

# Segmentação e classificação de tipologias florestais a partir de imagens QUICKBIRD

## Segmentation and classification of forest types with QUICKBIRD images

Adélia Maria Oliveira Sousa<sup>1</sup>

Paulo Mesquita<sup>2</sup>

Ana Cristina Gonçalves<sup>3</sup>

José Rafael Marques da Silva<sup>4</sup>

### Resumo

O principal objectivo deste estudo foi avaliar o potencial de imagens de alta resolução espacial (Quickbird) na identificação de espécies florestais para uma região do Alentejo (Portugal), utilizando o método de segmentação multi-resolução e classificação digital orientada a objeto. Foi criada uma máscara para a vegetação com base no índice de vegetação NDVI, e obtida uma classificação com a área coberta por espécie (*Q. suber* e *P. pinea*). A detecção da vegetação foi determinada com elevado rigor, sendo a data das imagens um fator muito importante para o resultado. A classificação foi validada através de dados de campo, e foi calculada a matriz de confusão e o coeficiente Kappa. Considerou-se a identificação e separação entre as duas classes florestais dominantes (*Q. suber* e *P. pinea*) e entre as três classes, as mesmas duas classes e outra que integra todos os restantes tipos de ocupação do solo. Os resultados foram satisfatórios, com Kappa de 79.3% e de 86,5 % respectivamente. Podemos concluir que a metodologia desenvolvida e os dados de alta resolução do satélite Quickbird produzem resultados satisfatórios para a identificação e separação das espécies florestais. A baixa resolução espectral deste satélite é compensada pela elevada resolução espacial e pela utilização de outros parâmetros de forma e de textura dos objetos individuais e entre objectos.

**Palavras-chave:** detecção remota; alta resolução; espécie florestal; segmentação multi-resolução; classificação orientada a objeto.

---

1 Universidade de Évora, Departamento de Engenharia Rural, Escola de Ciências e Tecnologia, Apartado 94, 7002-554; asousa@uevora.pt

2 Instituto de Ciências Agrárias e Ambientais Mediterrânicas, Apartado 94, 7002-554 Évora, Portugal; paulomesquita00@gmail.com

3 Universidade de Évora, Departamento de Engenharia Rural, Escola de Ciências e Tecnologia, Apartado 94, 7002-554; acag@uevora.pt

4 Universidade de Évora, Departamento de Engenharia Rural, Escola de Ciências e Tecnologia, Apartado 94, 7002-554; jmsilva@uevora.pt

## Abstract

The main goal of this study was to evaluate the potential of high spatial resolution images (Quickbird) to identify forest species in a region of Alentejo (Portugal), using the multi-resolution segmentation method and object-oriented classification. A vegetation mask was created based on vegetation index NDVI, and obtained a classification of forest species with the covered area (*Q. suber* and *P. pinea*). The detection of vegetation was determined with high accuracy, being the date of the images a very important factor for this result. The classification was validated by field data, and was calculated the confusion matrix and Kappa coefficient. We considered the identification and separation between for the two dominant forest classes (*Q. suber* e *P. pinea*) and the three classes, the same two classes and one that integrates include the other types of land. The results were satisfactory with Kappa of 79.3% and 86.5% respectively. It can be concluded that the methodology and data from high resolution Quickbird satellite produce satisfactory results for the identification and separation of forest species. The low spectral resolution of this satellite is compensated by the high spatial resolution and by the use of others parameters, such as form and texture of the individual objects and among them.

**Key words:** remote sensing; high resolution; forest types; multi-resolution segmentation; object-oriented; classification.

## Introdução

O setor florestal e a sua gestão é de elevada importância à escala mundial. Pela relevância das funções económicas, ambientais, sociais e culturais a ela associadas.

A floresta em Portugal ocupa 38% do território nacional, apresentando diferentes taxas de arborização nas várias regiões do País. A distribuição segundo as principais espécies, indica o pinheiro bravo (*Pinus pinaster*) como a espécie florestal predominante, com 29,1% da ocupação, equivalente a 976 mil hectares. Segue-se o Sobreiro (*Q. suber*) com 21,3%, o que corresponde a, aproximadamente de 713 mil hectares e o Eucalipto (*Eucalyptus globulus*) com 20,1%, o que corresponde a mais 672 mil hectares de floresta (DGF, 1999, 2001). Portugal, à escala Europeia, é um país especializado no setor florestal, com

uma contribuição elevada para o Produto Interno Bruto.

A avaliação da evolução da área florestal é, em Portugal, efetuada através do Inventário Florestal Nacional (IFN), elaborado em nível nacional, com periodicidade de dez anos. Os IFN são, normalmente, constituídos por duas partes: uma primeira, de análise da fotografia aérea, com determinação da ocupação, e para as formações florestais da composição e do grau de coberto; e uma segunda, por amostragem, com medição e recolha de dados biométricos em parcelas de área conhecida. A primeira permite, ainda, a produção de cartografia de ocupação do solo e a segunda a avaliação das existências.

Existem, no entanto, ocorrências que beneficiam com uma avaliação entre inventários, nomeadamente a caracterização das áreas percorridas por fogos e os

povoamentos jovens, principalmente os de espécies de crescimento rápido. Neste caso, a utilização da detecção remota, em conjunto com os sistemas de informação geográfica, são uma ferramenta que permite a atualização da informação com uma boa relação custo-benefício.

As imagens de satélite de resolução espacial moderada como as imagens LANDSAT têm sido utilizadas para produção de cartografia florestal, no entanto, nos últimos anos, as imagens de alta resolução espacial vieram trazer outro tipo de detalhe a esses estudos à escala regional. Espera-se que estas imagens sejam um desafio às fotografias aéreas, tendo em conta as suas características geométricas, operacionais e de custos. Novas questões emergem quanto às metodologias a serem utilizadas na classificação destes novos dados.

Um dos problemas associados às imagens de alta resolução espacial prende-se ao fato de apresentarem menor resolução espectral que, em última análise, poderá dificultar o processo de classificação. Nas imagens de elevada resolução, as entidades a classificar são usualmente constituídas por vários *pixels*, dessa forma, a abordagem passa por agrupar *pixels* com características semelhantes em objetos (segmentação) que posteriormente serão o alvo da classificação (classificação orientada a objeto).

O método de segmentação e classificação orientada a objeto, através do processo de segmentação, permite utilizar como unidade de classificação não o *pixel* individual, mas sim um conjunto de *pixels*, tornando assim possível o incremento de mais variáveis para o processo de classificação. Para além da informação espectral, é possível considerar outros parâmetros bem presentes nas imagens de alta resolução

espacial, tais como a forma, textura e relação entre objetos vizinhos (CASTILLEJO-GONZÁLEZ et al., 2009, p. 209).

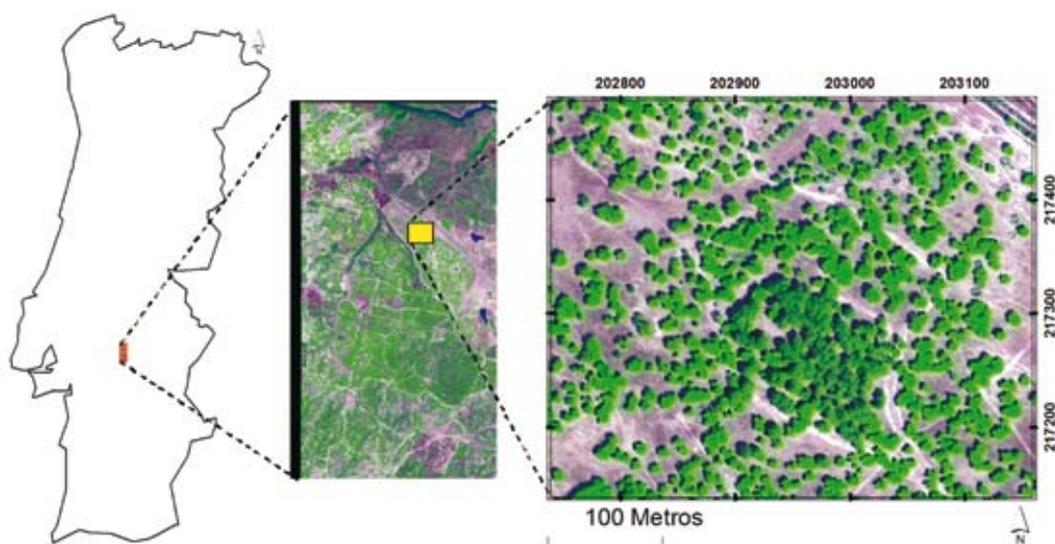
O objetivo deste estudo é: 1) desenvolver metodologia de classificação de imagens de alta resolução espacial utilizando o método de segmentação e classificação orientada a objeto, 2) analisar o potencial de imagens de alta resolução espacial na identificação de espécies florestais.

### Área de estudo e dados

A área de estudo localiza-se na região do Alto Alentejo, distrito de Évora, com uma dimensão aproximada de 2345 ha (3.5km por 6.7km, figura 1). Esta região caracteriza-se por um clima mediterrâneo, com um período seco de cerca de quatro meses. A precipitação média anual é de 550-650mm, concentrando-se sobretudo nos meses de Inverno (Novembro a Fevereiro). Quanto ao relevo, é marcado por planície, com pouca variação altimétrica. No Alentejo predomina o montado, povoamentos abertos de *Q. suber* e *Q. rotundifolia* e com menor área surge o *P. pinaster*, *P. pinea* e *E. globulus*, quer em formações puras quer em mistas.

Neste estudo, para a identificação de espécies florestais numa região no Sul de Portugal, a metodologia apresentada, bem como o trabalho de validação em campo, foram aplicados numa área de teste (Figura 1), de menores dimensões. Todas as figuras apresentadas corresponderão a essa área geográfica.

Foram utilizadas imagens de alta resolução espacial do satélite Quickbird da Digital Globe. As imagens foram adquiridas no modo "Pan-Sharpned" correspondendo à fusão da banda pancromática com as quatro bandas espectrais, b1-azul (0.45-0.52  $\mu\text{m}$ ),



**Figura I.** Localização da área de estudo. Área de teste da metodologia apresentada (Imagem em falsa cor, com a seguinte composição de bandas: Red (R) –b3, Green (G) – b4 e Blue (B) – b2)

**Figure I.** Study area. Test area of methodology developed. Bands composition (RGB – b3,b4 e b2)

b2-verde ( $0.52-0.60\mu\text{m}$ ), b3-vermelho (V) ( $0.63-0.69\mu\text{m}$ ) e b4- infravermelho próximo (IVP) ( $0.76-0.90\mu\text{m}$ ), resultando as 4 bandas com uma resolução espacial de 0.70 metros e resolução radiométrica de 16 bits.

## Metodologia

Procedeu-se primeiro à georreferenciação da imagem com base em pontos identificados no campo com um Sistema de Posicionamento Global (GPS) e em vértices geodésicos identificáveis na imagem. A seguir, procedeu-se ao processamento digital da imagem com o método de segmentação e classificação orientada a objeto, recorrendo ao programa Definiens Developer (versão 8.0.1).

As principais etapas percorridas foram: i) cálculo do NDVI, ii) criação de uma máscara de vegetação; iii) segmentação da vegetação em objectos homogêneos; iv) criação de legenda; v) classificação.

Inicialmente foi calculado o índice de vegetação Normalized Difference Vegetation Index,  $\text{NDVI} = (\text{IVP} - \text{V}) / (\text{IVP} + \text{V})$  (TUCKER, 1979, p. 128), que mostra ser um bom identificador da vegetação arbórea. Este índice é utilizado como banda adicional, com o objetivo de diminuir os problemas na discriminação da vegetação relativamente ao solo e arbustos de pequenas dimensões, resultantes da resolução espectral limitada do satélite Quickbird (KEY et al., 2001, p. 101).

Com base no NDVI, procedeu-se à obtenção de uma máscara de vegetação para a área de estudo. Realizou-se uma segmentação utilizando um limiar que maximiza o contraste entre regiões na imagem, resultando duas classes: as regiões claras representando a presença de vegetação e, em oposição, as regiões escuras representando todos os outros tipos de ocupação do solo. Quando executado ao nível do *pixel*, este algoritmo irá primeiro criar objetos quadrados de tamanho definido pelo utilizador; posteriormente,

cada um desses objetos irá ser avaliado em relação ao limiar fornecido, resultando numa máscara de vegetação, onde objetos com o valor 1 representam a vegetação e objetos com o valor 0 (zero) representam todos os outros tipos de ocupação do solo. No entanto, esta máscara de vegetação, inclui objetos mal classificados, com *pixels* que correspondem à sombra e, alguns, a solo (Figura 2a). Esta máscara foi refinada utilizando dois limiares: um aplicado à banda do infravermelho próximo e outro ao brilho, onde valores elevados de reflectância indicam a presença de vegetação e valores baixos de reflectância, representam os *pixels* de sombra e ainda de solo, resultando numa máscara de vegetação bastante precisa (Figura 2b).

Com vista à diferenciação entre espécies, foi executada uma segmentação multi-resolução para a área correspondente à máscara de vegetação com base em critérios aplicados apenas à banda do IVP.

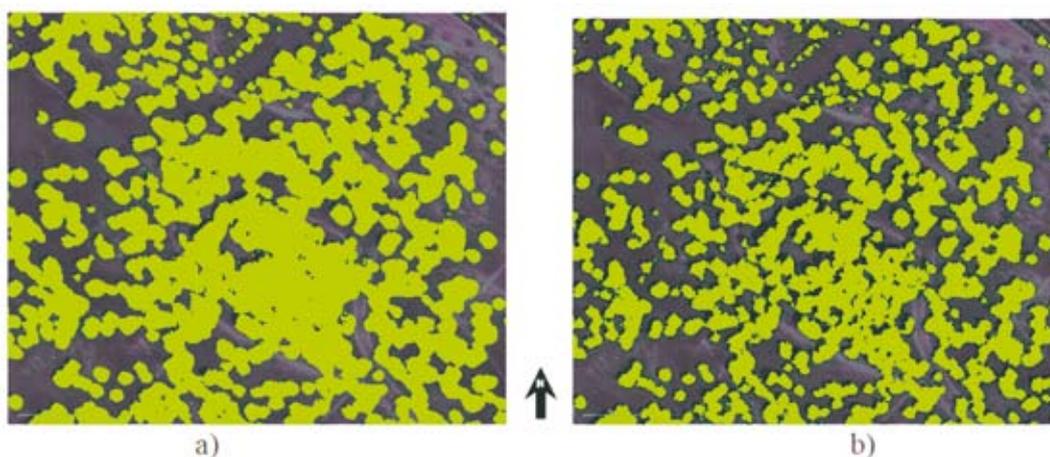
O método de segmentação multi-resolução, baseia-se num procedimento de otimização heurística, que calcula e minimiza localmente a média da heterogeneidade de regiões com características semelhantes para

toda a imagem (MATSUOKA; HAERTEL, 2007, p. 591).

Para a área de teste da metodologia foi definida a legenda com base nas espécies presentes (*Q. suber* e *P. pinea*), confirmadas com trabalho de campo. A informação recolhida serviu para a extração de áreas de treino para o processo de classificação.

No processo de classificação digital da imagem foi utilizado o algoritmo do vizinho mais próximo, com base nas características espectrais das quatro bandas e do NDVI de cada objeto (médias e desvio padrão) e ainda foram consideradas outras duas variáveis, como o brilho e a máxima diferença para as cinco bandas. Esta última variável calcula a diferença entre o valor médio mínimo e médio máximo dentro do objeto comparando os valores entre todas as variáveis disponíveis (ECOGNITION, 2010). O algoritmo do vizinho mais próximo estima a densidade local de padrões de treino na vizinhança de um padrão desconhecido durante o processo de classificação.

Após a classificação da imagem, realizou-se a avaliação da exatidão da classificação digital, com base na realização



**Figura 2.** a) Máscara de vegetação inicial; b) máscara de vegetação final, com remoção de sombras e áreas correspondentes a solo

**Figure 2.** a) First mask vegetation; b) final mask vegetation, without shadows and some soil areas

de trabalho de campo, onde se verificou qual a classe associada, a partir de uma amostragem aleatória (150 pontos). Essa informação foi comparada com os resultados da classificação da imagem através de uma matriz de confusão e do coeficiente Kappa.

A matriz de confusão (CONGALTON et al., 1983, p. 1673; STEHMAN, 1997, p.1221) apresenta o número de *pixels* classificados correctamente em oposição ao número de *pixels* previstos para cada classe no decorrer da classificação. Esta matriz permite obter informação da precisão global do respectivo classificador, bem como a precisão de classificação de cada classe, através do cálculo do coeficiente Kappa, dado pela seguinte equação:

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

onde N é o valor total de observações incluídas na matriz, r o número de linhas da matriz de confusão, x<sub>ii</sub> o número de observações na linha i e coluna i, x<sub>i+</sub> o número de observações na linha i e x<sub>+i</sub> o número total de observações na coluna i (FONSECA; FERNANDES, 2004, p. 120). Este coeficiente é uma estatística que indica o grau de concordância entre o resultado da classificação digital e a realidade no campo.

Foram calculados os erros de comissão e de omissão para se perceber a dificuldade de identificação das espécies para a classificação orientada a objeto com imagem de alta resolução espacial.

## Resultados e Discussão

O processo de segmentação multi-resolução da imagem, com base na máscara

de vegetação, originou objetos que delimitam com rigor elevado a projeção horizontal das copas de árvores individuais ou de manchas. De fato, nos povoamentos abertos, as árvores apresentam copas isoladas, no entanto, em formações mais densas é frequente haver uma imbricação das copas, sendo, por isso, difícil a sua individualização pé a pé, dando a análise por manchas resultados válidos.

Na figura 3 podem ser observados os objetos bem como a presença de duas espécies florestais com base na tonalidade da imagem em composição de falsa cor, uma espécie nitidamente mais clara, correspondendo a *Q. suber* e outra mais escura, correspondendo a *P. pinea*. Estes objetos resultantes vão ser alvo do processo de classificação digital para identificação das espécies presentes.

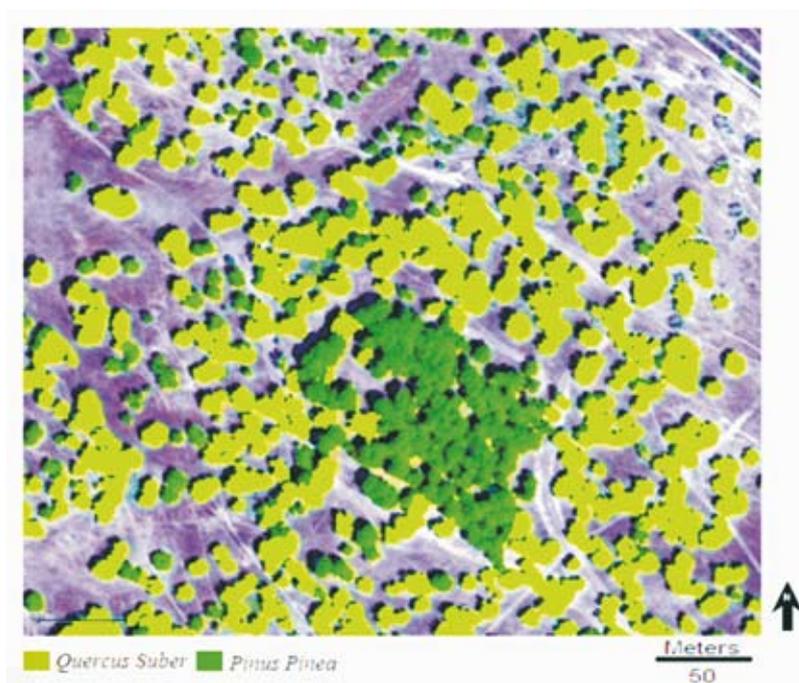
O resultado da classificação digital das duas espécies florestais, presentes na área de teste, apresenta-se com boa separação das classes (Figura 4). Podemos observar que os objectos de cor mais clara, no RGB, estão, na sua maioria, classificados como *Q. suber* e os de cor mais escura como *P. pinea*. Analisando a distribuição espacial das classes, verifica-se pequenos erros prováveis na classificação, por exemplo pequenas áreas de bordadura das copas de alguns objetos classificados como *Q. suber*. Uma explicação possível para este erro será a existência de alteração das características de parte da copa devido à sua irregularidade, que varia de individuo para individuo e há presença de sombra, visto que este problema surge ao lado da sombra. A grande vantagem da resolução espacial destas imagens, poderá assim também trazer alguns problemas, tendo em conta o elevado pormenor apresentado por este tipo de imagens.

Calculou-se a matriz de confusão e o coeficiente Kappa, para duas situações,



**Figura 3.** Segmentação multi-resolução com base na máscara de vegetação sobre uma imagem composta em falsa cor

**Figure 3.** Multi-resolution segmentation based on the vegetation mask on a false-color composite image



**Figura 4.** Imagem em falsa cor com sobreposição da classificação final para as duas espécies florestais

**Figure 4.** False color image with an overlap of the final classification for the two tree species

considerando: i) três classes florestais (*Q. suber*, *P. pinea* e outros) e ii) duas classes (*Q. suber* e *P. pinea*) onde a classe “outros” representa todos os restantes tipos de ocupação do solo (Tabela 1). No primeiro caso, verifica-se uma boa separação entre as três classes, com o coeficiente Kappa de 86.5% e uma exactidão global de 82.8%. Foram classificados incorrectamente 21% *pixels* de *P. pinea* e 10% dos *pixels* de *Q. suber*. Todas as classes mantêm pequena confusão entre si, a classe *P. pinea* tem igual confusão com as outras duas classes (*Q. suber* e “outros”) e a classe *Q. suber* tem maior confusão com a classe do *P. pinea*. Para a classe “outros” apenas 3% dos *pixels* foram mal classificados, podendo corresponder a *pixels* mistos de sombra e copa.

Para a segunda situação, onde se considerou apenas a separação entre as duas espécies florestais dominantes, ou seja, apenas foi considerada a área resultante da máscara de vegetação para a amostragem dos pontos, obteve-se um valor de Kappa satisfatório,

Kappa = 79.3%, e uma exactidão global de 91.1%. Os erros de comissão e omissão indicam uma pequena confusão na separação das duas classes, respectivamente com 16.7% e 11.8% para a classe de *P. pinea*.

## Conclusões

O método de classificação orientada a objeto com imagens Quickbird produz resultados satisfatórios na identificação das espécies *Q. suber* e *P. pinea*. Com o modo pan-sharpened das imagens Quickbird verificou-se que é possível isolar facilmente a vegetação das outras classes de ocupação do solo para regiões de clima Mediterrâneo. Um aspecto de grande importância para este facto, é o período de aquisição das imagens, já referido por (BUNTING; LUCAS, 2006, p. 246). Imagens registadas na época seca, permitem obter um maior contraste entre a vegetação arbórea e o substrato, diminuindo os erros possíveis de confusão com vegetação arbustiva com maior vigor.

**Tabela 1.** Matriz de confusão para a classificação considerando as 2 classes (indicada com fundo cinza) e as 3 classes. Erros de comissão (Co) e erros de omissão (Om)

**Table 1.** Confusion matrix for classification considering the two classes (background gray) and three classes. Commission errors (Co) and omission errors (Om)

	Pinus Pinea	Quercus Suber	Total (2 classes)	Erro de Co (2 classes)	Outros	Total (3 classes)	Erro de Co (3 classes)
Pinus Pinea	16	3	19	0.167	1	20	0.210
Quercus Suber	2	36	38	0.053	2	40	0.100
Total (2 classes)	18	39	57				
Erro de Om (2 classes)	0.118	0.077					
Outros	2	1			87	90	0.033
Total (3 classes)	20	40			90	150	
Erro de Om (3classes)	0.210	0.100			0.033		
<b>Kappa (2 classes) = 79.3%</b>							
<b>Kappa (3 classes) = 86.5%</b>							

As imagens Quickbird, apesar da sua elevada resolução espacial, poderão apresentar dificuldades na identificação e separação entre algumas espécies florestais devido à sua limitada resolução espectral, principalmente pela ausência de uma banda de infravermelho médio. No caso da região do Alentejo, poderá haver dificuldades na separação entre espécies dominantes (*Q. suber* e *Q. rotundifolia*), devido à sua semelhança espectral para os comprimentos de onda disponíveis.

A detecção, delineamento e identificação automática do coberto arbóreo com dados de detecção remota torna-se cada vez mais importante no sector florestal. Estes dados podem contribuir com a área

coberta e a distribuição espacial de espécies florestais de forma mais rápida e com menor custo, podendo torna-se uma mais valia na realização de inventário florestal, gestão de povoamentos florestais e estimativa de biomassa.

Num perspectiva futura de continuação do estudo aqui apresentado, serão objetivos: o isolamento de copa individual, minimizando assim a variação em cada objeto e entre objetos diferentes, facilitando o processo de classificação e definir uma relação entre a estimativa de biomassa por espécie e os dados obtidos por satélite, onde a relação entre a dimensão da copa e a altura da árvore, poderá ser um grande contributo.

## Referências

BUNTING, P.; LUCAS, R. The delineation of tree crowns in Australian mixed species forests using hyperspectral compact airborne spectrographic imager (CASI) data. **Remote Sensing of Environment**, v. 101, p. 230-248, 2006.

CONGALTON, R. G.; ODERWALD, R. G.; MEAD, R. A. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. **Photogrammetric Engineering and remote sensing**, Reston, v. 49, n. 12, p. 1671-1678, 1983.

CASTILLEJO-GONZÁLEZ, I. L.; LÓPEZ-GRANADOS, F.; GARCIA-FERRER, A.; PEÑA-BARRAGÁN, J. M.; JURADO-EXPÓSITO, M.; ORDEN, M. S.; GONZÁLES-AUDICANA, M. Object-and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. **Computers and Electronics in Agriculture**, v. 68, n. 2, p. 207-215, 2009.

DIRECÇÃO-GERAL DAS FLORESTAS - DGF. **Inventário Nacional Florestal**. Portugal Continental. 3ª Revisão, 1995-1998. Direcção-Geral das Florestas. Lisboa, 2001. (Relatório Final).

DIRECÇÃO-GERAL DAS FLORESTAS – DGF. **Manual de instruções para o trabalho de campo do inventário florestal nacional**, Lisboa, 1999.

ECOGNITION 8.0.1. **Reference Book**. Munich, Germany, Definiens AG, 2010.

FONSECA, A. D.; FERNANDES, J. C. **Detecção Remota**. Lisboa: Lidel, 2004.

KEY, T.; WARNER, T. A.; MCGRAW, J. B.; FAJVAN, M. A. A comparison of multispectral and multitemporal information in high spatial resolution imagery for classification of individual tree species in a temperate hardwood forest. **Remote Sensing of Environment**, v. 75, n. 1, p. 100-112, 2001.

MATSUOKA, J. V.; HAERTEL, V. Investigação do processo de segmentação utilizando o critério de ponderação de formas e cores aplicadas à imagem de áreas urbanas de alta resolução espacial do satélite Ikonos. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 13., Florianópolis. **Anais...** Florianópolis: INPE, 21-26 abril 2007, INPE, p. 589-596, 2007.

STEHMAN, S. V. Selecting and interpreting measures of thematic classification accuracy. **Remote Sensing of Environment**, v. 62, n. 1, p. 77-89, October 1997.

TUCKER, C. J. Red and photographic infra-red linear combinations for monitoring vegetation. **Remote Sensing of Environment**, v. 8, n. 2, p. 127-150, 1979.