

MÉTODOS PARA REDUÇÃO DO ESPAÇO N-DIMENSIONAL DE IMAGENS ORBITAIS

N-dimensional methods space reduction for orbital images

Thomaz Corrêa e Castro da Costa¹
Paulo De Marco²
Ricardo Seixas Brites³

¹Embrapa Milho e Sorgo
Rod. MG 424, km 45 – CEP 35701-970, Sete Lagoas - MG
thomaz@cnpes.embrapa.br

²Universidade Federal de Viçosa-UFV
Departamento de Biologia
Av. PhRofs s/n – CEP 37000-000, Viçosa - MG
pdmarco@ufv.br

³Universidade Católica de Brasília-UCB
SGAN 916 - Módulo B - Asa Norte - CEP 70790-160, Brasília - DF
rbrites@pos.ucb.br

RESUMO

Imagens orbitais contém informação redundante nas bandas espectrais e, no caso dos sensores hiperespectrais, tem também alta dimensionalidade (grande numero de bandas), dificultando seu processamento. Métodos de redução do espaço n-dimensional tem a finalidade de reduzir a dimensionalidade e a redundância em dados orbitais. Neste trabalho os principais métodos de redução do espaço n-dimensional foram comparados por meio de um estudo de caso com imagem Landsat TM, visando a avaliação destes métodos com relação a perda de informações que afetam a exatidão da classificação temática. Foram aplicados: seleção de bandas pela medida de separabilidade “divergência transformada”, Análise de informação mútua e índice V de Cramer; e transformação do espaço n-dimensional por Análise discriminante canônica, Análise de componentes principais, e Tasseled Cap. Os resultados indicaram que o índice V de Cramer se mostrou um método potencial para seleção de bandas. Já a análise de informação mútua apresentou resultados não satisfatórios; e a transformação por análise discriminante canônica foi sensivelmente superior as demais transformações.

Palavras chaves: Análise de informação mútua, entropia, análise canônica, Cramer’s V, divergência transformada

ABSTRACT

Orbital images contain redundancy information in spectral bands and, in hyperspectral sensors, with grate dimensionality (grate number of bands), difficult processing. In methods by reduction date the objective is reduce the dimensionality and redundancy at orbital dates. In this work the principal methods for reduction of the data dimensionality were compared using Landsat TM image for application case. The objective was available these methods with relation the information loss that affect the classification accuracy. The methods were: bands selections by divergence index, Information Analysis Mutual and Cramer’s V index, and Transformation of the n-dimensional space by Canonical Analysis, Principal Components analysis, and Tasseled Cap. The results indicated which Cramer’s V is the potential index by bands selection. The information analysis mutual was unstabled, selecting worst band combination and the transformation for canonical analysis was better midst transformations.

Keywords : Mutual Information Analysis, Entropy, Canonical Analysis, Cramer’s V, Transformed Divergence

1. INTRODUÇÃO

O conjunto ideal de dados para classificação/interpretação de imagens orbitais é aquele que permite a melhor separabilidade entre classes no espaço n-dimensional. Como dados de sensoriamento remoto apresentam redundância entre bandas e, no caso de sensores hiperespectrais, a alta dimensionalidade dos dados pode inviabilizar o processo de classificação (VENKATARAMAN, 2005), existem técnicas que visam a redução do espaço n-dimensional de imagens orbitais sem que isto afete consideravelmente a separabilidade.

Umam atuam na seleção de bandas, e outras transformam o espaço n-dimensional para um novo espaço, por meio de funções lineares, concentrando a maioria das informações nos primeiros componentes. Dentre as técnicas mais conhecidas estão os índices de separabilidade, a análise de componentes principais, análise de variáveis canônicas e Transformação de Tasseled Cap (THOMAS et al, 1987; SWAIN and DAVIS, 1978; JENSEN, 1996; BRITES, 1996; SPANNER et al, 1984).

Dentre as técnicas abordadas para redução da dimensionalidade em dados hiperespectrais, que se baseiam na otimização de funções critérios, que podem ser índices de separabilidade, uma bem conhecida é a “Sequential Forward Selection” (SFS), pela simplicidade computacional e conceitual, que consiste na inclusão de feições (bandas) até o alcance da máxima separabilidade. Outras técnicas, como a Steepest Ascent (SA), Fast Constrained Search (FCS), Projection Pursuit (PP) e Decision Boundary Feature Extraction (DBFE) propõe-se a melhorar a performance comparada a SFS (SERPICO et al., 2003).

Com relação aos índices de separabilidade, Mausel et al. (1990), citado por JENSEN (1996), avaliou quatro medidas de separabilidade para determinar qual delas poderia identificar mais precisamente a melhor subsérie de quatro das oito bandas originais para classificar seis classes de uso agrícola. Divergência transformada e a medida de distância de Jeffries-Matusita (JM) selecionaram a mesma subsérie de bandas que produziu a maior exatidão total. JENSEN (1996), com base nestes resultados, recomenda como norma geral o uso da divergência transformada.

Kavzoglu & Mather (2000) experimentaram as medidas de separabilidade Divergência, Divergência transformada, Distância de Battacharyya, Jeffries-Matusita, e Distância de Mahalanobis para a seleção de 8 bandas em um total de 24 bandas multisensores, na exatidão de classificações aplicando redes neurais, encontrando performances similares entre TD e J-M.

Com o lançamento dos sensores hiperespectrais, estas técnicas foram novamente acessadas na tentativa de solucionar o grande problema de dimensionalidade destes dados. ZORTEA et al. (2005) utilizou a distância média de Jeffries-Matusita (coeficiente exponencial de Battacharyya) para definir

partições, na redução da dimensionalidade de bandas espectrais do pixel, para segmentos contíguos, expressos pelas estatísticas média e dispersão, em algoritmo s top-down.

A transformação Tasseled Cap tem sido efetivamente usada para reduzir volume de dados e aumentar a interpretação de características físicas de classes na cena (Kauth et al., 1976; Crist, 1983 and 1984; Crist et al., 1983, 1984 a,b,c; Crist et al., 1986), em classificação não supervisionada (Wheeler, 1989), interpretação de contrastes espectrais em cenas urbanas (Goward et al., 1984), avaliação de risco de fogo em floresta (Cohen, 1989), e detecção de mudança em cobertura da terra (Malila, 1979; King et al., 1987) citados por LEE e PRICE (1994).

Uma metodologia denominada de Análise de Informação Mutua (AIM) (KULLBACK, 1959), não restrita a dados paramétricos, e não usual em sensoriamento remoto, também foi aplicada para seleção de bandas (CONESE e MASELLI, 1993). AIM utiliza o conceito de entropia, que permite quantificar a informação expressa pela amostra de treinamento (“verdade de campo”) sobre os dados de uma imagem, selecionando-se as bandas que contém a maior informação mútua com esta amostra.

A AIM vem sendo utilizada em diversos estudos para avaliar a complexidade de associações ecológicas (PHIPPS, 1981; DAVIS, 1990; Michaelsen et al., 1986 citado por CONESE e MASELLI, 1993). Foi aplicada também ao sensoriamento remoto para avaliar exatidão de imagens “soft” classificadas (FOODY, 1995; MASELLI et al., 1994).

CONESE e MASELLI (1993) aplicaram AIM para verificar a influência das condições ambientais na seleção de bandas. A combinação padrão (TM 3, TM4 e TM5) foi comparada às bandas escolhidas pela AIM em três áreas distintas, significando que as bandas selecionadas variaram de área para área, evidenciando a interferência do local para seleção de bandas. Os kappas obtidos com as bandas selecionadas pelo AIM foram significativamente superiores aos obtidos com as bandas padrão.

Neste trabalho o objetivo foi o de investigar métodos da literatura clássica de redução da dimensionalidade de dados orbitais, utilizando como estudo de caso uma redução de uma imagem Landsat TM para 3-dimensional, avaliada pela exatidão da classificação gerada.

2. Métodos para a Redução do Espaço n-dimensional de Imagens Orbitais

2.1 Medidas de Separabilidade

Índices de separabilidade tem duas finalidades principais: (1) Identificar pares de classes informacionais com pobre separabilidade em um conjunto de bandas; e (2) Recomendar uma redução no número de bandas para interpretação, sem prejudicar a separabilidade entre as classes. Os índices de

separabilidade mais conhecidos são: Distância Euclidiana (1)

$$E_{ij}^2 = (M_i - M_j)^T (M_i - M_j) \quad (1)$$

sendo M = vetor de médias da classe para n bandas.

que é uma medida de distância entre médias das classes i e j, e não considera em seus cálculos os parâmetros de dispersão, que expressam a extensão da distribuição das classes no espaço n-dimensional (THOMAS et al, 1987).

Os outros índices incluem em suas fórmulas matrizes de variância covariância, tendo como pressuposição a distribuição Gaussiana das classes.

A Distância de Mahalanobis (2), uma medida transformada de E, considera o parâmetro dispersão das classes (matrizes de variância covariância).

$$M_{ij} = (M_i - M_j)^T (\Sigma_i^{-1} + \Sigma_j^{-1})(M_i - M_j), \quad (2)$$

sendo Σ^{-1} =matriz inversa de variância-covariância para classe.

As medidas de separabilidade mais utilizadas em sensoriamento remoto são a Divergência, Distância de Jeffries-Matusita e Distância de Bhattacharyya, que foram desenvolvidas a partir das integrais da razão, da diferença e do produto, respectivamente, das funções de verossimilhança para as classes i e j, e suas fórmulas e deduções estão descritas em SWAIN e DAVIS (1978); Kailath (1967) citado por THOMAS et al. (1987); e BRUZZONE, L. et al. (1995).

A divergência é uma medida de separabilidade usual em sensoriamento remoto, deduzida da razão entre funções de verossimilhança para as classes, dada pela expressão (3):

$$L(ij / X) = \frac{|\Sigma_j|^{1/2} \exp[-1/2(X - M_i)^T \Sigma_i^{-1}(X - M_i)]}{|\Sigma_i|^{1/2} \exp[-1/2(X - M_j)^T \Sigma_j^{-1}(X - M_j)]}, \quad (3)$$

sendo X=vetor de pixel a ser classificado.

A dedução da divergência a partir de $L(ij/X)$ pode ser encontrada em SWAIN e DAVIS (1978) e THOMAS et al. (1987), chegando-se a expressão (4)

$$D_{ij} = \frac{1}{2} \text{tr}(\Sigma_i - \Sigma_j)(\Sigma_j^{-1} - \Sigma_i^{-1}) + \frac{1}{2} \text{tr}(M_i - M_j)^T (\Sigma_i^{-1} + \Sigma_j^{-1})(M_i - M_j) \quad (4)$$

Note que E e M são níveis de simplificação de D. Como D cresce indefinidamente, mesmo após a completa separabilidade entre classes ter sido alcançada, foi desenvolvido o índice Divergência Transformada, com a finalidade de estabelecer um valor assíntoto da função quando atingida a separabilidade total. Este índice, apresentado em SWAIN e DAVIS (1978), é dado pela seguinte expressão (5):

$$DT_{ij} = a[1 - \exp(D_{ij}/b)], \quad (5)$$

sendo a e b constantes escolhidas pelo usuário para avaliação de D.

Quando se deseja escolher a combinação de bandas que melhor separe todas as classes, calcula-se *DTmédia* a partir da expressão (6):

$$DT_{media} = \sum_{i=1}^M \sum_{j=1}^M p(i)p(j)DT_{ij} \quad (6)$$

2.2. Índice V de Cramer (OTT et al., 1983)

O índice V de Cramer é uma extensão da estatística Qui-quadrado, aplicado a tabelas de contingência com dados categóricos, para testar a relação de dependência entre variáveis nominais. Sua fórmula é dada por (7)

$$V \text{ de Cramer} = X^2/N(k-1), \quad (7)$$

Sendo:

X^2 a estatística Qui-quadrado entre duas variáveis;

N o total de observações;

k o menor valor entre linhas e colunas.

Com esta estrutura este índice pode também ser empregado para medir a associação entre as amostras de treinamento com as bandas de uma imagem orbital, ou seja, o cruzamento de dados informacionais com dados espectrais. Seu resultado é uma saída do módulo CROSSTAB do Software IDRISI (EASTMAN, 2003).

2.3. Análise de Informação Mutua (AIM)

Para compreender a AIM, considere uma população, no caso uma cena orbital, com o atributo "informação espectral", que é uma variável discreta, Y, com níveis M (valores de brilho ou níveis de cinza). A entropia desta cena com respeito a Y, é calculada pela expressão (8):

$$H(Y) = - \sum_{i=1}^M p(i) \ln p(i) \quad (8)$$

sendo p(i)= proporção de Y no nível de valor de brilho i.

Para uma cena com apenas um valor de brilho (um nível, 255 por exemplo), H(Y)=0, ou seja, a menor quantidade de informação na cena. Quanto maior os níveis e a equitatividade entre as frequências de pixel, maior será a entropia. A informação em termos absolutos é dada por (9)

$$I(Y)=N*H(Y) \quad (9)$$

sendo N = número de elementos da população, que neste caso é o total de pixels na imagem.

Esta é a avaliação de entropia para apenas uma variável, valor de brilho. Para selecionar bandas com maior informação mútua entre valores de brilho e dados reais é preciso desenvolver a entropia conjunta $H(Y, X)$, sendo a variável X , as classes informacionais (amostras de treinamento) com níveis L , dada pela fórmula (10), que é uma extensão da expressão 8 para distribuição conjunta entre X e Y .

$$H(Y, X) = - \sum_{i=1}^M \sum_{j=1}^L p(i, j) \ln p(i, j) \quad (10)$$

$p(i, j)$ é a proporção conjunta do nível i da variável Y com nível j da variável X .

Enquanto a entropia é uma forma de mensurar a complexidade do sistema, a informação mútua tem como objetivo identificar o nível de ordenamento deste sistema, com relação as variáveis envolvidas. A idéia é encontrar alguma ordem da informação Y controlada pela variável X , $p(Y/X) \neq p(Y)$, que significa dizer que existe um grau de diferenciação na informação espectral entre as classes de treinamento, no caso.

A dedução da informação mútua $I(Y, X)$, a partir de $H(Y, X)$, é dada em PHIPPS (1981), chegando-se a expressão (11)

$$I(Y, X) = N \ln N + \sum_{i=1}^M \sum_{j=1}^L F(i, j) \ln F(i, j) - \sum_{i=1}^M F(i, +) \ln F(i, +) - \sum_{j=1}^L F(+, j) \ln F(+, j) \quad (11)$$

sendo $F(i, j)$ =freqüência de Y e X com níveis i e j , $F(i, +)$; $F(+, j)$ =freqüências marginais dos níveis i e j respectivamente. Estes resultados são obtidos a partir da construção de uma tabela de contingência entre duas variáveis de interesse (Y, X).

Se Y e X são variáveis independentes, ou seja, não existe nenhum tipo de associação entre elas, $I(Y, X)=0$, enquanto que se Y e X são perfeitamente associados $I(Y, X)=I(Y)$ ou $I(Y, X)=I(X)$ (CONESE e MASELLI, 1993).

Desta forma pode-se medir a quantidade de informação mútua, IM , pela razão (12)

$$IM = I(Y, X) / I(X) \quad (12)$$

significando dizer que quanto maior a IM , maior a quantidade de informação associada entre X e Y . Focando esta análise para o problema de seleção de bandas, quanto maior for a associação da informação espectral com as classes de treinamento, maior será a entropia da banda analisada.

2.4 Transformação por Análise de Componentes Principais

A Análise de Componentes Principais (ACP) é um procedimento usual em processamento de imagens. A partir das bandas originais, correlacionadas entre si, onde estão dispostos os dados em um espaço n -

dimensional são calculados, a partir da matriz de variância-covariância do espaço multibandas, os autovalores (variâncias) e autovetores (coeficientes), para obter funções lineares ortogonais de forma a rotacionar os dados, retendo o maior número de informações para cada componente gerado, eliminando-se a redundância (MANLY, 1994). O primeiro principal componente agrupa a maior parte das informações, apresentando a maior variância (autovalor). Geralmente os três primeiros componentes conseguem agrupar mais que 95 % da variação dos dados, utilizando-se apenas estes para fins de interpretação. A expressão para o cálculo dos autovalores e autovetores é dada por (14)

$$(S-LI)D=0 \quad (14)$$

sendo S a matriz de variância-covariância das bandas; L uma matriz diagonal de I (autovalores), I a matriz identidade e D um vetor de d (autovetores).

2.5. Transformação pela Análise de Função Discriminante Canônica (ADC)

Assim como em ACP, o objetivo desta transformação é determinar combinações lineares das bandas originais que retenham a maior quantidade de informação nos primeiros componentes, deixando para os últimos componentes o ruído.

A principal diferença entre esta análise e a ACP é que, para ADC, os coeficientes das combinações lineares de n bandas (Z_i) rotacionam os eixos na direção que permite a maior separabilidade de classes (Figura b), enquanto a ACP rotaciona os eixos nas direções dos dados correlacionados de todas as feições presentes na cena. Os coeficientes da ADC são calculados a partir dos dados das amostras das classes (Figura a), enquanto que para ACP os coeficientes são calculados com os dados totais da cena.

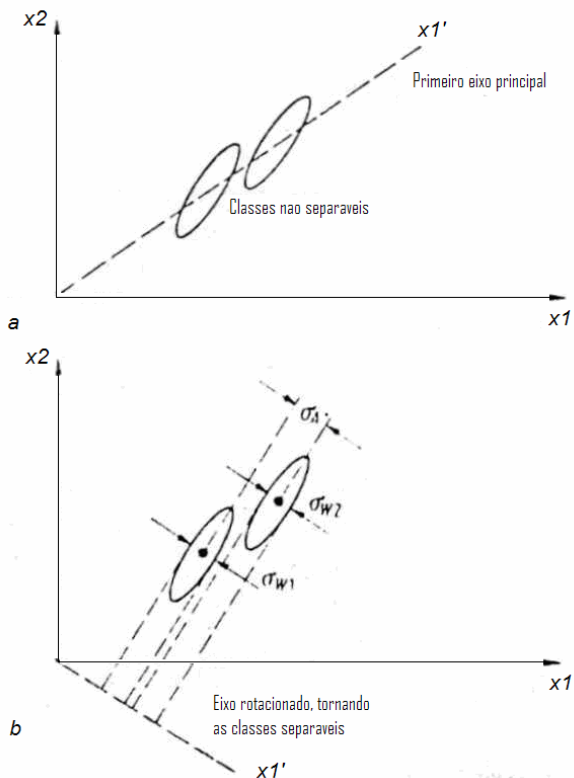


Figura. a) Classes hipotéticas não separáveis considerando os eixos das bandas originais, no espaço 2 dimensional; b) Classes separáveis após a rotação do eixo 1. Fonte: RICHARDS (1986).

A forma de escolha dos coeficientes de rotação dos eixos, d_{11} , d_{12} , ..., d_{1p} , é pela maximização de F obtida pela análise de variância para as classes (MANLY, 1994; RICHARDS, 1986) que, pelo exemplo da Figura, seria dado por $F = \frac{\mathcal{S}_A^2}{\mathcal{S}_w^2}$.

A primeira função, $Z_1 = d_{11}B_1 + d_{12}B_2$, considerando apenas 2 eixos, resulta no maior F gerado pela razão da variância entre grupos com a variância dentro dos grupos. A segunda função, $Z_2 = d_{21}B_1 + d_{22}B_2$, gera o segundo maior F , sendo Z_2 ortogonal (não correlacionado) com Z_1 , e assim sucessivamente até que a última função canônica seja gerada.

Os coeficientes da função discriminante canônica são obtidos da seguinte maneira (RICHARDS, 1986):

Sabendo que as matrizes de variância covariância intra e entre classes são dadas por $\mathcal{S}_w^2 = d' \mathcal{S}_w^v d$; e $\mathcal{S}_A^2 = d' \mathcal{S}_A^v d$, respectivamente, a maximização de F em relação a d é dada por

$\nabla F / \nabla d = 0$, resultando na expressão: $(\mathcal{S}_A - \lambda \mathcal{S}_w) d = 0$, ou na sua forma matricial (15):

$$(\mathcal{S}_A - \mathcal{L} \mathcal{S}_w) D = 0 \quad (15)$$

sendo \mathcal{L} uma matriz diagonal de \mathcal{I} e D um vetor de d .

Uma última condição exige que $D' \mathcal{S}_w D = 1$. Em RICHARDS (1986) é dado um exemplo numérico para cálculo dos coeficientes. Assim como em análise de componentes principais, são geralmente necessários até os três primeiros componentes.

2.6. Transformação Tasseled Cap

A técnica de compressão de dados "Tasseled Cap" surgiu como um índice de vegetação desenvolvido por Kauth e Thomas (1976), citados por JENSEN (1996), que foi construído por técnicas de ortogonalização seqüencial de Gram-Schmidt. Tasseled Cap transforma o espaço 4-dimensional original do MSS em um novo espaço ortogonal de quatro dimensões.

Posteriormente, Crist e Cicone (1984), citados por JENSEN (1996), verificaram que as seis bandas refletidas do TM também são dispersas em um espaço 3-dimensional, definindo dois planos perpendiculares e uma zona de transição entre os dois.

Na Tasseled Cap, áreas completamente vegetadas definem o "plano de vegetação", enquanto solos expostos definem o "plano de solos". Entre estes dois existem áreas parcialmente vegetadas, nas quais a vegetação e os solos são visíveis. A partir destes planos, são obtidos três componentes relevantes: (1) verdor (greenness); (2) brilho (brightness); e (3) umidade (wetness), e outros três descartados para fins de interpretação.

Um questionamento sobre o índice Tasseled Cap é com respeito a utilização dos seus coeficientes, gerados localmente, e que são recomendados para uso global (Huete et al., 1984; Ezra et al. 1984 citados por LEE e PRICE, 1994). JACKSON (1983) desenvolveu um procedimento para cálculo de coeficientes para locais específicos, que conforme resultados em LEE e PRICE (1994) não gerou exatidões significativamente melhores que resultados gerados por coeficientes globais da Transformação Tasseled Cap.

3. MATERIAL E MÉTODOS

A área está situada na microrregião de Viçosa-MG, com predominância do relevo ondulado. A tipologia vegetal é a Floresta Estacional Semidecidual, sob domínio da Mata Atlântica.

Foi utilizada uma subcena de 15 x 15 km, extraída da imagem Landsat5-TM, Órbita 217 Ponto 74, Quadrante Sul, de 10/10/94, composta de 5 bandas espectrais: TM1, 2, 3, 4, 5 e 7. A correção geométrica da imagem foi feita com pontos de controle coletados em OrtofotoCartas da CEMIG, Escala 1:10.000, utilizando-se a função de mapeamento quadrática, com reamostragem pelo vizinho mais próximo, atingindo um RMS < 15 metros.

A referência da área (Verdade de Campo) foi obtida com o apoio de 14 fotos aéreas não convencionais (duas câmeras Hasselblad mod. 553 ELX, com objetivas Zeiss Distigon 4/50 mm T, não métricas), na escala aproximada de 1:15.000.

A amostra de referência foi dimensionada pela fórmula da distribuição binomial: $n = Z^2 pq / E^2$, fixando-se os seguintes parâmetros: $p = 0,76$, exatidão estimada por amostragem piloto, $q = 1 - p$; $Z \sim N(0, 1) = 1,96$ para 95% de confiança, e $E = 0,05$, erro de amostragem,

resultando em n=280, mais 29 pontos excedentes, totalizando 309 pixels. O Quadro 1 mostra o número de pixels utilizados para treinamento e para referência.

QUADRO 1-Descrição das Classes de UT/CT, e amostras de treinamento e referência por classe.

CLASSE	Pixels Treinam	Pixels Refer.
1.AGRIC: Area agrícola, envolv. solo arado, e cult. Anuais em difer. Estágios de desenv.	233	27
2.CAFE: Cafeicultura, com expos. do solo	363	11
3.EUCAL: Eucaliptais em difer idades e tipos de manejo	248	11
4.LAGOA: Lagoas, geral. turvas e de peq. Dimensões	60	6
5.MTSEC: Mata Secundária em Estágio Médio a Avançado	534	65
6.PASTO: Difer tipos de pasto, Braquiária, Colômbio, limpos, sujeitos com predom. de Assa-peixe	611	115
7.REGEN: Mata Secundária em estágio inicial (regeneração)	71	21
8.SOLEX: solo exposto	66	4
9.URBAN: Área Urbana (Alta e média densidade de casas)	523	21
10.VGRIB: vegetação ribeirinha (maioria gramíneas)	59	9
11.SOMBR: sombras, de grande ocorrência devido a topografia acidentada da região	228	19
TOTAL	2996	309

O algoritmo de classificação adotado foi o de máxima verossimilhança e, para avaliar os resultados das classificações, utilizou-se o índice Kappa (K). Para testar a diferença entre os índices de exatidão (K_1 e K_2), o teste estatístico utilizado foi o de Z, conforme a expressão (16):

$$Z = \frac{\hat{K}_2 - \hat{K}_1}{\sqrt{V(\hat{K}_2) + V(\hat{K}_1)}}, \quad (16)$$

sendo (HUDSON e RAMM, 1987):

$$\hat{K} = \frac{q_1 - q_2}{1 - q_2}$$

$$v(\hat{K}) = \frac{1}{n} \left[\frac{q_1(1-q_1)}{(1-q_2)^2} + \frac{2(1-q_1)(2q_1q_2 - q_3)}{(1-q_2)^3} + \frac{(1-q_1)^2(q_4 - 4q_2^2)}{(1-q_2)^4} \right]$$

e

$$q_1 = \frac{1}{n} \sum_{i=1}^M n_{ii}$$

$$q_2 = \frac{1}{n^2} \sum_{i=1}^M n_{i+} n_{+i}$$

$$q_3 = \frac{1}{n^2} \sum_{i=1}^M n_{ii} (n_{i+} + n_{+i})$$

$$q_4 = \frac{1}{n^3} \sum_{j=1}^M n_{ij} (n_{j+} + n_{+i})^2$$

com n=número de amostras de referência; i=linha=coluna da matriz de confusão; e para o termo n_{j+} em q_4 , j=linha.

Os métodos empregados para redução de dados foram:

- Divergência Transformada, Índice V de Cramer e Análise de Informação Mutua, que atuam na seleção de bandas;
- Componentes Principais, Análise Discriminante Canônica, Transformação Tasseled Cap, que transformam o espaço n-dimensional para um novo espaço n-dimensional, por meio de funções lineares, concentrando a maioria das informações nos 3 primeiros componentes.

Divergência Transformada, Índice V de Cramer, Análise de Informação Mutua foram calculados em planilhas Excel com o auxílio de macros em visual basic, a partir de dados exportados do SIG IDRISI. A Análise Discriminante Canônica foi executada no Software Statistic for Windows, e após obter os coeficientes das funções lineares os mesmos foram aplicados às respectivas bandas por meio de módulos de adição e multiplicação entre imagens no IDRISI, gerando assim os novos componentes por transformação canônica.

E análise de componentes principais e transformação Tasseled Cap foram calculadas por meio de módulos do IDRISI.

Para a Análise de Informação Mútua não foi usado um mapa digitalizado de classes informacionais para gerar a informação mútua, mas os próprios polígonos de treinamento. Tinha-se como hipótese que as bandas que apresentassem maiores índices de informação mútua com polígonos de treinamento representariam as melhores informações para discernir entre as classes de treinamento. Isso simplificaria a aplicação do método. É preciso uma amostra razoavelmente grande de polígonos de treinamento, e que as amostras envolvidas representem fielmente o universo de classes daquela cena.

Para cada banda foi gerada uma planilha de frequência de valores de brilho por classe e efetuadas as operações para cálculo da entropia e entropia conjunta. A informação mútua foi calculada com relação a entropia das classes.

4. RESULTADOS E DISCUSSÃO

4.1. Métodos para seleção de bandas

A Divergência Transformada indicou a combinação das bandas 3, 4 e 5 como a de melhor separabilidade, com DTmédia próximo a DTmédia para seis bandas (Quadro 2).

QUADRO 2 - Hierarquia de Valores de Divergência Transformada Média (para todas as classes) considerando as seis bandas e todas as combinações possíveis de três bandas

Bandas	Dtmed	Bandas	Dtmed	Bandas	Dtmed
123457	13.87	247	12.61	457	11.94
345	13.59	347	12.59	147	11.91
245	13.46	257	12.58	123	11.84

357	12.95	157	12.22	137	11.70
135	12.77	234	12.21	127	11.60
145	12.69	125	12.18	235	11.02
134	12.65	124	12.03	237	10.32

Os índices de associação da informação entre classes e bandas gerados por Cramer's V e AIM são apresentados no Quadro 3. Os maiores valores para o Índice de Cramer's V foram obtidos também para as bandas 3, 4 e 5.

A Análise de Informação Mútua elegeu as bandas 2, 3 e 5, descartando a 4 com baixo valor de entropia com a amostra de treinamento, um resultado não esperado, pois a combinação 235 é a penúltima do ranking de divergência transformada para discriminação entre as classes.

Na tentativa de descobrir porque o método falhou para a banda 4, foram analisadas as matrizes para bandas 1 e 4. Verificou-se que a não padronização da amplitude dos valores de brilho entre bandas pode afetar o resultado. A banda 1 sofre o efeito de espalhamento da atmosfera, e também apresenta histograma de pequena amplitude, ou seja as classes apresentam pouca distinção entre si, mas na matriz da entropia seus valores de brilho apresentam alta frequência para as classes, mesmo ocorrendo mistura entre os valores de brilho. Estas frequências geram altos valores de $F(i,j) \cdot \ln F(i,j)$ na matriz conjunta, majorando a entropia conjunta. Por outro lado a banda 4 apresenta grande amplitude, e com frequências mais baixas, resultando em baixos valores de frequência conjunta. Desta forma a falta de padronização entre a amplitude das bandas não gera resultados realísticos de entropia conjunta.

QUADRO 3 - Índice Cramer's V e Análise de Informação Mútua para as bandas TM

	Cramer's V	AIM=I(X,Y)/I(X)
B1	0.09	0.46
B2	0.11	0.48
B3	0.16	0.51
B4	0.19	0.27
B5	0.27	0.50
B7	0.13	0.40

4.2. Métodos para transformação de bandas

A transformação n-dimensional justifica-se quando ocorre redundância nos dados. No Quadro 4 é mostrado o alto grau de correlação entre as bandas entre as bandas 1 e 2 (0,89), 1 e 3 (0,88) e 2 e 3 (0,95).

QUADRO 4- Matriz de Correlação entre as Seis Bandas TM

	B1	B2	B3	B4	B5	B7
B1	1.00					
B2	0.89	1.00				
B3	0.88	0.95	1.00			

B4	0.22	0.33	0.24	1.00		
B5	0.74	0.76	0.81	0.24	1.00	
B7	0.65	0.66	0.70	0.03	0.76	1.00

Com a aplicação da análise de componentes principais, os três primeiros componentes (C1, C2 e C3) garantiram 97,02% de toda a variação dos dados conforme apresentado no Quadro 5.

QUADRO 5 -Autovalores (Variâncias) e Percentagens Acumuladas dos Componentes Principais

Componente	Var	% var.	% var.acum.
C1	1387.69	86.58	86.57
C2	102.55	6.40	92.97
C3	65.00	4.05	97.02
C4	42.22	2.63	99.65
C5	3.28	0.20	99.85
C6	2.28	0.14	100.00

Para a transformação Canônica foram calculados os coeficientes (Quadro 6) para gerar cada componente.

QUADRO 6- Coeficientes para Transformação Canônica

Componentes	B1	B2	B3	B4	B5	B7
C1	0.174	0.232	0.182	-0.323	0.365	0.148
C2	0.286	0.471	0.478	-0.003	-1.524	0.107
C3	0.387	0.129	0.098	1.005	-0.381	-0.147
C4	1.613	2.167	-4.674	0.095	0.535	0.563
C5	1.898	-4.998	3.046	0.528	-1.147	1.071
C6	1.862	-1.644	0.606	-0.333	1.155	-1.909

E para gerar a transformação Tasseled Cap foram utilizados coeficientes globais, conforme apresentado no Quadro 7.

QUADRO 7 - Coeficientes Globais para Transformação Tasseled Cap (Crist et al., 1986)

Componentes	B1	B2	B3	B4	B5	B7
BRILHO	0.291	0.249	0.481	0.557	0.444	0.171
VERDOR	-0.273	-0.217	-0.551	0.772	0.073	-
UMIDADE	0.145	0.176	0.332	0.340	-	0.165
C4	-0.846	-0.073	-0.464	-0.003	0.621	0.419
C5	0.055	-0.023	0.034	-0.194	0.049	0.012
C6	0.119	-0.807	0.409	0.057	0.416	-
					-	0.782
					0.023	0.022

Nos Quadros 8 e 9 são apresentados os resultados finais de exatidão para os respectivos métodos de redução de dados. Sem considerar prova estatística, a combinação padrão 345, que foi escolhida pelo índice de divergência transformada e o índice Cramer's V, apresentou o melhor resultado de exatidão medido pelo índice Kappa. Esta combinação é geralmente considerada para descrever o espaço Euclidiano tridimensional (Crist and Cicone, 1984 citados por CONESE e MASELLI, 1993) e tem provado que são as bandas mais informativas para estudos de vegetação, sendo portanto indicadas como padrão (Horler e Ahern, 1986 citados por CONESE e MASELLI, 1993).

O segundo melhor resultado foi gerado com dados transformados por análise discriminante canônica, um resultado esperado na comparação com a transformação por Tasseled Cap e Componentes Principais, pois os coeficientes da análise canônica são construídos de forma a otimizar a discriminação entre classes, usando-se os próprios dados de treinamento.

Por último, a combinação TM235, selecionada pelo método da análise de informação mútua, teve o pior desempenho, com um Kappa de 0,578.

Considerando o teste estatístico Z, apresentado no Quadro 9, as diferenças significativas ocorreram na comparação entre TM235 e TM345, CC123, TasCap, sendo os demais resultados considerados estatisticamente iguais.

QUADRO 8 - Índice Kappa e sua Variância para as diferentes estratégias de seleção de dados para interpretação de imagens orbitais

Combinações de Bandas	Kappa	Var(Kappa)
TM345 (Diverg. Transf. e Índice Cramer's V)	0.7056	0.00082531
Comp. Canônicos C1, C2 e C3	0.6996	0.00082976
Tasseled Cap Brilho, Verdor e Umidade	0.6757	0.00086829
Comp. Principais C1, C2 e C3	0.6566	0.00089404
TM235 (AIM)	0.5780	0.00093675

QUADRO 9 - Teste Z para significância da exatidão entre estratégias de seleção de dados para interpretação de imagens orbitais

	Zcal	Zcal > 1,96 alfa 5%
TM345 vs CC123	-0.14748337	Ns
CC123 vs TasCap	-0.57999288	Ns
TasCap vs CP123	-0.45497732	Ns
CP123 vs TM235	-1.83697515	Ns
TM345 vs TasCap	-0.72655043	Ns
CC123 vs CP123	-1.03567864	Ns
TasCap vs TM235	-2.2995939	*
TM345 vs CP123	-1.18171823	Ns
CC123 vs TM235	-2.89318047	*
TM345 vs TM235	-3.03976723	*

4. CONCLUSÃO

O índice de Cramer's V atendeu a expectativa de seleção de bandas para a melhor separabilidade entre classes, por apresentar resultado idêntico ao gerado pela Divergência transformada, um índice de separabilidade usual para selecionar bandas.

Análise de informação mútua gerou resultados instáveis neste trabalho, não sendo indicada como um bom método para selecionar bandas.

A transformação por componentes canônicos embora com uma etapa adicional para cálculo dos coeficientes, apresenta resultados sensivelmente melhores que a análise por componentes principais e a transformação Tasseled Cap, devido a especificidade dos coeficientes gerados.

AGRADECIMENTOS

Este trabalho foi apoiado pelo Conselho de Desenvolvimento Científico e Tecnológico – CNPq no programa de Pós-graduação em Ciência Florestal na Universidade Federal de Viçosa.

REFERÊNCIAS BIBLIOGRÁFICAS

- BRITES, R. S., SOARES, V. P., RIBEIRO, C. A. A. S. Comparação de desempenho entre três índices de exatidão aplicados a classificações de imagens orbitais. In: VIII Simpósio Brasileiro de Sensoriamento Remoto, Salvador, 1996, **Anais. INPE**, 1996. (CD-ROM).
- BRUZZONE, L.; ROLI, F.; SERPICO, S. B. Na extension of the Jeffreys-Matusita Distance to multiclass cases for feature selection. **IEEE Transactions on Geoscience and Remote Sensing**, v. 33, n. 6, p. 1318-1321. 1995
- CONESE, C.; MASELLI, F. Selection of optimum bands from TM scenes through mutual information analysis. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 48, n. 3, p. 2-11. 1993.
- DAVIS, F. W.; DOZIER, J. Information analysis of a spatial database for ecological land classification. **Photogrammetric Engineering and Remote Sensing**, v. 56, n. 5, p. 605-613. 1990.
- EASTMAN, J. R. IDRISI Source Code 1987-2003. Worcester: Clark University, 1987-2003. 1 v.
- FOODY, G. M. Cross-entropy for the evaluation of the accuracy of a fuzzy land cover classification with fuzzy ground data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 50, n. 5, p. 2-12. 1995.
- HUDSON, W. D.; RAMM, C. W. Correct formulation of the Kappa coefficient of agreement. **Photogrammetric Engineering and Remote Sensing**, v. 53, n. 4, p. 421-422. 1987.
- JACKSON, R. D. Spectral indices in n-space. **Remote Sensing of Environment**, v. 13, p. 409-421. 1983.

- JENSEN, J. R. **Introductory digital image processing- a remote sensing perspective**. Englewood Cliffs: Prentice-Hall, 1996. 379p.
- KAVZOGLU, T.; MATHER, P.M. The use of feature selection techniques in the context of artificial neural networks. In 2000 Remote Sensing Society. Annual Conference, Leicester, UK, **Proceedings** p. 12-14. 2000.
- KULLBACK, S. **Information theory and statistics**. Wiley, New York, NY. 395p.1959.
- LEE, Re-Y.; PRICE, K. P. Effects of TM Tasseled Cap transformations and edge enhancement on land use/land cover classification accuracy. In: 1994 ASPRS/ACSM ANNUAL CONVENTION EXPOSITION TECHNICAL PAPERS, Nevada. **Proceedings**. p. 182-191. 1994.
- MANLY, B. F. J. **Multivariate statistical methods**. Thomson Press, New Delhi. 215p. 1994.
- MASELLI, F.; CONESE, C.; PETKOV, L. Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 49, n. 2, p. 13-20. 1994.
- OTT, L.; LARSON, R. F.; MENDENHALL, W. **Statistics: A Tool for the Social Sciences**. Boston: Duxbury Press. 1983.
- PHIPPS, M. Entropy and community pattern analysis. **Journal of Theoretical Biology**. v. 93, p. 253-273. 1981.
- RICHARDS, J.A. **Remote sensing digital image analysis: na introduction**. Springer-Verlag, New York, 281p. 1986.
- 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, 2003.
- SPANNER, M. A.; BRASS, J. A.; PETERSON, D. L. Feature selection and the information content of thematic mapper simulator data for forest structural assessment. **IEEE Transactions on Geoscience and Remote Sensing**, v. GE-22, n. 6, p. 482-489. 1984.
- SWAIN, P. H., DAVIS, S. M. **Remote sensing: the quantitative approach**. New York: McGraw-Hill, 396 p. 1978.
- THOMAS, I. L.; CHING, N. P.; BENNING, V. M.; DÁGUANNO, J. A. A review of multi-channel indices of class separability. **International Journal Remote Sensing**, v. 8, n. 3, p. 331-350. 1987.
- VENKATARAMAN, S. **Exploiting remotely sensed hyperspectral data via spectral band grouping for dimensionality reduction and multiclassifiers**. Mississippi State University, 63 p. 2005. (Thesis Master of Science).
- 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, p. 36-42, 2005.