ISPRS Journal of Photogrammetry and Remote Sensing 108 (2015) 128–137

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Road networks as collections of minimum cost paths

Jan Dirk Wegner *, Javier Alexander Montoya-Zegarra, Konrad Schindler

Photogrammetry and Remote Sensing, ETH Zürich, Switzerland

ARTICLE INFO

Article history:
Received 25 June 2014
Received in revised form 1 July 2015
Accepted 9 July 2015

Keywords:
Aerial
Multispectral
Urban
Networks
Extraction
High resolution

ABSTRACT

We present a probabilistic representation of network structures in images. Our target application is the extraction of urban roads from aerial images. Roads appear as thin, elongated, partially curved structures forming a loopy graph, and this complex layout requires a prior that goes beyond standard smoothness and co-occurrence assumptions. In the proposed model the network is represented as a union of 1D paths connecting distant (super-)pixels. A large set of putative candidate paths is constructed in such a way that they include the true network as much as possible, by searching for minimum cost paths in the foreground (*road*) likelihood. Selecting the optimal subset of candidate paths is posed as MAP inference in a higher-order conditional random field. Each path forms a higher-order clique with a type of clique potential, which attracts the member nodes of cliques with high cumulative road evidence to the foreground label. That formulation induces a robust P^N-Potts model, for which a global MAP solution can be found efficiently with graph cuts. Experiments with two road data sets show that the proposed model significantly improves per-pixel accuracies as well as the overall topological network quality with respect to several baselines.

1. Introduction

Despite more than three decades of research, automatic road extraction from remote sensing data remains to a large degree unsolved. Since the initial attempts in the mid-seventies (Bajcsy and Tavakoli, 1976) important progress has been made – see the overview papers (Heipke et al., 1997; Mayer et al., 2006) – but to our knowledge no fully automated road extraction system so far performs at a level that would allow operational use. In practice the extraction or update of roads is at most semi-automatic and requires a significant degree of user interaction, e.g. (Gerke et al., 2004; Zhang, 2004; Helmholz et al., 2012).

Factors that make the task challenging are strong illumination effects, appearance variations due to clutter and shadows, and occlusion by nearby buildings and vegetation. In urban environments these factors are compounded with highly variable road width, density and curvature, which makes the extraction particularly difficult. Even in "planned" cities with a regular grid layout (e.g., many American towns) nearby trees and buildings frequently cast shadows on roads or occlude them altogether. For older or more informally growing cities with irregular, narrow, winding roads the problem becomes much worse.

Road extraction in the presence of noisy and ambiguous low-level image evidence requires strong a priori knowledge. It turns out that formalizing the structural properties of roads (and also other networks, e.g., in medical image processing) in a prior is difficult. Existing models are usually either too restricted to faithfully describe the network, or too complex for stable and efficient inference (see overview in Section 2).

We seek a compromise between these extremes and develop a model of the road network which is on the one hand expressive (e.g., it does not impose a tree structure or require piecewise straight roads), and on the other hand amenable to powerful inference algorithms (i.e., it does not require expensive all-purpose solvers like MCMC or Gibbs sampling). The proposed method follows the recover-and-select strategy: an over-complete collection of potential road segments is generated, which is subsequently pruned to those segments which cover parts of the road network. The segments, which we call *paths*, are found by minimum cost path computation based on local features. The pruning step, in which incorrect paths are suppressed, is formulated as MAP inference in a higher-order conditional random field (CRF), constructed in such a way that it allows for efficient global energy minimization. The conservative *recover* step ensures high completeness (recall), while the *select* step aims to maximize correctness (precision), by explicitly including long-range connections via higher-order CRF potentials.

* Corresponding author.

http://dx.doi.org/10.1016/j.isprsjprs.2015.07.002
0924-2716/

To our knowledge, this is the first attempt to combine the classical idea of minimum cost paths with the comprehensive global modeling capabilities of CRFs, for road network extraction in particular and for other loopy, undirected networks in general. A preliminary version of this work appeared in Wegner et al. (2013). In that work local stretches of road were confined to lie on straight line segments. Here, we extend the model to allow for arbitrary (minimum-cost) paths that naturally adapt to more complex road shapes (e.g., sharp bends). Moreover, we present a much expanded experimental evaluation.

2. Related work

Road extraction in rural or suburban areas is often approached in a rule-based fashion, i.e. one attempts to explicitly formulate an exhaustive set of rules for delineating the road network (Doucette et al., 2004; Mena and Malpica, 2005; Poullis and You, 2010; Grote et al., 2012; Ünsalan and Sirmacek, 2012; Miao et al., 2013). Common to all these approaches is a heuristic processing pipeline consisting of multiple sequential or intertwined steps, with a rather large set of parameters that need to be re-tuned for each new scene.

To bridge the gap between low-level road cues and high-level road network layout in a more principled way Stoica et al. (2004) (later followed by Lacoste et al. (2005) and Lafarge et al. (2010)) have introduced marked point processes, a comprehensive probabilistic framework to impose connectivity priors. In Chai et al. (2013)) the authors extend the original idea of sampling line-segments by explicitly modeling junctions with point processes. However, the corresponding objective functions can only be minimized with all-purpose solvers like simulated annealing and/or reversible jump Markov Chain Monte Carlo (RJMCMC). They are thus on one hand computationally very expensive and on the other hand risk not finding a satisfactory optimum (e.g. due to poor mixing of the chain).

All previously mentioned approaches primarily focus on rural and suburban scenes with relatively sparse and mostly unoccluded road networks. Only few works deal with road extraction in more complex urban areas. Hinz and Baumgartner (2003) have developed a detailed heuristic model for roads and their context in scale-space, using evidence from multiple overlapping aerial images. More recently, Youn et al. (2008) combine an orthophoto and airborne laser scanning data to extract wide, largely unoccluded roads that follow a grid pattern. Similar to Hinz and Baumgartner (2003) and Grote et al. (2012) they design a hierarchical framework which constructs longer road pieces from initial segments, but no high-level connectivity is imposed, thus many gaps remain.

Here, we argue that a particularly important property of the road network is its connectedness. Hierarchical bottom-up procedures that iteratively assemble short pieces of road to longer ones must base their decisions primarily on local shape constraints, whereas they account for connectedness at a very late stage (or not at all). It seems more intuitive to view road networks as a collection of smooth, connected long-range paths without strong restrictions on the local shape. Probably the first work to model roads via minimum cost paths is (Fischler et al., 1981). They use an A^*-type algorithm to iteratively find roads based on per-pixel scores generated with a line detector. Since this early attempt various groups have proposed *semi-automated* road tracking approaches, in which single roads (mostly in rural areas) are traced after manual selection of starting nodes. Technical implementations of this idea include Kalman filtering (Vosselman and de Knecht, 1995), extended Kalman filtering and/or particle filtering (Movaghati et al., 2010), heuristic rule-based tracing (Baumgartner et al., 2002), and shortest path computation by dynamic programming (Gruen and Li, 1995; Gruen and Li, 1997; Dal Poz et al., 2010; Dal Poz et al., 2012).

To our knowledge, minimum cost paths for *automated* road network extraction have not been followed up in recent years in remote sensing, until recently Türetken et al. (2012) tested their method, originally developed for vessel tree extraction in medical imagery, on road networks. In *medical imaging*, many researchers have used minimum cost paths to model 2D and 3D tree structures, e.g. (Li and Yezzi, 2007; Türetken et al., 2011; Benmansour and Cohen, 2011; Bas and Erdogmus, 2011; Zhao et al., 2011). Generally, these approaches first detect local cues, which are then connected to elongated tubes via minimum cost paths. Model-based criteria ensure that all branches fit into a global tree topology, either in a bottom-up or in a top-down fashion. Bottom-up methods try to initially extract only correct network pieces, thereby accepting low completeness, followed by insertion of missing links, e.g. (Bas and Erdogmus, 2011; Wang et al., 2011). Top-down methods proceed the other way round, and first generate an overly complete network, by allowing all potential paths at the risk of a high false alarm rate. Subsequently, erroneous links are pruned to obtain a correct network, e.g. (Li and Yezzi, 2007; Türetken et al., 2012). Bottom-up methods are usually fast to compute iteratively but often fail to bridge large gaps, whereas top-down techniques have problems when it comes to suppressing "shortcuts" through the background. The approach proposed here follows the top-down strategy and aims for high *topological completeness*, i.e. our objective is to extract the complete urban road network *as far as possible*, which is crucial for navigation applications such as personal navigation systems or vehicle routing.

A method similar in spirit to ours is (Türetken et al., 2012), which was extended to graphs with cycles in Türetken et al. (2013). They locally compute road (resp. tube) likelihoods at each pixel (or voxel, if applied to stacks of medical images) and connect seed points via minimum cost paths at several scales. The resulting graph is broken down into short overlapping segments, and a network graph through the set of segments is found with mixed integer programming. Although originally developed for a medical application, the experiments also demonstrate promising performance for suburban road networks in aerial images.

For completeness we also mention a body of literature, starting with (Laptev et al., 2000), that uses the term "road extraction" for the delineation of roads with different variants and extensions of active contour models ("snakes"). For example, Butenuth and Heipke (2012) extend standard snakes to explicitly model the network topology, including junctions, with so-called network snakes. Wang et al. (2011) apply a similar snake formulation to iteratively reconstruct tree-like tubular structures from medical image stacks. However, snakes are a local optimizer, and mainly useful to delineate roads more precisely once their approximate layout is known. We therefore rather see them as a potential geometric refinement *after* extraction.

3. Network extraction

In our approach the road network is thought of as the union of many elongated *paths*. In this way, network extraction can be cast as the search for a set of paths that together cover the entire network. The proposed method follows the recover-and-select strategy:

- In the *recover* step a large, over-complete set of potential *candidate paths* is generated, by finding the most road-like connections between many different pairs of seed points. The aim of candidate generation is high recall, ideally the candidate set covers the entire road network, at the cost of also containing many false positives that do not lie (completely) on roads.