

Classificação de imagens aéreas de alta-resolução utilizando Redes Neurais Artificiais e dados de varredura a laser

Tatiana Sussel Gonçalves Mendes¹
Aluir Porfírio Dal Poz¹

¹ Universidade Estadual Paulista – FCT/UNESP
R. Roberto Simonsen, 305 - 19060-900 - Presidente Prudente - SP
tatisussel@gmail.com; aluir@fct.unesp.br

Abstract. The problem of urban road network extraction from digital image can be simplified by detecting RoI (Region of Interest) corresponding to streets using image classification procedure. The use of only radiometric data in image classification process can result in overlapping classes, due to objects that have similar spectral characteristics. The use of additional information (e.g. laser scanner data) can contribute to the distinction between these objects. In order to isolate regions corresponding to streets in urban environments, the goal of this work is to evaluate the result of the classification by Artificial Neural Networks, using two data sets. The first one uses only the RGB high-resolution aerial images and the second one combines RGB images with an image representing the aboveground objects, which was obtained from the laser scanner altimetry data. The analysis of the results showed that the latter data set allows better results, as it reduces the confusion between classes representing mainly streets and building roofs.

Palavras-chave: street detection, Artificial Neural Networks, normalized Digital Surface Model, detecção de vias, Redes Neurais Artificiais, Modelo Digital de Superfície normalizado.

1. Introdução

A disponibilidade de imagens de alta-resolução tem motivado e aumentado o sentido de urgência para a automação de extração de feições cartográficas (MOKHTARZADE e ZOEJ, 2007). A extração automática de malha viária urbana a partir de imagens digitais é uma tarefa extremamente complexa, uma vez que as cenas urbanas envolvem uma grande quantidade de objetos que se relacionam com as vias. O problema pode ser simplificado detectando regiões que supostamente correspondem às vias, a partir de uma prévia classificação da imagem (SHACKELFORD e DAVIS, 2003). Usando métodos de classificação não-supervisionada, a classe correspondente às vias é isolada a partir de imagens IKONOS em Gao e Wu (2004) e a partir de imagens *pan-sharpened* QuickBird em Zhang e Wang (2004). Ambos os trabalhos realizam a extração da malha viária, concentrando os processamentos nessa região de vias isolada. Mokhtarzade e Zoj (2007) detectam regiões de vias em imagens *pan-sharpened* IKONOS e QuickBird usando uma classificação por Redes Neurais Artificiais.

A utilização de apenas dados radiométricos na classificação de imagens pode resultar em sobreposição de classes, devido aos objetos que possuem comportamento espectral semelhante. O uso de informações de dados de altura, provenientes do sistema de varredura a laser, pode contribuir para a distinção entre estes objetos, como por exemplo, gramas de árvores, telhados vermelhos de solo exposto e telhados cinzas de vias. Segundo Galo (2000), a utilização de Redes Neurais Artificiais permite incorporar dados adquiridos de outras fontes, mesmo que não apresentem natureza espectral, no processo de classificação de imagens.

No contexto de isolar regiões de vias em ambientes urbanos, o objetivo deste trabalho é avaliar o resultado da classificação por Redes Neurais Artificiais, usando dois conjuntos de dados. O primeiro, que usa somente as imagens R, G e B de uma imagem aérea colorida de alta-resolução e o segundo, que além dessas imagens, é adicionada uma imagem que representa os objetos elevados, a qual foi obtida a partir de dados altimétricos de varredura a laser. Os resultados obtidos serão avaliados enfatizando a classe vias, foco principal do presente trabalho, permitindo verificar a contribuição dos dados de varredura a laser no processo de classificação para detectar regiões correspondentes às vias.

2. Classificação por Redes Neurais Artificiais

As Redes Neurais Artificiais são compostas de unidades simples de processamento, os chamados neurônios artificiais que, segundo Gonzalez e Woods (2008), foram inicialmente construídos com base no funcionamento biológico de um neurônio e são organizados de forma similar à maneira pela qual, acredita-se, que os neurônios estão interligados no cérebro.

A arquitetura de uma rede refere-se à maneira como os neurônios estão organizados e, portanto, pode afetar o desempenho da rede (BRONDINO e SILVA, 2005). Numa arquitetura de rede de múltiplas camadas (*Perceptron Multilayer*), formada por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída, cada neurônio de uma camada é conectado a todos os neurônios da próxima camada, no entanto, os neurônios de uma mesma camada não são interligados, conforme pode ser visto na Figura 1, que apresenta um modelo de uma Rede Neural Artificial *Perceptron Multilayer* com duas camadas ocultas.

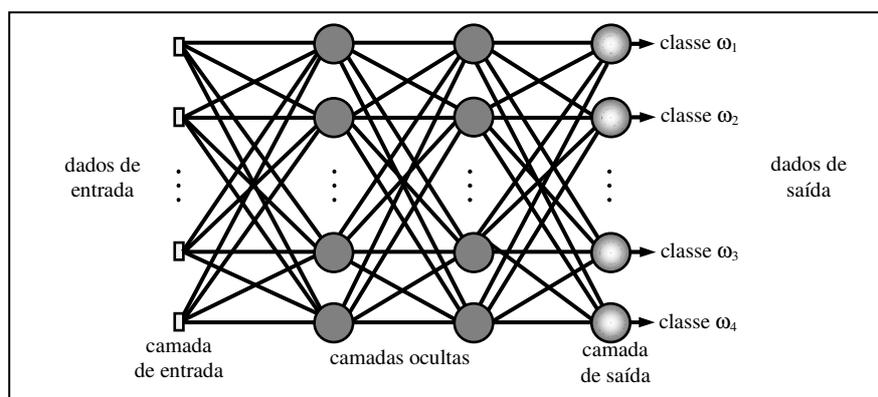


Figura 1. Modelo de uma Rede Neural Artificial *Perceptron Multilayer* (Adaptado de Mokhtarzade e Zoj (2007)).

Segundo Schowengerdt (2007), a capacidade de discriminação de uma Rede Neural Artificial está contida em seus pesos. Durante o treinamento, os pesos são iterativamente ajustados de modo a permitir que a rede discrimine os padrões de interesse. O algoritmo de treinamento mais comum é o *Backpropagation*, que é designado para minimizar a função erro entre a saída calculada por uma arquitetura e a saída desejada. A função erro é apresentada na Equação 1 (GONZALEZ e WOODS, 2008).

$$E = \frac{1}{2} \sum_{j=1}^L (d_j - o_j)^2, \quad (1)$$

sendo que E é o erro quadrático entre as saídas desejadas d_j e as saídas calculadas o_j nos neurônios em uma camada de saída. A apresentação de um padrão de treinamento na entrada da rede permite saber qual deveria ser a resposta desejada d_j de cada neurônio na camada de saída. L é o número de neurônios na camada de saída.

Em um processo iterativo, os ajustes dos pesos são calculados e adicionados aos valores anteriores. Assim, para um peso W_{ij} (peso do neurônio da camada i para o neurônio na camada j), o seu peso ajustado na iteração $(t+1)$ é dado $W_{ij}(t+1)$ conforme a Equação 2 (BOCANEGRA, 2002).

$$\begin{cases} W_{ij}(t+1) = W_{ij}(t) + \Delta W_{ij} \\ \Delta w_{ij} = \eta \delta_j o_i \end{cases}, \quad (2)$$

sendo que η é a taxa de aprendizado, cujo valor determina a proporção do ajuste dos pesos. Assim, se η for grande, o valor de ΔW_{ij} também será grande, e portanto, os pesos sofrerão uma maior alteração; δ_j é a diferença (erro) entre o valor calculado para o neurônio da camada j e a saída desejada.

3. Material e Métodos

3.1 Material

Os dados para a realização do presente trabalho correspondem a uma imagem aérea de alta-resolução e dados geométricos de varredura a laser (nuvem de pontos laser), ambos de uma mesma região. Esta região de estudo corresponde a uma área urbana da cidade de Curitiba-PR e apresenta edificações, ruas pavimentadas e vegetações de diferentes tamanhos.

A Figura 2 mostra um recorte da imagem aérea colorida. Nela, observa-se que as vias aparecem em tons de cinza, com a presença de alguns carros em seu interior. As edificações geralmente apresentam-se com telhados cinzas e vermelhos (de cerâmica).

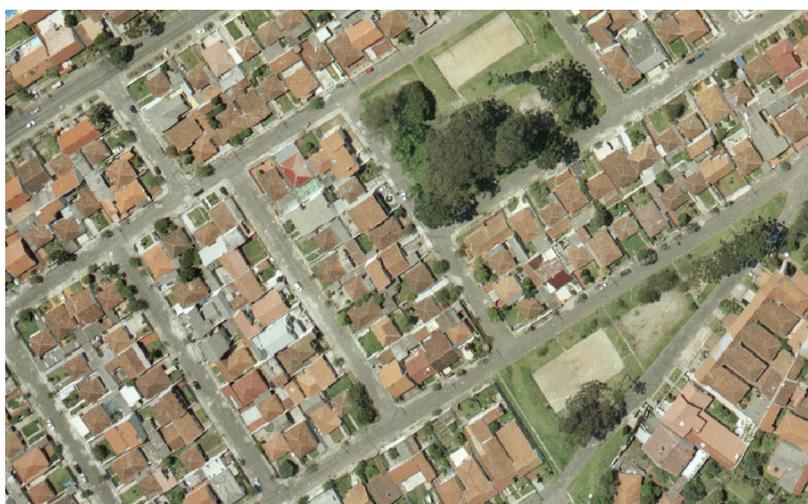


Figura 2. Recorte da imagem aérea colorida.

Para classificar a imagem colorida utilizou-se o software IDRISI Andes. Os dados de varredura a laser foram interpolados usando o método vizinho mais próximo, disponível no software SURFER e a imagem que representa os objetos elevados foi obtida usando o software HALCON da MVTec.

3.2 Metodologia

3.2.1 Pré-Processamento dos Dados

Primeiramente, para o processo da classificação de imagem, o recorte da imagem aérea colorida foi decomposto em imagens R, G e B.

Usando os dados geométricos brutos do sistema de varredura a laser foi gerado um MDSn (Modelo Digital de Superfície normalizado) da área correspondente à região apresentada na Figura 2, obtendo assim uma representação dos objetos elevados em uma superfície plana. A Figura 3 mostra a imagem que representa o MDSn, na qual observa-se que os pontos mais elevados (árvores e edificações) apresentam-se em tonalidade mais clara, enquanto os pontos mais baixos, correspondentes às ruas, praças e calçadas, apresentam-se em tonalidade bem escura.

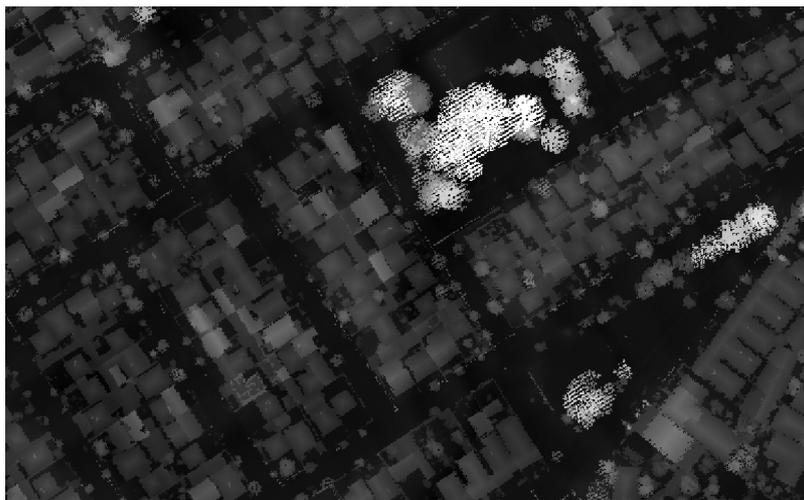


Figura 3. Imagem representando o Modelo Digital de Superfície normalizado.

A partir da análise do histograma da imagem que representa o MDSn (Figura 3), é possível escolher uma limiar que separa os objetos elevados dos restantes, gerando assim a imagem apresentada na Figura 4.

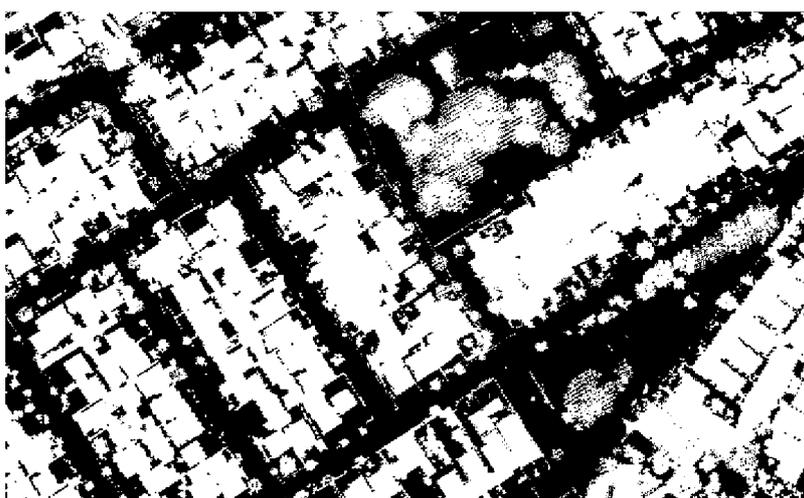


Figura 4. Imagem representando os objetos elevados.

3.2.2 Classificação por Redes Neurais Artificiais.

Anteriormente à classificação foram definidas as classes de informação presentes na imagem a ser classificada, sendo 6 classes (*vias, gramas, árvores, telhado cinza, telhado vermelho e concreto*). Para cada uma dessas classes foram coletadas amostras de treinamento.

Para a classificação por RNA, foi usado o algoritmo *Backpropagation*, implementado e disponível no software IDRISI. Para otimizar o processamento da rede foi inserido o termo *momentum* com um valor de 0,05. O valor adotado para a taxa de aprendizagem foi de 0,010.

Como dados de entrada, foram definidos dois conjuntos de dados constituindo-se de: (1) imagens R, G e B da imagem aérea colorida; (2) imagens R, G e B e da imagem que representa os objetos elevados (Figura 4). A arquitetura da rede para cada um dos dois conjuntos de dados foi testada variando o número de camadas ocultas e o número de nós dentro dessas camadas. Para selecionar a arquitetura mais adequada, avaliaram-se previamente os resultados obtidos através de uma análise visual e de índices fornecidos pelo próprio software IDRISI, o qual indica a qualidade da arquitetura da rede escolhida, baseando-se nos dados de treinamento e validação.

3.2.3 Avaliação dos Resultados

Para avaliar e comparar os resultados obtidos a partir das duas classificações fez-se necessário obter dados de referência para, então, serem usados para gerar a matriz de erro. Esses dados foram obtidos a partir de uma amostragem estratificada não alinhada, que combina as vantagens da aleatoriedade e do padrão sistemático.

Usando os dados de referência fez-se a validação cruzada das imagens geradas a partir da classificação, gerando assim a matriz de erro. A partir dessa matriz, foi calculado o índice Kappa, que segundo Congalton e Green (1999), mede a concordância entre os dados classificados e os dados de referência. Além disso, análises sobre a matriz de erro foram realizadas, uma vez que esta mostra a quantidade de pixels que foram classificados de forma correta, sendo possível comparar os resultados obtidos para a classificação usando como dados de entrada as imagens R, G e B e a classificação incluindo a imagem de altura.

A partir da análise dos resultados obtidos foi verificado se a imagem de altura proveniente dos dados laser contribui para melhorar a acurácia da classificação, principalmente em relação à classe *vias*, cuja classe é o maior interesse no presente trabalho.

4. Resultados e Discussões

A classificação por Redes Neurais Artificiais foi realizada testando diferentes arquiteturas de rede. A seleção da arquitetura considerou a análise do gráfico do Erro Médio Quadrático, a análise da diferença entre o EMQ de treinamento e de validação, a taxa de acurácia atingida e o número de iterações. Dentre as diversas arquiteturas testadas, duas arquiteturas que apresentaram melhores desempenhos foram escolhidas. Uma usando as imagens R, G e B da imagem colorida e a outra usando as imagens R, G, B e imagem de altura.

Para a classificação usando somente as imagens R, G e B da imagem colorida como dados de entrada, a arquitetura [3-18-24-6] foi a que apresentou melhor desempenho. Nessa arquitetura tem-se 3 neurônios na camada de entrada (R, G e B), duas camadas ocultas com 18 e 24 neurônios, respectivamente, na primeira e na segunda camada e 6 neurônios na camada de saída (classes de informação). Com 10.000 iterações, esta arquitetura obteve uma acurácia de 86,17%, um EMQ das amostras de treinamento igual a 0,003014 e para as mostras de validação igual a 0,003272. O valor do Kappa obtido usando apenas os dados de treinamento para a validação foi igual a 0,76. A Figura 5 mostra o resultado da classificação, juntamente com a legenda apresentando as classes de informação resultantes.



Figura 5. Classificação usando como entrada as imagens R, G e B.

A partir da análise visual do resultado da classificação (Figura 5) observa-se que a classe *vias*, em sua maioria, apresenta-se com pixels pulverizados e misturados com pixels representando edificações de telhados cinza. Isso ocorre devido à similaridade espectral que estes dois objetos apresentam. Da mesma forma, ocorre também com as áreas de concreto. No interior das quadras, é possível verificar que algumas edificações, cujos telhados correspondem à cor cinza, foram classificadas como *vias*.

Além da avaliação da arquitetura da rede, foi feita a validação cruzada, baseando-se em amostras de validação que foram interpretadas e coletadas na imagem. A Tabela 1 mostra a matriz de erro obtida na avaliação da acurácia da classificação.

Tabela 1. Matriz de erro da classificação usando como dados de entrada as imagens R, G e B.

		Dados de referência						TOTAL	Erro de Inclusão
		<i>vias</i>	<i>grama</i>	<i>árvores</i>	<i>telhado cinza</i>	<i>telhado vermelho</i>	<i>concreto</i>		
Classificação	<i>vias</i>	9692	56	7	842	86	0	10683	0,0927
	<i>grama</i>	640	2371	65	45	47	0	3168	0,2515
	<i>árvores</i>	59	316	3069	312	185	0	3941	0,2212
	<i>telhado cinza</i>	1922	14	121	1136	154	291	3638	0,6877
	<i>telhado vermelho</i>	286	18	14	111	7645	38	8112	0,0575
	<i>concreto</i>	52	0	0	57	99	531	739	0,2814
	TOTAL	12651	2775	3276	2503	8216	860	30281	
Erro de Omissão	0,2338	0,1455	0,0631	0,5461	0,0694	0,3825			

Com base nos valores da matriz de erro (Tabela 1), foi obtido um valor igual a 0,74 para o índice Kappa, cujo valor considera a qualidade da classificação como muito boa, segundo Landis e Koch (1977).

Analisado a matriz erro é possível confirmar a análise visual anteriormente realizada. O maior erro de inclusão (0,68) ocorreu na classe *telhado cinza*, uma vez que um grande número de pixels correspondentes a outras classes foram incorretamente classificados como pertencentes à essa classe. A classe que mais contribuiu para esse erro foi a classe *vias* com 1922 pixels, que deveriam ser classificados como *vias* e foram incorretamente classificados como *telhado cinza*, em seguida tem-se a classe *concreto*, contribuindo com 291 pixels. Ao analisar o erro de omissão, justamente a classe *telhado cinza* apresentou o maior erro. Isso significa que muitos pixels pertencentes a essa classe foram erroneamente classificados como pertencentes a outras classes, como por exemplo, à classe *vias*, com um total de 842 pixels, sendo a que mais contribui para este erro. Esse resultado é justificado pela similaridade espectral entre edificações de telhados cinza e as *vias*, havendo, portanto, uma maior confusão dos pixels representando estas classes.

Agora, além das imagens R, G e B da imagem colorida, foi incluída uma quarta informação, a imagem de altura proveniente dos dados de varredura a laser. Para este conjunto de dados foram testadas diferentes arquiteturas da rede, sendo que a arquitetura [4-5-5-6] foi a que apresentou um melhor desempenho. Nessa arquitetura tem-se 4 neurônios na camada de entrada (R, G, B e altura), duas camadas ocultas com 15 e 5 neurônios, respectivamente, na primeira e na segunda camada e 6 neurônios na camada de saída. Com 10.000 iterações esta arquitetura obteve uma acurácia de 92,17%, um EMQ das amostras de treinamento igual a 0,001857 e 0,002519 para as amostras de validação. O valor do Kappa obtido usando apenas os dados de treinamento foi igual a 0,86. A Figura 6 mostra o resultado da classificação.

Analisando visualmente o resultado da classificação (Figura 6) tem-se que o aspecto de pixels pulverizados na classe *vias* foi reduzido, quando comparado com a classificação realizada anteriormente (Figura 5). Nota-se ainda que quantidade de pixels correspondente a classe *vias* no interior da quadras também foi reduzida consideravelmente. Isso ocorre pela contribuição da informação de altura que permite distinguir objetos que possuem comportamentos espectrais semelhantes, mas que possuem alturas diferentes. Outra

importante diferença está em relação à classe *concreto*, que nesta classificação aparece ao redor das quadras, justamente onde se encontram as calçadas. Na classificação anterior (Figura 5), esses pixels foram geralmente classificados como *telhado cinza*.



Figura 6. Classificação usando como entrada as imagens R, G, B e imagem de altura.

A validação cruzada foi realizada usando os dados de referência e a imagem resultante da classificação, gerando assim a matriz erro, a qual é apresentada na Tabela 2. A partir da matriz erro calculou-se o índice Kappa, obtendo um valor igual a 0,86, cujo valor é maior que o valor do Kappa obtido na validação da classificação anterior.

Tabela 2. Matriz de erro da classificação usando como dados de entrada as imagens R, G, B e a imagem de altura.

Classificação	Dados de referência						TOTAL	Erro de Inclusão
	vias	grama	árvores	telhado cinza	telhado vermelho	concreto		
vias	11384	71	11	1	23	56	11546	0.0140
grama	562	2469	151	24	82	1	3289	0.2493
árvores	26	174	2950	310	184	0	3644	0.1904
telhado cinza	385	42	143	2062	347	14	2993	0.3110
telhado vermelho	14	19	21	106	7579	31	7770	0.0245
concreto	280	0	0	0	1	758	1039	0.2704
TOTAL	12651	2775	3276	2503	8216	860	30281	
Erro de Omissão	0.1001	0.1102	0.0995	0.1761	0.0775	0.1186		

Os erros de inclusão e de omissão foram calculados baseando-se na matriz erro. O maior erro de inclusão (0,3110) ocorreu na classe *telhado cinza*, no entanto, em menor valor quando comparado com o resultado anterior (Tabela 1). As classes *vias* (com 385 pixels) e *telhado vermelho* (com 347 pixels) foram as que mais contribuíram para este erro. Quando novamente faz-se a comparação com o resultado anterior, verifica-se que, na primeira classificação, 1922 pixels foram erroneamente classificados como *telhado cinza*, sendo que na verdade, são *vias*. Enquanto que na segunda classificação, o número de pixels caiu para 385. O erro de omissão mais expressivo ocorreu para a classe *telhado cinza*, com um valor igual a 0,1761, cujo valor é bem menor do que o apresentado na classificação anterior (Tabela 1). Verifica-se, ainda nesta classificação, que a classe *vias* não contribui para esse erro.

5. Conclusão

O presente trabalho apresentou uma aplicação de classificação por Redes Neurais Artificiais, com o objetivo de classificar uma cena urbana combinando uma imagem aérea

colorida de alta-resolução e uma imagem de altura, obtida a partir de dados de varredura a laser.

A partir dos resultados obtidos pode-se verificar que a classificação que inclui como dados de entrada a imagem de altura obteve um melhor resultado para o índice Kappa, confirmando a contribuição desta imagem para separar os objetos em ambientes urbanos. Além disso, ao analisar as matrizes de erro geradas, tem-se que a imagem de altura ajudou a diminuir a confusão existente entre as classes que representam objetos que possuem comportamento espectral semelhantes.

É importante ressaltar que o principal foco deste trabalho é em relação à classe *vias*, justamente onde as análises foram enfatizadas. Desta forma pode-se concluir que a informação adicional proveniente dos dados de varredura a laser contribui para classificar os pixels pertencentes à classe *vias* com maior acurácia. A localização prévia da região de vias tem como consequência diminuir a complexidade de uma subsequente extração automática da malha viária.

Agradecimentos

Ao CNPq, por concessão de bolsa de doutorado do primeiro autor (Edital MCT/CNPq nº 27/2007), ao LACTEC pelo fornecimento dos dados de varredura a LASER e à mestrandia Monique Sacardo Ferreira pelo apoio prestado durante a execução deste trabalho.

Referências Bibliográficas

Brondino, N. C. M.; Silva, A. N. R. Da teoria à prática: introduzindo técnicas emergentes na avaliação de imóveis em massa. **Revista Minerva – Pesquisa & Tecnologia**, 2 (2), p. 107–117, 2005.

Congalton, R.; Green, K. **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices**. Boca Raton: Lewis Press, 1999. Cap. 3, p. 12-14.

Galo, M. L. B. T. **Aplicação de redes neurais artificiais e sensoriamento remoto na caracterização ambiental do Parque Estadual Morro do Diabo**. 2000. Tese (Doutorado em Ciências da Engenharia Ambiental) - Escola de Engenharia de São Carlos. Universidade de São Paulo, São Carlos. 2000.

Gao, J.; Wu, L. Automatic Extraction of Road Networks in Urban Areas from Ikonos Imagery Based on Spatial Reasoning. In: **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science**, Istanbul, v.35, Comm. III/3, 2004, p. 331-337.

Gonzalez, R. C.; Woods, R. E. **Digital Image Processing**. 3th Edition. Prentice Hall. 2008. 954 p.

Landis, J.; Koch, G. G. The Measurements of observer agreement for categorical data. **Biometrics**, v.33, n.3, p. 159-179, 1977.

Mokhtarzade. M.; Zoej, M. J. V. Road detection from high-resolution satellite images using artificial neural networks. **International Journal of Applied Earth Observation and Geoinformation**, 9, p.32-40, 2007.

Shackelford, A. K.; Davis, C. H. Urban Road Network Extraction from High-Resolution Multispectral Data. In: **2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas**. p. 142-146, 2003.

Schowengerdt, R. A. **Remote Sensing: Models and Methods for Image Processing**. 3rd Edition. San Diego: Academic Press, 2007. 515 p.

Zhang, Y.; Wang, R. Multi-resolution and multi-spectral image fusion for urban object extraction. In: **XXth ISPRS CONGRESS GEO-IMAGERY BRIDGING CONTINENTS**, 2004, Istanbul, Vol. XXXV, Part B3, Commission 3, p. 960-966, 2004.