation, mainly by the availability of a large volume of data and its implications in the field of computational cost and processing techniques. From this, the main objective of this paper was to investigate the discriminant analysis associated with object orientation can accurately classify the types of vegetation in an image with high spatial resolution. Therefore, reconciled the computing environment of geographical information systems and statistical package for process execution. The results indicated that there was a good match between the classes predicted by the conditional probability of belonging discriminant analysis for objects in the types of vegetation cover.

Keywords: Classification, discriminant analysis, object orientation

*Universidade Federal de Minas Gerais – UFMG. alessandrorc1@yahoo.com.br

1- INTRODUÇÃO

A identificação, classificação e monitoramento da cobertura da terra são atividades relevantes por várias razões, especialmente porque constituem uma informação essencial para o estudo dos ciclos geoquímicos globais e para as pesquisas em sistemas terrestres, como a modelagem de ecossistemas e do impacto resultante das atividades antrópicas (ZHANG et al., 2003). Para subsidiar essas atividades, dados de sensoriamento remoto orbital cada vez mais têm sido usados.

Nesse contexto, o mapeamento da cobertura vegetal tem se destacado, na medida em que existe um esforço por parte da comunidade científica em melhorar a acurácia destes estudos (HANSEN et al., 2002). A utilização de imagens de satélite derivadas de sensores com melhores resoluções temporais, espaciais, espectrais e radiométricas podem melhorar a precisão desses mapeamentos, aumentando a confiança da classificação das diversas formações vegetais.

Por outro lado, a maior disponibilização e utilização desses sistemas sensores, sobretudo aqueles de alta resolução espectral e espacial, impõe sofisticação e robustez aos métodos quantitativos empregados no Processamento Digital de Imagens - PDI, cuja finalidade é aprimorar o aspecto visual de certas feições estruturais e extrair informações para o analista humano (GONZALEZ e WOODS, 2000), o que subsidia, dentre outros, o mapeamento da cobertura vegetal. Sob o aspecto do processamento, esse mapeamento é obtido através do emprego de um sistema de medidas baseado na imagem (VAN DER HEIJDEN, 1994) que possibilitam o reconhecimento de padrões e a identificação de classes de objetos a partir de um vetor de n variáveis aleatórias ($x_1, x_2, x_3 \dots x_n$).

O principal aspecto desafiante, referente à classificação e utilizando esse tipo de imagem, diz respeito à disponibilidade de um grande volume de dados e suas implicações no campo do custo computacional e das técnicas de processamento. Isso se deve ao fato de que com o aumento da resolução espacial e espectral, os detalhes e as características dos objetos são cada vez mais distintos (MING et al., 2009). Com isso, a aplicação de técnicas convencionais de classificação que consideram apenas as características individuais de cada pixel da imagem (denominadas pixel-a-pixel) podem não ser suficientes para a distinção adequada das feições fisiográficas da paisagem, ou seja, reconhecer classes, devido a tendência de haver uma alta variância intra-classe (GONZALEZ e WOODS, 2000), comprometendo significativamente o resultado da classificação.

Uma opção de técnica para tratamento desses dados para classificação seria incorporar nesse processo outros aspectos, atributos e parâmetros além da característica

espectral, passando assim a abordar não mais os pixels individualmente, mas os objetos da imagem. Essa abordagem de classificação é denominada orientada a objeto (BLASCHKE, 2010).

Esse método é destacado principalmente por utilizar como unidade de processamento os objetos da imagem que compartilham atributos semelhantes, tais como: nível de cinza, características espectrais, textura, tamanho, forma, compacidade, informação do contexto dos objetos adjacentes etc. (CÂMARA et al., 1996; GONZALEZ e WOODS, 2000; LIU et al., 2006; BATTZ et al., 2008; MING et al., 2009; BLASCHKE, 2010), sendo necessário para tal o emprego de técnicas de segmentação que permitirão a extração desses objetos.

Os algoritmos segmentadores dividem as imagens em segmentos (objetos) internamente homogêneos e externamente dissimilares (GONZALEZ e WOODS, 2000) baseado em alguma característica específica, de modo que tenham coesão interna e sejam heterogêneos entre si. Identificados os objetos, são extraídas suas características gerais por meio de n -variáveis morfométricas, espectrais e texturais, de acordo com o algoritmo utilizado.

A partir da extração desses atributos é possível aplicar técnicas estatísticas multivariadas exploratórias de sintetização (ou simplificação) da estrutura de variabilidade dos dados para discriminação e classificação desses objetos identificados na imagem, na medida em que se tem um conjunto n -dimensional de variáveis, medidas simultaneamente, que os caracteriza e podem ser utilizadas para distingui-los,

O objetivo imediato da discriminação e classificação, próprio da estatística, mas que pode ser estendido ao processamento digital de imagens, segundo Johnson e Wichern (2002), é:

i) Descrever, de forma gráfica ou algébrica, as características diferenciais dos objetos (observações) de vários conjuntos (populações) conhecidos, tentando assim encontrar discriminantes cujos valores numéricos maximizam a separação entre esses conjuntos;

ii) Classificar objetos (observações) dentro de dois ou mais grupos previamente determinados, com ênfase em identificar a regra derivada desse processo, que pode ser utilizada para atribuir novos objetos a esses grupos.

Nesse sentido, uma técnica multivariada direcionada a separação de distintos conjuntos de observações e a alocação de novos objetos em grupos previamente definidos (JOHNSON e WICHERN, 2002), que pode ser aplicada como método supervisionado de

classificação de imagens de satélite orientado a objeto é análise discriminante (LOBO et al., 1998; THENKABAIL et al., 2004; THESSLER et al., 2008).

A análise discriminante é um conjunto de métodos e ferramentas utilizado para distinção entre grupos de Π_j populações e determinar como alocar novas observações nestes grupos (HÄRDLE e SIMAR, 2007). É uma técnica que pode ser utilizada para classificação de elementos de uma amostra ou população, em que é necessário que os grupos para os quais cada elemento amostral pode ser classificado sejam predefinidos e suas características gerais conhecidas a priori (MINGOTI, 2005). A partir disso é possível elaborar uma função matemática chamada de regra de classificação ou discriminação, que é utilizada para classificar novos elementos amostrais nos grupos já existentes.

O modelo discriminante aplicado para a classificação dos objetos da imagem de satélite buscaria por uma combinação linear das variáveis independentes, como as características espectrais, textura e forma dos objetos da imagem, que melhor os separaria entre as classes predefinidas de cobertura vegetal, sendo que os parâmetros do modelo seriam estimados com base nos objetos de amostra identificados previamente.

A partir do exposto, o principal objetivo deste artigo foi investigar se a análise discriminante associada à orientação a objeto pode classificar corretamente os tipos de cobertura vegetal em uma imagem de alta resolução espacial, além da identificação das variáveis que melhor discriminam esses tipos mutuamente exclusivos.

1.1. Análise Discriminante

Segundo Härdle e Simar (2007), o problema da discriminação entre dois ou mais grupos, visando posterior classificação, foi inicialmente abordado por Fisher. Consiste em obter funções matemáticas capazes de classificar um indivíduo X (uma observação X) em uma de várias populações π_i , ($i=1, 2, \dots, g$), com base em medidas de um número p de características, buscando minimizar a probabilidade de má classificação, isto é, minimizar a probabilidade de classificar erroneamente um indivíduo em uma população π_i , quando realmente pertence a população π_j , ($i \neq j$) $i, j=1, 2, \dots, g$.

A análise discriminante envolve determinar uma variável estatística que é a combinação linear de duas ou mais variáveis independentes que discriminarão melhor entre grupos definidos a priori. A ideia de Fisher foi transformar as observações multivariadas X em observações univariadas Y tal que as Y 's obtidas a partir das populações π_1 e π_2 fossem o mais “separadas” possíveis.

A discriminação é conseguida estabelecendo-se os pesos da variável estatística para cada variável para maximizar a variância entre grupos relativa à variância dentro dos grupos. A combinação linear para uma análise discriminante, também conhecida como a função discriminante, é determinada de uma equação que assume a seguinte forma:

$$z_{jk} = a + w_1x_{1k} + w_2x_{2k} + \dots + w_nx_{nk} \quad (1)$$

onde

z_{jk} = escore Z discriminante da função discriminante j para o objeto k

a = intercepto

w_i = peso discriminante para a variável independente i

x_{ik} = objeto k na variável X_i .

2. MATERIAIS E MÉTODOS

Para análise da aplicação da análise discriminante como classificador supervisionado para o mapeamento da cobertura vegetal foi utilizada a imagem 2329621 de 23/06/2010 do instrumento sensor REIS (RapidEye Earth Imaging System) a bordo da constelação de satélites RapidEye. O sensor (pushbroom imager) multiespectral REIS obtém imagens da Terra em cinco faixas espectrais, sendo elas, o Azul (440–510nm), Verde (520–590nm), Vermelho (630–685nm), Red-Edge (690–730nm), sensível a alterações do teor de clorofila das plantas e Infravermelho Próximo (760–850nm), com uma área imageada 25 km² por cena, 5 metros de resolução espacial para o produto ortorretificado e resolução Radiométrica de 16 bits.

A escolha da cena utilizada teve como requisito principal a existência de pelo menos quatro tipos de cobertura vegetal, identificadas a partir da sobreposição da grade de imagens RapidEye ao mapa de cobertura vegetal 2009¹ do Instituto Estadual de Florestas de Minas Gerais - IEF.

Foi necessário realizar o pré-processamento da cena no que diz respeito ao registro, correção radiométrica e aplicação de contrastes nas bandas utilizadas. Para melhorar a segmentação e classificação dos objetos foi gerado o Normalized Difference Vegetation

¹Disponível em: <http://geosisemanet.meioambiente.mg.gov.br/zee/>

Index (NDVI), índice que envolve a diferença e a soma entre as bandas do infravermelho próximo (NIR) e do vermelho (RED), segundo a razão:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (2)$$

O fluxo de trabalho relacionado ao processamento da classificação foi dividido em duas etapas, claramente demarcadas pelo processo e o ambiente computacional de processamento: a segmentação, em que foram gerados os objetos da imagem e extraídos seus atributos; e a classificação, quando foi aplicada a análise discriminante utilizando os objetos da imagem.

A primeira etapa foi realizada no software ENVI 5.0, utilizando o segmentador Edge, que está baseado no algoritmo de detecção de bacias por imersão (watershed) de Vincent e Soille (1991). O procedimento de segmentação por detecção de bacias pressupõe uma representação topográfica para a imagem, ou seja, para uma dada imagem gradiente, o valor de nível digital de cada pixel equivale a um valor de elevação naquele ponto. O operador watershed é uma ferramenta poderosa para detectar contornos de objetos (POLIDORIO et al., 2005). A ideia é considerar o objeto a ser segmentado como sendo uma bacia de retenção. As linhas de partição de águas resultante definem o contorno do objeto.

Alinhado com o objetivo do trabalho, esse processo de segmentação utilizou uma máscara de cobertura vegetal, gerada a partir de uma segmentação previa baseada em regra utilizando o NDVI para filtrar os segmentos associados à cobertura do solo com vegetação. Após identificação dos objetos da imagem (exemplificado pela Figura 1), o algoritmo efetuou a extração de suas características gerais a partir dos seguintes atributos:

- atributos espectrais: média (AVG_Bx), máximo (MAX_Bx), mínimo (MIN_Bx) e desvio padrão (STD_Bx) de todas as bandas, incluindo o NDVI;
- atributos de textura: média do intervalo de dados na janela 3x3 (TXRAN_Bx), média do valor dos pixels na janela 3x3 (TXAVG_Bx), variância média dos pixels na janela 3x3 (TXVAR_Bx) e média da entropia do valor dos pixels de todas as bandas (TXENT_Bx), incluindo o NDVI;
- espaço de cor: a composição colorida das bandas 3R2G1B foi transformada para o espaço IHS, formando três novas camadas de informação para a imagem:

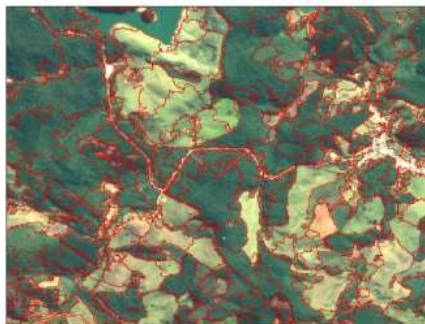


Fig. 1 - Exemplo dos segmentos extraídos

Intensity, Hue e Saturation. Para essas, também foram mensuradas todas as métricas acima.

Ao final dessa etapa, totalizaram 66.495 objetos (observações) com um conjunto de 72 variáveis. Antes da conversão dessa base de dados para o formato apropriado para a segunda etapa do processo, foram criados p-vetores de amostras, em que $p = n$ classes. Um total de 477 objetos foram classificados a priori como pertencente a uma das classes de cobertura vegetação existente na região: Floresta estacional semidecidual Montana; Campo; Campo rupestre; e Pasto.

A segunda etapa foi realizada utilizando o pacote estatístico *Statistical Package of the Social Sciences* (SPSS® V. 19). Nessa etapa foi desenvolvida a análise discriminante utilizando os objetos extraídos da imagem.

Uma vez selecionado o conjunto q de variáveis independentes, as funções de classificação, ou funções discriminantes lineares de Fisher, foram calculadas utilizando (SPSS, 2010):

$$b_{ij} = (n - g) \sum_{i=1}^q w_{ij}^* \bar{X}_{ij} \quad i=1, 2, \dots, q, j=1, 2, \dots, g \quad (3)$$

para os coeficientes e,

$$a_j = \log p_j - \frac{1}{2} \sum_{i=1}^q b_{ij} \bar{X}_{ij} \quad j=1, 2, \dots, g \quad (4)$$

para a constante, onde n é a soma total de pesos, q o número de variáveis selecionadas, g o número de grupos, a média da i \bar{X}_{ij} no grupo j , e w_{ij} a soma dos quadrados dentro dos grupo e produto cruzado da matriz (W).

O processo básico para classificação dos objetos seguiu os seguintes passos (SPSS, 2010):

- Se X é o $1 \times q$ vetor de variáveis discriminantes para o objeto, o $1 \times m$ vetor de valores da função discriminante canônica é $f = XB + a$;
- Uma distância Qui-quadrado de cada centróide é calculado:

$$\chi_j^2 = (\mathbf{f} - \bar{\mathbf{f}}_j) \mathbf{D}_j^{-1} (\mathbf{f} - \bar{\mathbf{f}}_j)' \quad (5)$$

onde \mathbf{D}_j é a matriz de covariância da função discriminante canônica para o grupo j , e o centroide do grupo.

- A classificação, ou probabilidade a posteriori, $P(G_j | X)$ é dada por:

$$g_j = \log P_j - \frac{1}{2} (\log |\mathbf{D}_j| + \chi_j^2) \quad (6)$$

onde P_j é a probabilidade a priori para o grupo j . O objeto é atribuído ao grupo para o qual $P(G_j | X)$ é o mais alto.

Com o objetivo de modelar a análise discriminante, diversas estatísticas e testes foram utilizados para analisar o poder de discriminação do modelo e avaliar os pressupostos para aplicação da técnica, como o lambda de Wilks, a estatística Box's M , o teste F , Qui-quadrado, a correlação canônica e o autovalor.

O lambda de Wilks, que varia de 0 a 1, propicia a avaliação da existência de diferenças de médias entre os grupos para cada variável e assim a significância da função discriminante. Valores elevados desta estatística indicam ausência de diferenças entre os grupos.

A estatística M de Box testa a hipótese nula de que as matrizes de covariância não diferem entre os grupos formados. Espera-se que este teste não seja significativa para que a hipótese nula de que os grupos não diferem possa ser mantida. Utilizou-se F como aproximação para o teste a um nível de significância de 0,05.

Identificadas as funções discriminantes e classificadas as observações, verificou-se a acurácia da classificação através da matriz de erro (CONGALTON e GREEN, 2009), que

indicou o percentual de acerto da classificação comparando os objetos com categoria definida a priori com o resultado do processo.

Por fim, os resultados foram migrados para o ambiente de Sistemas de Informação Geográfica – SIG, retornando os objetos, agora classificados, para a estrutura vetorial. Como faz parte do escopo desse trabalho a aplicação da técnica como classificador supervisionado para o mapeamento da cobertura vegetal, também foram feitas análises referentes à precisão da classificação das distintas categorias temáticas utilizando amostras de avaliação. Nesse sentido, foram selecionadas amostras de referência, de forma supervisionada na imagem e coletadas em campo, a partir das quais foi gerada a matriz de erro e o índice Kappa, calculado seguindo a expressão:

$$k = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \cdot x_{+i})} \quad (7)$$

3. RESULTADOS E DISCUSSÃO

Considerando que na análise discriminante o objetivo passa por prever os membros de um grupo, é prudente verificar a existência de diferenças entre os grupos a partir das médias de cada variável independente, utilizando para isso a análise de variância. A Tabela 1 apresenta os resultados do Lambda de Wilks, a aproximação pela estatística F e a significância dos testes.

O Lambda de Wilks, testando a hipótese H0 de que as médias das variáveis nos grupos são iguais, foi significativa (p-valor > 0,05) apenas para a variável txent_b6 (média da entropia do valor dos pixels do NDVI), indicando uma forte evidência de que essa variável não é significativa na diferenciação entre os grupos. As demais variáveis foram significantes, sendo que a avg_b2 (média da banda 2) e a txavg_b2 (textura média da banda 2) são as de maior importância para a função discriminante, o que é sugerido pelos baixos valores do Lambda de Wilks e altos da estatística F. Esse teste de igualdade de médias sugere com significância estatística que os valores médios dos pixels e de textura média das bandas 2, 3 e 4, do NDVI e de Saturação (b9) do espaço de cor IHS são boas variáveis, comparadas as demais, para discriminar os tipos de cobertura vegetal para a imagem que esta sendo analisada.

O teste M de Box (Box's M) nos permite avaliar uma das pressuposições da análise discriminante, que é a homogeneidade das matrizes de variância e covariância, em cada um dos grupos e para cada uma das variáveis de estudo, testando a hipótese nula (H0) de que as matrizes são homogêneas. Embora o resultado desse teste não ter sido significativo ao nível de significância de 0,05, a princípio sugerindo a não homogeneidade das matrizes, e considerando que ele é muito sensível ao tamanho amostral e pelas diferenças de tamanho que algumas amostras podem apresentar, seu resultado deve ser interpretado em conjunto com o Log Determinants (SPSS, 2010). Como essa medida de variabilidade dos grupos obteve valores relativamente iguais entre as classes de cobertura, considerou-se então a não violação do pressuposto.

TABELA 1 - TESTE DE IGUALDADE DE MÉDIAS DOS GRUPOS

Variável	Lambda de Wilks	F	df1	df2	p-valor	Variável	Lambda de Wilks	F	df1	df2	p-valor
avg_b2	,185	813,702	3	554	,000	std_b2	,624	111,310	3	554	,000
txavg_b2	,186	810,115	3	554	,000	txvar_b9	,626	110,235	3	554	,000
txavg_b3	,210	695,865	3	554	,000	txvar_b4	,629	109,059	3	554	,000
avg_b3	,210	695,602	3	554	,000	txran_b4	,643	102,513	3	554	,000
avg_b9	,225	634,475	3	554	,000	max_b1	,649	100,006	3	554	,000
txavg_b9	,227	630,566	3	554	,000	txent_b2	,655	97,136	3	554	,000
txavg_b4	,258	530,921	3	554	,000	txavg_b5	,655	97,081	3	554	,000
avg_b4	,259	529,404	3	554	,000	max_b6	,658	96,051	3	554	,000
txavg_b1	,305	420,277	3	554	,000	max_b9	,659	95,766	3	554	,000
avg_b1	,306	419,674	3	554	,000	avg_b5	,659	95,375	3	554	,000
txent_b3	,343	353,890	3	554	,000	min_b1	,671	90,371	3	554	,000
txran_b3	,364	322,556	3	554	,000	txent_b9	,672	90,301	3	554	,000
txavg_b6	,401	275,922	3	554	,000	txvar_b1	,708	76,139	3	554	,000
avg_b6	,410	265,971	3	554	,000	txvar_b5	,732	67,509	3	554	,000
std_b9	,483	197,356	3	554	,000	max_b8	,737	66,018	3	554	,000
txavg_b8	,485	195,958	3	554	,000	max_b5	,746	62,744	3	554	,000
avg_b8	,486	195,302	3	554	,000	max_b7	,757	59,293	3	554	,000
avg_b7	,497	186,551	3	554	,000	txran_b6	,799	46,321	3	554	,000
txavg_b7	,499	185,746	3	554	,000	min_b5	,803	45,356	3	554	,000
std_b3	,499	185,235	3	554	,000	txent_b1	,808	43,828	3	554	,000
max_b3	,521	169,594	3	554	,000	std_b4	,816	41,636	3	554	,000
txran_b2	,525	166,922	3	554	,000	txent_b5	,832	37,342	3	554	,000
txran_b9	,530	163,537	3	554	,000	txent_b4	,839	35,458	3	554	,000
min_b6	,533	162,090	3	554	,000	txran_b7	,846	33,639	3	554	,000
min_b8	,547	153,115	3	554	,000	txent_b8	,850	32,686	3	554	,000
min_b2	,548	152,305	3	554	,000	std_b7	,850	32,638	3	554	,000
txvar_b3	,549	151,741	3	554	,000	std_b6	,868	28,191	3	554	,000
min_b3	,555	148,250	3	554	,000	txvar_b6	,873	26,978	3	554	,000
max_b2	,557	147,071	3	554	,000	txran_b8	,881	24,909	3	554	,000
txran_b1	,583	131,945	3	554	,000	txvar_b8	,884	24,266	3	554	,000
min_b4	,584	131,768	3	554	,000	std_b5	,886	23,694	3	554	,000
min_b9	,588	129,252	3	554	,000	txvar_b7	,896	21,337	3	554	,000
max_b4	,594	126,179	3	554	,000	min_b7	,933	13,178	3	554	,000
txvar_b2	,603	121,783	3	554	,000	txent_b7	,951	9,468	3	554	,000
txran_b5	,609	118,629	3	554	,000	std_b8	,959	7,958	3	554	,000
std_b1	,613	116,708	3	554	,000	txent_b6	,992	1,502	3	554	,213

Tendo em vista o número de grupos ($g = 4$), foram estabelecidas três ($g - 1$) funções discriminantes de Fisher (Z_1 , Z_2 e Z_3) com as combinações lineares das 72 variáveis independentes e o correspondente vetor W de pesos, estimados de modo que a

variabilidade dos escores da discriminante seja máxima entre os grupos e mínima dentro dos grupos. Como medida relativa da eficácia de cada função discriminante, os Autovalores (Eigenvalues) são apresentados na Tabela 2.

TABELA 2. AUTOVALORES (EIGENVALUES)

Função	Autovalor	% de Variância	% Acumulado	Correlação Canônica	Correlação Canônica ²
1	11,823	67,8	67,8	,960	,922
2	4,261	24,4	92,3	,900	,810
3	1,344	7,7	100,0	,757	,573

O autovalor para a primeira função (11,823) indica que ela tem o maior poder de explicação da variância da variável dependente, com o percentual de variância de 67,8%, demonstrando assim sua maior contribuição para demonstrar as diferenças entre os grupos. A correlação canônica, medida de associação múltipla entre as variáveis dependentes e as funções discriminantes, sugere que o modelo explica, através da primeira função, 92,2% (Correlação Canônica²) da variação da variável de agrupamento. O percentual acumulado indica que quase toda a variância explicada pelo modelo é devido às duas primeiras funções discriminantes.

Em termos de significância, as três funções foram testadas em conjunto utilizando o Lambda de Wilks e o Qui-quadrado, apresentados na Tabela 3.

TABELA 3. LAMBDA DE WILKS E QUI-QUADRADO

Teste de Funções	Lambda de Wilks	Qui-quadrado	df	Sig.
1 até 3	,006	2655,703	183	,000
2 até 3	,081	1317,569	120	,000
3	,427	446,723	59	,000

Os resultados sugerem que as três funções são estatisticamente significantes para o modelo e assim para a separação das classes de cobertura vegetal, tendo em vista a rejeição da hipótese nula (H0) de igualdade das médias entre os grupos a um nível de significância de 0,05.

A partir da matriz estrutura (Tabela 4), que indica a correlação entre de cada variável independente com as funções discriminantes canônicas padronizadas, é possível interpretar a contribuição que cada variável forneceu para cada função.

TABELA 4. CORRELAÇÃO ENTRE AS VARIÁVEIS E AS FUNÇÕES

Variável	Funções		
	1	2	3
avg_b2	-,593*	-,139	,348
txavg_b2	-,592*	-,147	,343
txavg_b3	-,531*	-,317	,180
avg_b5	-,528*	-,315	,184
avg_b9	-,528*	-,132	,227
txavg_b9	-,527*	-,136	,225
avg_b6	,231	,422*	,192
txavg_b6*	,241	,419*	,189
txran_b5	,104	,329*	-,198
max_b6	,122	,283*	,052
txavg_b8	,234	,283*	-,248
avg_b8	,232	,282*	-,249
txran_b9	-,227	-,088	-,426*
txran_b1	-,167	-,203	-,396*
txran_b2	-,247	,038	-,363*
txvar_b9	-,190	-,048	-,347*
txvar_b1	-,114	-,158	-,337*
txent_b1	-,084	-,019	-,336*

A Tabela 4 apresenta apenas as variáveis que tem as maiores correlações com as funções, sendo que as demais foram suprimidas devido a baixa correlação.

*. Menores correlações entre a variável e a função

Nesta análise, considerando a imagem e os objetos com classificação a priori selecionados, há indícios, a partir dos coeficientes de correlação, de que a variável avg_b2 (média da banda 2) é a de maior relevância para a determinação da primeira função discriminante, a variável avg_b6 (média do NDVI) para a segunda função e a txran_b9 (textura média do intervalo de dados na janela 3x3) para a terceira função. A partir dessas correlações, as funções podem ser identificadas da seguinte forma:

Função 1: Espectral e textura do visível e cor

Função 2: Espectral e textura do infravermelho e cor

Função 3: Textura do visível e cor

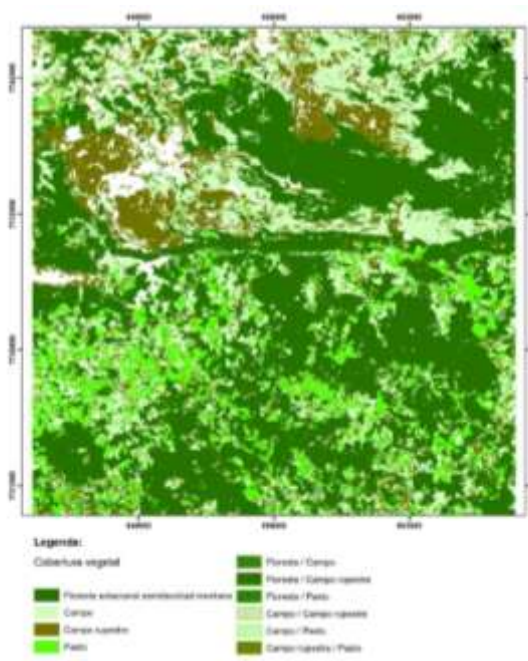


Fig. 2 - Mapa de cobertura vegetal

Os objetos foram diretamente classificados na classe para o qual $P(G_j|X)$ era igual ou superior a 75%, que indica, de maneira inversa, uma chance razoavelmente baixa do objeto ser classificado erroneamente. Para os demais objetos com $P(G_j|X)$ inferior à 75% para qualquer classe, cujas características não são suficientes para os discriminar, o que significa que podem ter atributos de mais de uma classe como uma transição ou mistura, foram identificadas as classes com maiores probabilidades e considerados sub-classes entre estas. Nesse sentido, foram formadas as seguintes classes derivadas das identificadas a priori na imagem: Floresta/Campo; Floresta / Campo rupestre; Floresta / Pasto; Campo / Campo rupestre; Campo / Pasto; e Campo rupestre / Pasto. A Figura 3 exemplifica a ocorrência da subclasse Campo / Campo rupestre em uma região de transição.

A matriz de erro com amostras de avaliação e o índice Kappa (Tabela 5) sugerem que o resultado da análise discriminante foi significativo em termos de acurácia temática da classificação.

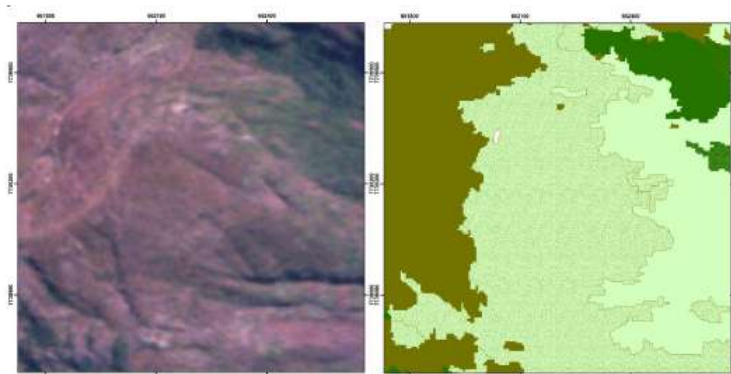


Fig. 3 - Exemplo da ocorrência da subclasse Campo / Campo rupestre

TABELA 5. MATRIZ DE ERRO

		Referência				Σ
		Floresta	Campo	Campo rupestre	Pasto	
Classificação	Classes					
	Floresta	329	7	5	1	342
	Campo	10	156	7	3	176
	Campo rupestre	0	13	131	0	144
	Pasto	1	4	0	128	133
Σ		340	180	143	132	795
		Floresta	Campo	Campo rupestre	Pasto	
Acurácia do Produtor		0,9676	0,8667	0,9161	0,9697	
Acurácia do Usuário		0,9620	0,8364	0,9097	0,9624	
Acurácia Total		0,9358				
Kappa		0,909				

A acurácia do produtor sugere que o processo de classificação teve uma boa habilidade preditora dos objetos para todas as classes, sendo a menor capacidade associada ao reconhecimento da classe campo (0,8667). Por outro lado, é importante chamar a atenção que uma alta acurácia do produtor pode ocorrer pela grande quantidade de objetos classificados, por exemplo como floresta, no mapa. O parâmetro que pode ser utilizado conjuntamente para avaliar a precisão é a acurácia do usuário. Reforçando o bom desempenho da classificação da cobertura vegetal apresentada, ela mostra que 96,20% dos objetos chamados de floresta no mapa são realmente floresta na referência.

Por fim, essa análise indica que o mapa resultado da classificação tem uma acurácia de 93%, o que significa que 93% das amostras de avaliação possuem correspondência entre a classe predita e aquela que é de fato, conforme referência. Esse é considerado um alto grau de assertividade (CONGALTON e GREEN, 2009), sendo suficiente para validar um processo de classificação.

4. CONSIDERAÇÕES FINAIS

O processo de classificação de imagens orbitais da superfície terrestre é especialmente complexo quando se trata de alta resolução espacial e espectral, dada a alta dimensionalidade possibilitada e a alta variância derivada dessas representações. Pesquisas relativamente recentes apontam para orientada a objeto como solução em termos de abordagem para segmentação desse tipo de imagem e experiências com classificadores são cada vez mais necessárias. Este estudo baseou-se na hipótese de que um classificador linear poderia não somente capturar informações que são úteis para fins de discriminação, mas também classificar corretamente os tipos de cobertura vegetal em uma imagem de alta resolução espacial, aplicando a análise discriminante de Fisher para tal.

Foi utilizado um conjunto de variáveis que caracterizavam, em termos de atributos espectrais, texturais e de cor, os objetos extraídos da imagem utilizando um algoritmo segmentador orientado a objetos. A partir desses foi desenvolvido um modelo saturado (incluindo todas as variáveis) discriminante, parametrizado por objetos amostrados cujas classes de cobertura vegetal foram identificadas a priori, gerando funções que foram utilizadas para prever a classe dos demais objetos.

O processo permitiu uma análise exploratória das variáveis, identificando aquelas com maior poder discriminante e assim o grau de contribuição de cada uma para a distinção dos tipos de cobertura.

As medidas de acurácia temática indicaram que houve uma correspondência satisfatória entre as classes preditas pela probabilidade condicional de pertencimento para os objetos e as amostras de referência coletadas em campo e supervisionadamente na imagem. Um ganho importante da aplicação da análise discriminante na classificação de imagens orbitais que deve ser destacado foi a possibilidade, utilizando essas probabilidades, de reconhecer subclasses intermediárias à aquelas identificadas a priori. Os objetos dessas subclasses podem ser utilizados posteriormente para estudos de modelos de mistura espectral, na análise da fragmentação da paisagem, em pesquisas de zonas de transição, dentre outros.

Por fim, embora os resultados sejam significantes e sugerirem que a análise discriminante associada à orientação a objeto pode classificar corretamente os tipos de cobertura vegetal em uma imagem de alta resolução espacial, deve haver cuidado na extrapolação desse modelo discriminante para outras imagens, tendo em vista que

condições atmosféricas e aspectos do pré-processamento, como correções geométricas, podem alterar as características da imagem.

REFÊRENCIAS BIBLIOGRÁFICAS

BATTZ, M.; HOFFMAN, C.; WILLHAUCK, G. Progressing from Object-Based to Object-Oriented Image Analysis. In *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*; Blaschke, T., Lang, S., Hay, G.J., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 29-42.

BLASCHKE, T. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, n. 65, p. 2-16, 2010.

CÂMARA, G.; SOUZA, R.C.M.; FREITAS, U.M.; GARRIDO, J. Spring: Integrating remote sensing and GIS by object-oriented data modeling. *Computers & Graphics*, n. 20(3), p. 395-403, 1996.

CONGALTON, R. G.; GREEN, K. Assessing the accuracy of remotely sensed data: principles and practices. New York: Lewis Publishers, 2009. 137 p.

GONZALEZ, C. R.; WOODS, E. R. *Processamento de Imagens Digitais*. São Paulo: Ed Edgard Blucher, 2000. 332 p.

HANSEN, M.C.; DEFRIES, R.S.; TOWNSHED, J.R.G.; MARUFU, L.; SOHLBERG, R. Development of MODIS tree cover validation data set for Western Province, Zambia. *Remote Sensing of Environment*, v. 83, n. 1-2, p. 195-213, 2002.

HÄRDLE, W.; SIMAR, L. *Applied multivariate statistical analysis*. Berlin: Springer, 2007. 2 ed. 455 p.

JOHNSON, R.A.; WICHERN, D.W. *Applied multivariate statistical analysis*. New Jersey: Pearson Prentice Hall, 2002. 6 ed. 395 p.

LIU, Y.; LI, M.; MAO, L.; XU, F.; HUANG, S. Review of Remotely Sensed Imagery Classification Patterns Based on Object-oriented Image Analysis. *Chinese Geographical Science*. n. 16, p. 282-288, 2006

LOBO, A.; GULLISON, R. E. Mapping the tropical landscape of Beni (Bolivia) from Landsat-TM imagery: beyond the „forest/non-forest“ legend. In F. Dallmeier, & J. A. Comiskey (Eds.), *Forest biodiversity research, monitoring and modeling: Conceptual background and old world case studies* (pp. 159–181). Paris: UNESCO. 1998.

MING, D.; WANG, Q.; LUO, J.; SHEN, Z. Evaluation of high spatial resolution remote sensing image segmentation algorithms. In: *International Congress on Image and Signal Processing*, 2, 2009.

MINGOTI, S.A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte: Editora UFMG, 2005. 295 p.

POLIDORIO, A.M.; TOMMASELLI, A.M.G.; IMAI, N.N.; GALO, M.L.B.T. Segmentação de corpos d'água em imagens multiespectrais e temporais usando watershed com marcadores automaticamente definidos. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 12, 2005, Goiânia. Anais... São José dos Campos: INPE, 2005. Artigos, p. 4249-4257. CD-ROM, On-line. ISBN 85-17-00018-8. Disponível em: <http://martem.dpi.inpe.br/col/ltid.inpe.br/sbsr/2005/02.12.16.31/doc/@sumario.htm>. Acesso em: 01 out. 2012.

STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES (SPSS). Versão 19.0. Chicago: IBM Company, 2010.

THENKABAIL, P. S.; ENCLONA, E. A.; ASHTON, M. S.; LEGG, C.; DE DIEU, M. J. Hyperion, IKONOS, ALI, and ETM plus sensors in the study of African rainforests. *Remote Sensing of Environment*, n. 90, p. 23–43, 2008.

THESSLER, S.; SESNIE, S.; BENDANA, Z.S.R.; ROUKOLAINEN, K.; TOMPPO, E.; FINEGAN, B. Usin K-nn and discriminant analyses to classify rain forest types in Landsat TM image over northern Costa Rica. *Remote Sensing of Environmental*, n. 112, p. 2485-2494, 2018.

VAN DER HEIJDEN, F. Image based measurement systems: object recognition and parameter estimation. United Kingdom: John Wiley and Sons, 1994. 293 p.

VINCENT, L.; SOILLE, P. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. n. 13, p. 583-598, 1991.

ZHANG, X; FRIEDL, M.A.; SCHAAF, C.B.; STRAHLER, A.H.; HODGES, J.C.F.; GAO, F.; REED, B.C.; HUETE, A. Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, v. 84, p. 471-475, 2003.