

ABSTRACT

Orbital remote sensing is considered as an important technique for urban research. Digital remote sensing data can be processed in order to enhance the urban targets, which allows the precise visualization of city limits. In this context, this work evaluates different digital image processing techniques applicable to urban studies, selecting the best by means of Kappa Statistic. The following techniques were evaluated: contrast-stretch, principal components, band ratios and vegetation index, all applied to both dry and wet seasons (Landsat Pass date of 08/06/86 and 12/09/85). For both periods the Principal Component technique was the one that best defined the limits of city of São José dos Campos, São Paulo state, Brazil.

1. INTRODUÇÃO

O sensoriamento remoto orbital é aceito como uma importante técnica de auxílio às pesquisas urbanas pois o monitoramento da expansão de cidades pode ser realizado em tempo quase real.

A constante produção de dados, obtidos por sensores localizados a bordo do satélite de recursos naturais série LANDSAT, sobre uma mesma área, possibilita captar as tendências de expansão urbana a um custo relativamente baixo quando comparado a outros tipos de levantamento, como, por exemplo, levantamentos aerofotogramétricos.

As imagens produzidas pelo sensor TM são processadas e convertidas em formas mais convenientes aos usuários, tais como imagens fotográficas, fitas compatíveis com o computador (CCT) ou transparências positivas.

O produto em CCT torna possível o realçamento de características específicas de alvos localizados em uma cena, dependendo do objetivo do pesquisador, através do tratamento automático ou processamento digital de imagens.

Algumas técnicas de processamento digital, aplicadas às áreas urbanas, refinam a imagem permitindo a visualização mais precisa dos limites urbanos, facilitando, assim, os estudos sobre a expansão urbana.

Neste contexto, o objetivo deste trabalho é analisar e avaliar as diferentes técnicas de processamento digital utilizando a estatística Kappa para selecionar a mais indicada para a delimitação urbana.

2. ESTATÍSTICA KAPPA: UMA MEDIDA DE CONCORDÂNCIA ENTRE CLASSIFICAÇÕES

Para se avaliar a exatidão de uma classificação têm sido utilizadas técnicas de análise multivariadas discreta.

A classificação dos dados é discreta porque ou os dados pertencem a uma dada categoria de uso do solo ou não. Por exemplo, um pixel pode ser classificado como pinus ou água e não metade pinus e metade água (Congalton et al., 1983).

Esta exatidão de classificação é representada, de uma maneira geral, em uma matriz de similaridade, a qual é utilizada para avaliar o desempenho de uma classificação (Congalton e Mead, 1983).

Uma matriz de similaridade é um conjunto de números colocados em linhas e colunas; as colunas, geralmente, correspondem aos dados de referência (verdade terrestre -V.T.) e as linhas aos resultados da identificação ou da classificação R.I..

Diversos pesquisadores na área de sensoriamento remoto tem investigado vários coeficientes que representam a exatidão de uma classificação temática como um todo ou a exatidão de classes individuais.

A estatística KAPPA pode ser adotada pela comunidade do sensoriamento remoto como uma medida de exatidão de classificação temática e categorias individuais pois faz uma boa avaliação do grau de concordância entre classes de uma matriz de similaridade, segundo Rosenfield e Fitzpatrick-Lins (1985).

Dallemand (1987) afirma que a estatística KAPPA inclui ao mesmo tempo a porcentagem de identificação correta e o erro de inclusão.

Segundo Fleiss (1981), a estatística KAPPA (K) é uma medida de concordância com propriedades desejáveis. Dizemos que existe uma concordância completa entre duas classificações se $K=i$. Se a concordância observada é maior ou igual a concordância devida ao acaso, $K>0$ e se a concordância observada é menor ou igual que a concordância devida ao acaso, $K<0$.

A fórmula da estatística KAPPA para uma matriz é baseada na ordenação dos dados, como na Tabela 2.1.

TABELA 2.1

MATRIZ DE SIMILARIDADE

classificador A		classificador B				
	1	2	...	K	total	
i	p ₁₁	p ₁₂	...	p _{1k}	p _{1.}	
2	p ₂₁	p ₂₂	...	p _{2k}	p _{2.}	
.						
.						
.						
K	p _{k1}	p _{k2}	...	p _{kk}	p _{k.}	
total	p _{.1}	p _{.2}	...	p _{.k}	i	

A proporção total da concordância observada é:

$$P_o = p_{11} + p_{22} + \dots + p_{kk} \quad (2.1)$$

e a proporção total da concordância esperada é:

$$P_e = p_{1.} \cdot p_{.1} + p_{2.} \cdot p_{.2} + \dots + p_{k.} \cdot p_{.k} \quad (2.2)$$

O valor de KAPPA é, então:

$$K = \frac{P_o - P_e}{1 - P_e} \quad (2.3)$$

Segundo FLEISS (1981), para testar a hipótese de que as classificações são independentes ($K=0$), a variância estimada para KAPPA é dada por:

$$VAR(K) = \frac{P_e + P_e^2 - \sum P_i \cdot P_i (P_i + P_i)}{(1 - P_e)^2 N} \quad (2.4)$$

onde n = número de amostras submetidas aos dois classificadores.

A hipótese de que a concordância entre as classificações é grande, pode ser verificada realizando-se um teste de significância, onde:

$$Z = \frac{K}{S.e.(K)} \quad (2.5)$$

dado $S.e.(K) = \sqrt{Var(K)}$

Para a Tabela de distribuição normal, rejeita-se a hipótese se Z for suficientemente grande para um nível de significância de 0.05.

3. METODOLOGIA

Foram utilizados, para a realização deste trabalho os seguintes processamentos: ampliação linear de contraste, razão de canais, índice de vegetação e principais componentes.

A ampliação linear de contraste foi realizada sobre as bandas individuais 1 e 3 do período seco e 3 do período chuvoso.

A razão de canais utilizou as bandas 5 e 7 para ambos os períodos.

Utilizando-se as bandas 3 e 4 dos períodos seco e chuvoso, realizou-se o índice de vegetação.

A técnica dos principais componentes utilizou as bandas 1, 2, 3, 4, 5 e 7 dos dois períodos.

Através do programa MATERR, desenvolvido por Medeiros, Ponzoni e Moreira (Medeiros, 1987), foi possível calcular a estatística Kappa. Este cálculo foi realizado comparando-se a imagem de referência, onde estavam definidas as duas classes segundo a verdade terrestre, com as imagens resultantes dos processamentos.

O programa MATERR faz a comparação entre duas classificações em dois canais do I-100, podendo-se admitir que uma destas classificações seja a verdade terrestre. É necessário, para isso, que o número de classes seja igual nas duas classificações.

Automaticamente, este programa fornece uma matriz de similaridade, somatório da diagonal, somatório total, precisão total, porcentagem de erros de inclusão e de omissão, análise de concordância entre as duas classificações comparadas através da estatística KAPPA para a matriz de similaridade, como um todo, e para cada classe, separadamente.

Realizando-se um teste de significância, utilizando-se a estatística z, foi possível ordenar os valores de KAPPA para cada um dos processamentos e determinar qual das imagens/produto obtidas é a melhor para cada período analisado.

É importante lembrar que cada imagem/produto foi classificada através da regra de decisão por máxima verossimelhança para poder ser comparada com a imagem da verdade terrestre.

Gerou-se apenas uma imagem referência para as duas datas pois não foram encontradas mudanças espaciais significativas entre os dois períodos.

3. RESULTADOS OBTIDOS

Para verificar a exatidão dos resultados das classificações com a imagem padrão através do programa MATERR é necessário que as duas classes consideradas (área urbana e área não urbana), tanto nas classificações quanto na imagem padrão, possuam o mesmo valor digital. Esta padronização é realizada através do programa Geração de Cores (GECOR), implementado no I-100.

Na Tabela 3.1 são apresentados os resultados obtidos para cada período.

TABELA 3.1
VALORES OBTIDOS PELA ESTATÍSTICA KAPPA

Período	Produto	R refer. x produto	S ²
Período Seco	banda 1	0,04658	0,0003582
	banda 3	0,0778	0,0006452
	razão de canais	0,1026	0,0012943
	índice de veg.	0,0938	0,000711
	princ. compon.	0,2364	0,0010636
Período Chuvoso	banda 3	0,1929	0,0006971
	razão de canais	0,3543	0,0012495
	índice de veg.	0,1109	0,0007476
	princ. compon.	0,4482	0,0017454

Levando-se em consideração o nível de significância $\alpha = 0,05$, os valores de z , para os correspondentes índices de Kappa, maiores que 1.96 tornam a hipótese H_0 rejeitada.

Para o período seco obteve-se a seguinte ordenação dos valores de Kappa segundo o nível de significância:

TABELA 3.2

VALORES DE KAPPA PARA O PERÍODO SECO

Z	K	Produto
7,2486	0,2364	Princ. Componentes
3,5182	0,0936	Índice de Veget.
3,0628	0,0778	Banda 3
2,8519	0,1026	Razão de Canais
2,4612	0,0466	Banda 1

O índice de vegetação e a banda 3 apesar de possuírem valores para Kappa inferiores ao valor obtido pela razão de canais, alcançaram maiores valores para z .

Visualmente, a razão de canais apresentou melhor resultado do que o índice de vegetação e a banda 3, obtendo um bom desempenho na classificação.

Porém, devido, principalmente, ao valor da variância da razão de canais, que influencia no cálculo de z , sua posição foi alterada.

A banda 1 obteve o menor valor para Z e para K . Isto apenas comprova a análise feita no início deste capítulo acerca desta banda. A área urbana, nesta banda não é bem definida, ou seja, devido a forte influência atmosférica esta classe, em muitos pontos, confunde-se com a outra categoria (área não-urbana).

O método principais componentes obteve o maior valor para K , reforçando uma análise feita, segundo a qual este método permitiu a melhor definição e classificação da área urbana.

Para o período chuvoso obteve-se a seguinte ordenação dos valores de Kappa segundo o nível de significância:

TABELA 3.3

VALORES DE KAPPA PARA O PERÍODO CHUVOSO

Z	Kp	PRODUTO
10,7281	0,4482	Princ.Comp.
10,0233	0,3543	"Ratio"
7,3060	0,1929	Banda 3
4,0559	0,1109	Banda 1

Neste período a ordenação dos melhores produtos coincidiu com os maiores valores de Kappa.

O maior valor de Z (10,7281) foi obtido pelo método principais componentes. Visualmente, este foi o processamento que conseguiu melhor desempenho na definição dos limites urbanos.

A razão de canais obteve o segundo maior valor para Z (10,0233), confirmando o bom desempenho visual e numérico da classificação.

O índice de vegetação obteve menores valores para Z e K do que a razão de canais e a banda 3, contradizendo os resultados de classificação MAXVER. Isto pode ser explicado pela própria essência da estatística Kappa. Por si só, a imagem pode demonstrar uma situação que, quando comparada com a verdade terrestre, torna-se irreal.

Desta forma, levando-se em consideração os resultados obtidos através da análise qualitativa e quantitativa da classificação MAXVER e os obtidos pela estatística Kappa, os valores indicaram que o método principais componentes para os dois períodos dos estudados é o que mais coincidiu com a verdade terrestre.

Estes resultados se devem, principalmente, às características deste tipo de redimensionamento de dados. Devido à concentração de informações nas três primeiras componentes e a utilização conjunta destas 3 componentes na obtenção de composições coloridas, as informações podem ser melhor manipuladas e, conseqüentemente, os resultados obtidos podem ser considerados satisfatórios.

Para estudos urbanos esta técnica, geralmente, apresenta bons resultados, facilitando não só a delimitação da área urbana como também realçando alguns alvos urbanos importantes como vegetação intra-urbana e asfalto, ressaltando o arruamento das cidades.

Como demonstraram Duggin et al. (1986), a utilização da transformação por principais componentes permite um maior número de detalhes de feições urbanas, realçando a separabilidade espectral entre os diversos alvos.

Deve ser ressaltado, ainda, que os resultados obtidos pelo período chuvoso foram melhores que os obtidos pelo período seco.

Isto se deve, principalmente, a maior nitidez dos limites urbanos no período chuvoso, o que aumenta o contraste entre a vegetação dos arredores e a área construída, como foi observado por Murai (1974) e Foresti (1978).

4. REFERÊNCIA BIBLIOGRÁFICA

BISHOP, Y.M.M.S.E.F.; HOLLAND, P.W. Discrete multivariate analysis theory and practice. Cambridge, MA, MIT Press 1975m 575 p.

CONGALTON, R.G.; MEAD, R.A. A quantitative method to test for consistency and conectess in photointerpretation. Photogrammetric Engineering and Remote Sensing, 42(1):66-74, Jan. 1983.

——— A review of the discrete multivariate analysis techniques used in assessing accuracy of remotely sensed data from error matrices. IEEE Transactions on Geoscience and Remote Sensing, GI-24(1):169-174, Jan. 1986.

- CONGALTON, R.G.; ODERWALD, R.G.; MEAD, R.A. Assessing LANDSAT classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering and Remote Sensing*, 42(12):1671-1678, Dec. 1983.
- DALLEMAND, J.F. Identificação de culturas de inverno por interpretação de dados SPOT e LANDSAT TM no noroeste do Paraná. Dissertação de Mestrado em Sensoriamento Remoto, São José dos Campos, INPE, 1987. (INPE-4334-TDL/296).
- DUGGIN, M.J.; ROWNTREE, R.; EMMONS, M.; HUBARD, N.; ODELL, A.W.; SAKHAVAT, H.; LINDSAY, J. The use on multirate multichannel radiance data in urban feature analyses. *REmote Sensing of Environment*, 20(1):95-105, Ag. 1985.
- FLEISS, J.L. Statistical methods for rates and proportions. New York, John Willey, 1981, p. 321.
- FORESTI, C. Estimativas populacionais e de crescimento de áreas urbanas no Estado de São Paulo, com utilização de imagens LANDSAT. Dissertação de Mestrado em Sensoriamento Remoto, São José dos Campos, INPE, 1978 (INPE-1298-TPT/095).
- MEDEIROS, J.S. Desenvolvimento metodológico para a detecção de alteração da cobertura vegetal através de dados MSS LANDSAT. São José dos Campos, INPE, 1987 (INPE - 4123-TDL/262).
- MURAI, S. Estimation of population density in Tokyo district from ERST-1 data. In: INTERNATIONAL SYMPOSIUM ON REMOTE SENSING OF ENVIRONMENT, 9., Proceedings, V.1, Ann Arbor, Michigan, 00-00 ap, ERIM, 1974, p. 13-22.