Computers, Environment and Urban Systems 41 (2013) 289–298

Contents lists available at SciVerse ScienceDirect

Computers, Environment and Urban Systems

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

Application of random sets to model uncertainty of road polygons extracted from airborne laser points

Liang Zhou *, Alfred Stein

University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC), Hengelosestraat 99, P.O. Box 217, 7500 AE Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 13 January 2012
Received in revised form 29 June 2012
Accepted 30 June 2012
Available online 25 July 2012

Keywords:
Uncertainty modelling
Road extraction
Random sets
Laser scanning
Dependency

ABSTRACT

High point densities obtained by today's laser scanning systems enable the extraction of features that are traditionally mapped by photogrammetry or land surveying. While significant progress has been made in the extraction of roads from dense point clouds, little research has been performed on modelling uncertainty in extracted road polygons. In this paper random sets are used to model this uncertainty. Based on the accuracy reported by the data provider, positional errors in laser points are simulated first by a Markov Chain Monte Carlo method. An algorithm is developed next to detect the positions of road polygons in the simulated data and integrating the random sets for the uncertainty modelling. This algorithm is adapted to point data with different densities and variable distributions. Uncertainty modelling includes modelling of the dependence between the vertices of a road polygon. Road polygons constructed from vertices with different truncated normal distributions along with their uncertain line segments are represented by random sets, and their parameters are estimated. The effect of distributions on the area of the mean set is analysed and validated by a set of reference data collected from GPS measurements and image digitising. Results show that random sets provide useful spatial information on uncertainties using their basic parameters like the core, mean and support set. The study shows that random sets are well-suited to model the uncertainty of road polygons extracted from point data.

1. Introduction

LiDAR is a terrain and urban information acquisition technique based on laser technology. The use of LiDAR data for urban modelling and visualisation has received much attention recently. Specifically, digital city modelling is benefiting from realistic visualisations. Advantages such as short data acquisition and processing times, relatively high accuracy and point density, and low costs have caused LiDAR to be preferred over traditional aerial photogrammetric techniques.

Laser scanners nowadays can acquire point clouds with densities of 20–50 pts m^{-2} from airborne platforms. These point densities enable the use of laser scanning data for various mapping tasks. Studies on the use of laser points typically focused on the applications such as DTM generation (Kraus & Pfeifer, 2001), 3D building modelling (Brenner, 2005; Oude Elberink & Vosselman, 2009; Pu & Vosselman, 2009), and change detection (Matikainen, Hyyppä, & Hyyppä, 2003). In the context of road furniture and forest inventories, algorithms for the detection of pole-like objects have been developed (Brenner, 2009; Pfeifer, Gorte, & Winterhalder, 2004; Rutzinger, Pratihast, Oude Elberink, & Vosselman, 2010). Recent research shows that objects such as traffic islands and pavements can be extracted from airborne point clouds (Zhou & Vosselman, 2012). Positional accuracy of the extracted objects is improved by fitting a sigmoid-shaped surface. Uncertainty, however, still exists in the modelled road polygons due to the limitation of designed model and various point densities and distributions in different study areas. As the shapes of road polygons are diverse in urban areas, the designed mathematical models have their limitations in precisely modelling these diverse shapes. Hence, properly modelling this uncertainty in the extracted road polygons will be interesting for data users and stimulate other applications in urban environments.

As a basic GIS operation, the area of a polygon is calculated from the coordinates of the vertices representing its boundary. Therefore uncertainty in the coordinate values leads to uncertainty in area calculation (Van Oort, Stein, Bregt, De Bruin, & Kuipers, 2005). Traditional probability theory has been used for modelling error propagation in spatial objects. Positional uncertainties are mainly caused by measurement errors (Zhang & Goodchild, 2002). Analytical approaches have been developed for modelling uncertain points (Thapa & Bossler, 1992) and lines (Shi & Liu, 2000). Models for uncertain polygons are mostly based on models for points and lines, as the uncertain location of the outline of a polygon is specified by the joint probability distribution function

* Corresponding author. Tel.: +31 53 4874 345; fax: +31 53 4874 335.
E-mail addresses: zhou19153@itc.nl (L. Zhou), stein@itc.nl (A. Stein).

0198-9715/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compenvurbsys.2012.06.006

of its primitive points (Zhao, Stein, & Chen, 2010). Research has been conducted to model the uncertainty of individual polygons (Bondesson, Stahl, & Holm, 1998; Chrisman & Yandell, 1988; Griffith, 1989; Liu & Tong, 2005; Magnussen, 1996; Næset, 1999; Shahin, 1997). Prisley, Gregoire and Smith (1989) derived variance and covariance equations from an area equation that required the coordinate values of each polygon's centroid. Based on the work of Chrisman and Yandell (1988), Van Oort et al., 2005 derived variance and covariance equations, which are independent of polygon centroid coordinates. In most of the above studies, assumptions were made that the uncertainty in the area of a polygon originated from the uncertainty of the vertices defining its outline. Liu and Tong (2005) developed an error model to describe the influence of line uncertainty on the uncertainty of polygon area where the effect of uncertainty in parametric curves on the uncertainty of polygon area, however, was not explicitly modelled. The main reason is that an error model for parametric curves is complicated as the error may be either in the endpoints or in the parameters (Chrisman & Yandell, 1988). Since spatial objects derived from remote sensing data may have gradual transition boundaries (Stein, Hamm, & Ye, 2009), estimating the probability density function (pdf) for each polygon vertex may be difficult. In addition, georeferencing of remote sensing data and manual digitization may introduce correlations between boundary points (Heuvelink, Brown, & Van Loon, 2007).

Alternatively, uncertain spatial objects can be modelled by random sets. Random sets were originally developed for the study of randomly varying geometrical shapes (Stoyan & Stoyan, 1994) and for image segmentation (Epifanio & Soille, 2007). Zhao et al. (2010) used random sets to model the uncertainty of wetland derived from satellite images. It quantifies extensional uncertainty of spatial objects and models the broad boundaries extracted from images (Zhao, Stein, Chen, & Zhang, 2011). Its parameterization is adapted for monitoring seasonal dynamics of wetland variation and interannual changes of wetland inundation extents (Zhao, Stein, & Chen, 2011). In this way, a natural entity with uncertainties is treated as random sets in population space. The objects derived by digitizing, thresholding or segmentation of image data were modelled as a set of observations in sample space. A statistical analysis helps to understand the characteristics of the sample.

The aim of this paper is to examine the feasibility of employing random sets to model the uncertainty of road polygons derived from airborne laser scanning data. This is realised in three steps. First, the uncertainties in airborne laser points are simulated and their effect on the uncertainties of derived polygons is explored. Second, random sets theory is used to model the uncertainties of derived polygons. Third, a statistical analysis is employed to estimate the characteristics of the derived polygons. The study is applied on a set of six road polygons from the city of Enschede, The Netherlands.

2. Data description

Airborne laser scanning data used in this study were acquired with a FLI-MAP 400 system (Fugro Aerial Mapping., 2011) with forward, nadir, and backward looking scan directions. The system consists of an airborne laser scanner, two digital cameras and two video cameras. The dataset contains 15 strips with a point density of 20 pts m^{-2} and was recorded at a flight height of 275 m above ground in Enschede, The Netherlands. The systematic errors (offsets between strips) are in the order of 4–8 cm for the *X*, *Y* coordinates and 2–3 cm for the *Z* coordinates. The stochastic platform positioning error is approximately 2–3 cm for *X*, *Y* and *Z* coordinates. Planimetric standard deviations of 2 cm have been achieved with a little additional calibration following Vosselman (2008). This is superior to the accuracy reported in the platform specification of 5 cm accuracy in both horizontal and vertical directions at the 95% confidence interval. Fig. 1 highlights the six road polygons digitised from the orthophoto, being the subject of the study. More details are provided in Sections 3 and 4.

3. Methodology

The quality of objects derived from remote sensing data depends upon properties of the input data and the processing steps. Uncertainty modelling of the output objects should include errors in the input data as well as errors caused by data processing methods. In this study the positional errors in laser points were assumed to have a bivariate normal distribution and were simulated by using Markov Chain Monte Carlo (MCMC) simulations (Besag, 2001). A previously developed method (Zhou & Vosselman, 2012) was used to detect and model the road polygons from the error-contaminated data. Every extracted road polygon consists of a sequence of vertices and the uncertainty of each vertex is represented by its uncertain line segment. Random sets were used to model the uncertainties in the extracted road polygons and results of statistical analysis were discussed. The impact of viable distances between two neighbouring vertices and their dependency on the uncertainty of derived road polygons is analysed by means of different experiments.

3.1. Simulating the positional errors in laser points

Markov Chain Monte Carlo methods consist of a class of algorithms for sampling from probability distributions. To conduct an MCMC simulation, the probability distribution of random variables should be known. Hunter et al. (1996) found that the standard deviation supplied by the data producer is useful to simulate random errors. In this research, it is assumed that the horizontal error of each LiDAR point has a bivariate normal distribution with a mean of zero and a standard deviation (SD) known from data provider. Since the vertical error of LiDAR points has less influence on the positional uncertainty of a polygon, we do not take it into account. The Metropolis–Hastings algorithm is applied on the MCMC method for obtaining a sequence of random samples from a probability distribution. As the laser points have been calibrated to a high degree of accuracy, a planimetric standard deviation of 2 cm was achieved. For each laser point, a sequence of 1000 random samples was selected by the Metropolis–Hastings algorithm from a bivariate normal distribution $X \sim \mathcal{N}(\mu, \Sigma)$, where $X = [x, y]$, $u = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma = \begin{Bmatrix} 4 & 0 \\ 0 & 4 \end{Bmatrix}$ (Fig. 2).

3.2. Mapping road polygons from simulated data

In urban areas, curbstones often separate the road surface from the adjacent pavement. These curbstones are mapped using a three step procedure. First, the locations with small height jumps near the terrain surface are detected. Second, midpoints of high and low points on either side of the height jump are generated, and are put into a sequence to obtain a polygon describing the approximate curbstone location (Zhou & Vosselman, 2012). A sigmoidal function (Eq. (1)) is then fitted to the simulated points near the polygon to increase the accuracy. We use this function to describe the height *Z* as a function of the location *x* perpendicular to the road side. In this equation *W* is the slope parameter, x_{ip} is the inflection point, and Z_t and Z_b are the top and bottom height (Fig. 3a). Further, the longitudinal shape of the curbstone is assumed to follow a cubic polynomial $x_{ip} = a_0 + a_1 y_{ip} + a_2 y_{ip}^2 + a_3 y_{ip}^3$ where the *y*-direction is taken parallel to the local road side