ISPRS Journal of Photogrammetry and Remote Sensing 66 (2011) 762–775

Contents lists available at SciVerse ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Markov-random-field-based super-resolution mapping for identification of urban trees in VHR images

Juan P. Ardila *, Valentyn A. Tolpekin, Wietske Bijker, Alfred Stein

University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), Department of Earth Observation Science, P.O. Box 217, 7500AE Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 20 January 2011
Received in revised form 11 August 2011
Accepted 16 August 2011
Available online 29 September 2011

Keywords:
Image classification
Markov random field
Super resolution mapping
Urban trees
Contextual classification

ABSTRACT

Identification of tree crowns from remote sensing requires detailed spectral information and submeter spatial resolution imagery. Traditional pixel-based classification techniques do not fully exploit the spatial and spectral characteristics of remote sensing datasets. We propose a contextual and probabilistic method for detection of tree crowns in urban areas using a Markov random field based super resolution mapping (SRM) approach in very high resolution images. Our method defines an objective energy function in terms of the conditional probabilities of panchromatic and multispectral images and it locally optimizes the labeling of tree crown pixels. Energy and model parameter values are estimated from multiple implementations of SRM in tuning areas and the method is applied in QuickBird images to produce a 0.6 m tree crown map in a city of The Netherlands. The SRM output shows an identification rate of 66% and commission and omission errors in small trees and shrub areas. The method outperforms tree crown identification results obtained with maximum likelihood, support vector machines and SRM at nominal resolution (2.4 m) approaches.

1. Introduction

Municipalities and landscape designers need detailed inventories of urban vegetation to reach informed decisions about conservation, planting and maintenance of urban trees (McHale et al., 2007). Moreover, this information is needed to quantify the benefits of trees and urban forestry activities. Specific characteristics of cities such as the complexity of the urban environment, the rapid changes occurring in urban infrastructure and the limited accessibility to private areas make difficult the timely acquisition of urban forest information (Ward and Johnson, 2007; McPherson et al., 1997).

With the increasing availability of color infrared (CIR) and very high resolution (VHR) satellite imagery, vegetation in urban areas can be continuously observed and monitored cost-effectively. As such, remote sensing solutions have been sought for the quantification of urban forest resources. Myeong et al. (2006) used NDVI series of Landsat images to estimate carbon storage of urban trees; Walker and Briggs (2007) used VHR aerial images to map the urban forest in Phoenix and Walton et al. (2010) used a combination of aerial and satellite imagery to estimate the canopy cover in urban areas.

Most of those studies, however, concentrate on the overall estimation of vegetation extent at a metropolitan scale level, and only few have focused on a detailed identification of tree crowns for urban forestry inventories. Currently, determination of location and crown size of individual trees is only possible with ground surveying or with the expensive combination of optical imagery and elevation data derived from laser scanning (LiDAR) or overlapping photographs (Brandtberg, 2007). Automated extraction of locations of individual trees from VHR imagery remains a challenge even for homogeneous growing environments outside cities (Gougeon and Leckie, 2006; Hirschmugl et al., 2007; Ouma and Tateishi, 2008; Daliakopoulos et al., 2009). This is a direct consequence of the limited spatial resolution of VHR–CIR imagery and the low spectral separability of tree crowns with respect to other vegetation surfaces in the complex urban space.

In this work VHR images are those with a spatial resolution finer than 1 m in the panchromatic mode (e.g., Geoeye, WorldView-2, and QuickBird). While infrared information is fundamental for detection of vegetation in remote sensing, the current spatial resolution of VHR imagery is insufficient to discriminate tree crowns from other vegetation life-forms such as grasses and shrubs. For instance, the multispectral image mode of QuickBird (QB), and WorldView-2 has a spatial resolution about 2 m for nadir observations, whereas small and young trees have a crown diameter starting at nearly 0.5 m. Table 1 presents an estimate of the number of image pixels representing a single tree in commercially

* Corresponding author. Tel.: +31 0 53 487 4444.
E-mail addresses: ardila14710@itc.nl (J.P. Ardila), tolpekin@itc.nl (V.A. Tolpekin), bijker@itc.nl (W. Bijker), stein@itc.nl (A. Stein).

0924-2716/$ - see front matter © 2011 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2011.08.002

Table 1
Number of pixels representing a tree crown in multispectral modes of Ikonos and QuickBird images.

Crown diameter (m)	Number of pixels			
	Ikonos		QuickBird	
	MS	PAN	MS	PAN
Small (3)	<1	7	1	19
Medium (10)	5	78	14	218
Large (20)	19	314	54	872

available VHR imagery. As shown in the table, a small tree crown is captured by less than two pixels in the multispectral mode, resulting in an L-resolution-case (Woodcock and Strahler, 1987) with a large proportion of mixed pixels. Meanwhile, the panchromatic image mode, captured simultaneously and with a spatial resolution four times finer than the multispectral mode, captures approximately 16 times more pixels for a tree crown.

Alternatives have been sought to harness the spatial and spectral attributes of panchromatic and multispectral image modes. For instance, image fusion techniques such as pansharpening produce a synthetic image where the multispectral information is brought to the spatial resolution detail of the panchromatic mode. This, however, modifies the spectral information and is known to produce image artifacts (Du et al., 2007). Soft classifiers and sub-pixel methods have been used to produce information on the class-fraction composition within individual pixels, yet these methods do not inform on the actual spatial distribution of class fractions and require an accurate characterization of the pure spectral response of land cover classes (Richards and Jia, 2006).

Super resolution mapping (SRM) has been proposed as an image classification method that produces hard classification maps at a finer resolution than that of the original input image (Atkinson, 2009). It usually involves the generation of class fractions with a sub-pixel or soft classification method, followed by the generation of the fine resolution map using spatial optimization models. These models define the local spatial distribution of fine resolved pixels according to prior information on the mapped features. SRM has been implemented using approaches such as neural networks (Tatem et al., 2002), spatial optimization (Atkinson, 2005) and geostatistical techniques (Boucher et al., 2008). Another SRM approach (Kasetkasem et al., 2005), tested on synthetic data, spatially optimized the class fractions using a Markov random field (MRF) after the definition of an objective energy function. Few SRM studies, however, have considered the potential contribution of panchromatic information for classification purposes. Moreover, SRM results are limited by the quality of class fractions estimated from external sources or from sub-pixel classification techniques such as linear unmixing (Tatem et al., 2002; Atkinson, 2009). The applicability of linear unmixing for tree crown detection is questionable because: (a) tree crowns are characterized by large spectral variance in VHR images, (b) VHR images have a limited number of spectral bands, and (c) tree crowns have a similar spectral response with other understory surfaces e.g., shrubs and grasses.

MRF-based SRM is an attractive image analysis approach for the identification of tree crowns in urban areas as it addresses two main issues constraining this task in VHR satellite images. First, since SRM produces classification maps with a finer spatial detail than that of the input image, trees not evident at the native image resolution could be identified, while larger trees and group of trees should be better delineated and resolved in the classification output. Second, incorporation of contextual information with an MRF approach results promising to overcome mixed pixel effects and the large within-class spectral variance inherent to vegetation in VHR images.

This work focuses on tree crown detection in VHR–CIR images under the scope of a practical implementation of detailed tree inventories in urban areas. We aim to develop MRF-based SRM for identification of tree crowns in VHR satellite images. With this approach, we exploit the multispectral and panchromatic information of VHR imagery and optimize the spatial correlation between pixels of a fine-classified map that does not rely on linear unmixing or a pansharpening method. This paper extends the work of Kasetkasem et al. (2005) and developments introduced by Tolpekin and Stein (2009). Furthermore, we apply the MRF-based SRM model to a QuickBird image of a residential area in The Netherlands and present a detailed study on energy minimization and a thorough assessment of pixel-based and object-based performance of the SRM method as compared with alternative pixel-based detection methods.

In the rest of this document Section 2 presents the conceptual and mathematical approach for MRF-based SRM, while Section 3 presents the parameter optimization and implementation of the method in QuickBird images. Sections 4 examines the performance and accuracy of the obtained results against alternative classifiers. The paper ends with a discussion of the proposed method, a short concluding section, and an appendix describing the implemented energy minimization algorithm.

2. Super resolution mapping

We consider the classification of a multispectral image y that consists of K spectral bands with spatial resolution R and pixel locations $b_i \in B$, where B is a $M_1 \times M_2$ pixel matrix. In addition we assume a panchromatic image z with finer spatial resolution $r < R$. The super-resolution map (SR map) c is defined on the set of pixel locations A and covers the same extent on the ground as y and z with spatial resolution r. The scale factor $S = R/r$ is an integer for common VHR images. Hence each pixel b_i corresponds to the area on the ground covered by S^2 finer resolution pixels $a_{j|i},\ j = 1, \ldots, S^2$.

We assume the existence of a multispectral image x defined on the set of pixels A with K spectral bands and fine spatial resolution r. Image x is not observed directly while images y and z are considered as spatial and spectral degraded observations of x respectively. Furthermore, we assume that each pixel in x can be assigned to a unique class: $c(a_{j|i}) = \alpha$, with $\alpha \in 1, 2, \ldots, L$. The relationship between y and x, and z and x are established by the degradation models:

$$y_k(b_i) = \frac{1}{S^2}\sum_{j=1}^{S^2} x_k(a_{j|i}),\ \ k = 1, \ldots, K \tag{1}$$

$$z(a_{j|i}) = \frac{1}{K}\sum_{k}^{K} x_k(a_{j|i}) \tag{2}$$

for $b_i \in B$, $a_{j|i} \in A$, and $y_k\ (b_i)$ being the feature vector for b_i in K-dimensional feature space. Note that Eqs. (1) and (2) cannot be solved for x because for $K > 1$ and $S > 1$ the number of equations $(M_1M_2(K + S^2))$ is smaller than the number of unknown values $(M_1M_2KS^2)$. We therefore do not intend to estimate image x; instead we aim to find the SR map c that corresponds to the maximum a posteriori probability (MAP) solution $P(c|y,z)$ for c given observed data y and z. Note that this setup does not constrain the SR map c to the estimated class fractions of a soft-classification method, but it rather optimizes the c map regarding the spatial distribution of class labeled pixels and the spectral properties of y and z images. Thus, according to Bayes' theorem, $P(c|y,z)$ is computed from prior probability $P(c)$ and conditional probabilities $P(y|c)$ and $P(z|c)$ as

$$P(c|y,z) \propto P(c)P(y|c)P(z|c) \tag{3}$$