Recognition of Road Network in Aerial Images using CRF and Marked Point Process

XUANZI YONG[1] & UWE SÖRGEL[2]

Abstract: In this paper, we propose a method to combine a Marked Point Process with Conditional Random Fields for the purpose of road network extraction, in which stochastic process and probabilistic graphical model classification are coupled together, aiming at extracting road network faster and more efficiently. In our method, superpixel segmentation is conducted first and the resulting segments are regarded as the basic units or primitives for classification. The features are mainly derived from the attributes of an ellipse circumscribing the superpixel, which are fed into a conditional random field operating on the superpixel graph. On the next step, each superpixel segment is modeled as a directed point in space. Instead of usual uniform sampling as prior, we take advantage of the labeling results of the previous CRF classification, which provides for each segment a belief of being as a part of road. The spatial interactions between point segments are defined by means of the direction of the superpixel ellipse. Compared with the traditional marked point process, much less input points are considered, the sampling procedure is reduced leading to faster convergence of the optimizing process.

1 Introduction

The knowledge of road network offers plenty useful information for human daily life such as routing and navigation, as well as applications in commercial and industry area, for example, hot topics like unmanned aerial vehicles and self-driving cars, in use for purposes such as surveillance and control. To be useful, this kind of information need to be updated correctly on time. In order to minimize or even avoid such manual operations, the problem of automatic road network recognition has been widely investigated for many years.

However, this topic is still a challenging. There are many difficult factors preventing to get a good solution. The appearances of roads are various. We might deal with a countryside alley, which is narrow and curved, or a crowded road on metropolis, which is full of moving mobiles and parked with parking cars on the side. Moreover, the complex environment around road area may cause mistakes, for example, buildings with concrete roofs or the traditional grey-green roof tiles. These road-like background objects should be eliminated. On the other hand, a core purpose of road networks is to provide good connectivity, which is one of the kern point especially for routing and navigation. But in practice, the shadows of buildings and even the crowns of trees may divide the road into parts.

In this work, we propose a new point process framework for road network extraction. It contains three steps as image segmentation, CRF pre-classification and point processing. The segments are

[1] Technische Universität Darmstadt, Fachgebiet Fernerkundung und Bildanalyse, Franziska-Braun-Str.7, D-64287 Darmstadt, E-Mail: yong@geod.tu-darmstadt.de

[2] Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Str. 24D, D-70174 Stuttgart, E-Mail: soergel@ifp.uni-stuttgart.de

modelled as shaped points which inherit the feature of the underlying segment. Also, a CRFs as well as MPP model is constructed for road recognition. While common MPP model the objects as random numbers at some random location, the sampling process is quite time-consuming. Our proposed method uses a simple CRF to derive an initial segmentation and classification, leading to a significant lower number of objects compared to the given pixel grid. Then we turn to MPP sampling, which ensures that relevant points are sampled more frequently, leading to significant speed up the simulation process and convergence to final result.

2 Related works

The problem of automated road recognition has been widely investigated for many years. The proposed resolutions range from image segmentation (SONG & CIVCO 2004) to line detection (JIA et al. 2005), from mathematical morphology (VALERO et al. 2010) to advanced statistic models (WEGNER et al. 2013).

Conditional random fields (CRF) are widely used for image classification and labeling (ZHANG & JIA 2012; ZHONG et al. 2014). HUANG et al. (2014) proposed an object-based CRFs model for road extraction, in which each object is represented by the color, the gradient of histogram and the texture. Object-based methods are robust to noise and can make full use of spatial information. A road network can also be represented in a graph as collection of line segments. WEGNER et al. (2015) introduced a higher-order CRF model, where the clique consists of all pixels along minimum cost paths which enables to overcome to some extent occasional false negative labeling in between longer chains of true positive segments.

Marked Point Process (MPP) model objects within a stochastic framework and are adapted by incorporation of geometrical constraints in the solution in a manner which enables dealing with correlated and geometrical noise (DESCOMBES & ZERUBIA 2002). Models based on marked point process have provided convincing results in various applications like population counting, structure detection and texture representation (LAFARGE et al. 2010). The recognition targets are created as parametric objects, e.g., in (STOICA et al. 2004), roads are treated as sets of line segments with length and orientation. The small line segments are random sampled and moved to construct a road network by concerning the interaction between line segments. In this way, the figure of road can be remain well. CHAI et al. (2013) proposed a new point process taking junction point into consideration, which provides a structurally-coherent solution and performs functional well on different network extraction applications.

Besides aerial imagery, many researches are concerning with other kinds of remote sensing data for road network extraction, e.g. satellite imagery (SINGH & GARG 2013; YUAN et al. 2011), LiDAR (ZHAO & YOU 2012) , SAR images (LU et al. 2014; HE et al. 2013; LIU et al. 2013), 3D Point clouds (BOYKO & FUNKHOUSER 2011) and others (WANG et al. 2012).

Fig. 1: Superpixel segmentations with different capacity. First two images are segmented by SEEDs and the last two by SLIC.

3 Problem formulation

3.1 Image segmentation and points description

Since the remote sensing data are always in a large scale, we first generate superpixel segments and treat them as the entities for subsequent processing in this work. In this way, the search space is simplified and the computational load therefore reduced.

Usually a road network is regarded as a set of several long and narrow connected parts. In order to extract these properties from our data, we take use of the circumscribed ellipse to describe such irregular shaped segments. And the following features are taken into account, which characterize an ellipse:

- center coordinate, which refers to the location of the segments,
- rotation angel, which points out potentially, if the segment is a part of road, the orientation of the road,
- axes, which represent the scale of the segment.

Besides, we define a feature *ratio* as the quotient of minor axis to major axis of a circumscribed ellipse, $ratio \in (0,1]$, which describe the shape of the segment. A low ratio value indicates an elongated shape.

These superpixel segments are hereinafter referred to as points. And these points could be treated as “line segments”. Thus the image is represented as a set of marked points $X = \{x_i\}$, $x_i = (w_i, g_i, o_i, r_i)$, where w_i, g_i, o_i, r_i are location, gray value intensity, orientation and ratio of the point respectively. These parametric objects will be used later in marked point process.

3.2 Pre-classification using Conditional Random Fields

Conditional random field is a probabilistic model for labelling. We construct a simple pairwise CRF to pre-classify the image. The graph energy function is defined on unitary as well as pairwise cliques as:

$$E = \sum_{i \in \mathcal{V}} \Psi_i(x_i) + \sum_{(i,j) \in \mathcal{E}} \Psi_{ij}(x_i, y_j), \qquad (1)$$

where $\mathcal{V}$ refers to the set of points and $\mathcal{E}$ represents the set of adjacent point pairs.

The unary energy Ψ_i is derived by training a Random Forest Classifier (BREIMAN 2001). Considering the local attributes, the grayscale intensity as well as the ratio of the points are chosen

as feature of a point. The orientations of the points are still ignored here, because it is not useful for unary potential which does not consider context yet.
The binary energy $\Psi_{i,j}$ is formed by a contrast sensitive Potts model as:

$$\Psi_{i,j}(x_i, x_j) = \begin{cases} 0 & , \quad if\ x_i = x_j, \\ \theta_1 + (1 - \theta_1)e^{\theta_2 \|I_i - I_j\|^2}, & \quad otherwise, \end{cases} \tag{2}$$

where θ_1 and θ_2 are hyperparameters learned during training. I_i and I_j are the feature vectors of point i and j, respectively. The extracted grey values and the orientation properties are applied as features. The model punishes the labelling inconsistence between adjacent points with similar optical performance and a high likelihood to be connected road parts.
Finally, alpha-expansion graph cut algorithm (BOYKOV et al. 2001) is applied to infer the labelling. Many researchers have been working on the topic of CRFs for road extraction and shown a better result (WEGNER et al. 2013). Since in our framework, CRFs are only used as pre-classification, we take only simply model and no more details are discussed here.

3.3 Marked Point Process

A Point process F describes a random configuration of points in a continuous bounded set K. For $n \in \mathbb{N}$, let Ω_n be the set of all unordered configurations $\omega = \{\omega_1, \dots \omega_n\}$ that consists n unordered points $\omega_i \in K$. A point process is a measurable map from a probability space into the set of configurations Ω. Marked Point Process is extended from a point process with marks. Adding each point p_i with additional parameters m_i associates the point with an object $x_i = (p_i, m_i\}$, thus the geometric features of images can be modelled.
Comparing to (LACOSTE et al. 2005), the problem of geometric constrains in this work are simplified. The spatial interaction behavior can be ignored, while the relationship between two different points are only adjacent or not.

3.3.1 Data energy

A probability density $h(\cdot)$ can define the consistency of data and spatial independence between points. The total energy of the model is defined as:

$$U(x) = \sum_{x_i \in X} D(x_i) + \sum_{x_i, x_j \in X} V(x_i, x_j), \tag{3}$$

where $D(x_i)$ is the unary data energy measuring the coherence of the object x_i with respect to image and $V(x_i, x_j)$ measures the quality of a pairwise interaction between objects.

As shown in Fig. 1, the pixels on the road usually have high responses ratios and similar colors. Considering these two assumptions, the data consistency intensity is defined as: $h_d(x) \propto h_g(x) \cdot h_r(x)$, where h_g and h_r are the grayscale and ratio value distributions learned from samples by means of normalized histograms. And the data consistency energy is represented as: $D(x_i) = -\log h_d(x_i)$.

With respect to the graph connectivity, three different criteria are considered: the orientation similarity, the degree of adjacency and a self-defined distance between two points. The pairwise energy is defined as:

$$V(x_i, x_j) = E_{conn}(x_i, x_j) \cdot E_{adj}(x_i, x_j) \cdot E_{dist}(x_i, x_j). \tag{4}$$

Unlike the complex interaction model in (STOICA et al. 2004), we are interested only in the local neighborhood. As introduced in Sec. 3.2, we penalty a pair of adjacent points with great difference in orientation with

$$E_{conn}(x_i, x_j) = 1 - \sin(\|o_i - o_j\|), \tag{5}$$

where o_i is the orientation of x_i.

In some cases rooftops of buildings look quite similar to adjacent road segments. They feature similar or even parallel orientation like the road segments. In order to filter such objects, we enforce a collinearity constraint along roads by calculating the distance from a point to the other adjacent point located line. For a pair of points, the distance from a point to another point standing line is not symmetric. Hence we take the energy as the average of the bidirectional distance:

$$E_{dist}(x_i, x_j) = \frac{1}{2}\left(dist(x_i, x_j) + dist(x_j, x_i)\right), \tag{6}$$

where $dist(i, j)$ is the function calculating the distance from i to the line decided by j and its orientation.

E_{adj} is a term referring to the sum of lengths of the edges that two points have in common. As can be seen from Figure 1, the dark gray segments are connected on the shorter side and thereby construct a road network. Therefore, we decide the adjacent relevant energy as:

$$E_{adj}(x_i, x_j) = \min\left\{\frac{c}{N_{i,j}}, 1\right\}, \tag{7}$$

where $N_{i,j}$ refers to the number of adjacent pixels in x_i and x_j, and c is a factor to ensure the value of E_{adj} in a proper range.

Thus two points in the near with similar orientation can be characterized together.

3.3.2 Simulation

The maximum density is usually estimated by a RJMCMC sampler (GREEN 1995) coupled with a simulated annealing in CHAI et al. (2013) and STOICA et al. (2004). In this way, the global optimization can search over varying dimensions. This sampler simulates a discrete Markov Chain $(X_t)_{t\in\mathbb{N}}$ on the configuration space Ω, converging towards an invariant measure specified by U. The conventional RJMCMC sampler performs successively on objects, but takes long time, especially for large scale problems.

Conventionally, in point process, a Poisson or a uniform distribution are used as reference distribution. For Poisson process, the number of points follows a discrete Poisson distribution and the position of them is uniformly and independently distributed. In image analysis applications, uniform distribution is used, while all the points are treated equally in an image. VERDIE & LAFARGE (2012) proposed a new parallel sampler integrated with data-driven partition, which allow the points non-uniformly distributed in the image. In our model, the reference distribution is determined by the results from previous CRF classification. After CRFs pre-classification, the image parts are labelled as roads or backgrounds. With these characteristics, the points of images are distinguished by sampler and treated as non-uniform distributed. The points classified as roads are signed with high possibility. And the interested points with high possibility should be sampled frequently and has a high acceptancy.

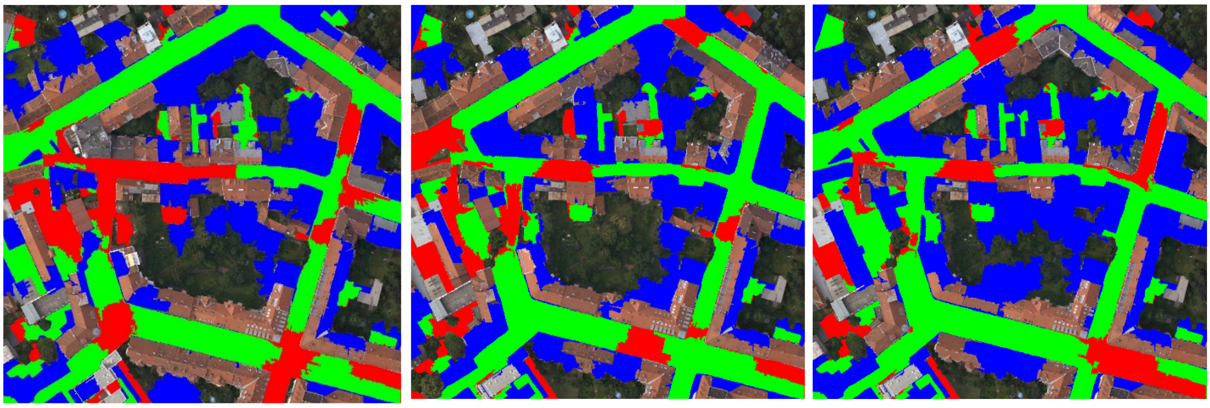

Fig. 2: Proposed road extraction results. From left to right [1]: using simple CRFs; [2]: using MPP with CRFs; [3]: using Poisson point process. True positives are displayed green, false positives blue and false negatives red

4 Experiments

We evaluate our experiments on the dataset of the city Graz, Austria. Each image from the dataset is a tile of the aerial true mosaics with 1000×1000 pixels. The simulation process are dealt with the open source C++ library librjmcmc (BRÉDIF & TOURNAIRE 2012).

The images are first segmented into superpixels. We apply two algorithms SEEDs (VAN DEN BERGH et al. 2012) and SLIC (ACHANTA et al. 2012) respectively. As shown in Fig. 1, SLIC offers a more regular and compacter superpixels with a low computation. The superpixel generated by SEEDs are usually so irregular, that the major axes of their circumscribed ellipses are long. Hence, the points with low ratios are dominant, which has a bad effect on the quality of results.

For the purpose of studying the effect of the number of points in the model, we segment the image into about 200 superpixels and 600. From Fig.1, more details are separated out such as the parked cars on roads. Since the scale of each superpixel is smaller than before, most of the points are closer to the similar. On this aspect, the shape feature like ratios and orientation cannot describe the attributes of the points reasonable and hence it is not proper to our model. On the other hand, we focus on a fast resolution of the road network extraction problem. Considering the time and space computation, the segmentation leading to a lower number of segments is taken in this work.

The quality of results are measured by the following generally used standard criteria for road network extraction. Completeness, defined as: Completeness = TP/(TP+FN), is the fraction of the road parts that are successfully extracted. And correctness estimates the percentage of the road parts extracted in results, as correctness = TP/(TP+FP). Quality is a balance accuracy in the form: quality = TP/(TP+FN+FP). Noticing that we only deal with the evaluation process pixel-wise.

First of all, comparing to the conventional methods, this proposal accelerate the simulation process in a significant degree. Due to the low data capacity, that 1000×1000 pixels are represented only as about 200 points, the results come out in about 75s on average, and for the same situation, other methods need at least 4 minutes (CHAI et al. 2013; VERDIE & LAFARGE 2012).

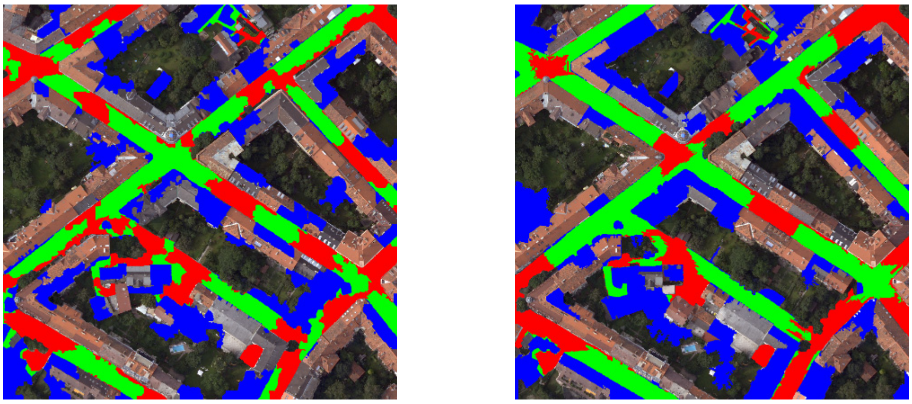

Fig. 3: Proposed road extraction results using SEEDs (left) and SLIC (right). The color indications are the same as in Fig.2.

Tab. 1: Quantitative evaluation of the proposed method. All numbers are percentages.

	Completeness	Correctness	Quality
[1]CRFs	64.80	46.21	36.94
[2]MPP	65.30	52.10	40.80
[3]Poisson	74.45	44.09	38.30

The road network extraction results are displayed in Figure 2 and estimated as Table 1. We first take use of our defined CRFs to get the pre-classification results, which is recorded as [1] CRFs. Then the marked point process is used and the results are marked as [2] MPP. It shows that the proposed MPP improves the result from CRFs on all the considered criteria. Some missing road parts are filled in and new paths are found on the left side of the image. Furthermore, we apply Poisson point process to the model with our proposed parametric points. We can find that with Poisson point process [3] provides a better completeness, by which more road parts are recognized, but the non-road background are false recognized either.
In addition, we evaluate the performance of our framework with different superpixel segmentation methods and the extraction results are shown in Figure 3 using SEEDs and SLIC respectively. Obviously using SLIC superpixels provides a better result than the points generated by SEEDs. The former extracts a network with a better connectivity, while SLIC segmentation usually provides more regular superpixels.

5 Conclusion and future work

We have proposed a new framework for extracting road networks from aerial images. In this work, shape features are extracted from the superpixel segments and then used in a simple pairwise CRFs model for image classification. The algorithms improves the sampling process by adopting the classification results and avoids complex geometric constrains. Comparing to the conventional marked point process relevant methods, the method improves the performance in terms of computing times and data scale.

The current results still have some problems needed to be fix. A better pre-classification results will lead to a more reasonable resolution. Some strategies are provided in many researches, e.g. using higher-order CRF, which can model the total energy of the road graph and ensure more about the connectivity. In the future work, we intend to study on more sophisticated CRFs as well as MPP model based on our point and feature.

6 References

ACHANTA, R., SHAJI, A., SMITH, K., LUCCHI, A., FUA, P. & SÜSSTRUNK, S., 2012: SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. IEEE Transactions on Pattern Analysis and Machine Intelligence **34** (11), 2274-2281.

BOYKO, A. & FUNKHOUSER, T., 2011: Extracting roads from dense point clouds in large scale urban environment. ISPRS Journal of Photogrammetry and Remote Sensing **66** (6), 2-12.

BOYKOV, Y., VEKSLER, O. & ZABIH, R., 2001: Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence **23** (11), 1222-1239.

BRÉDIF, M. & TOURNAIRE, O., 2012: Librjmcmc: An open-source generic c++ library for stochastic optimization. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **39** (B3).

BREIMAN, L., 2001: Random Forests. Machine Learning **45** (1), 5-32.

CHAI, D., FÖRSTNER, W. & LAFARGE, F., 2013: Recovering Line-Networks in Images by Junction-Point Processes. Conference on Computer Vision and Pattern Recognition (CVPR), 1894-1901.

DAL POZ, A.P., GALLIS, R.A., DA SILVA, J. & MARTINS, É., July 2012: Object-Space Road Extraction in Rural Areas Using Stereoscopic Aerial Images. IEEE Geoscience and Remote Sensing Letters **9** (4), 654-658.

DESCOMBES, X. & ZERUBIA, J., 2002: Marked point process in image analysis. IEEE Signal Processing Magazine **19** (5), 77-84.

GREEN, P.J., 1995: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika **82** (4), 711-732.

HE, C., YANG, F., YIN, S., DENG, X. & LIAO, M., 2013: Stereoscopic Road Network Extraction by Decision-Level Fusion of Optical and SAR Imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **6** (5), 2221-2228.

HU, X., TAO, C. V. & HU, Y., 2004: Automatic road extraction from dense urban area by integrated processing of high resolution imagery and LIDAR data. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **35** (Part B3), 288-292.

HUANG, Z., XU, F., LEI, L. & NIE, H., 2014: Object-based conditional random fields for road extraction from remote sensing image. IOP Conference Series: Earth and Environmental Science **17** (1).

JIA, C.-L., JI, K.-F., JIANG, Y.-M. & KUANG, G.-Y., 2005: Road extraction from high-resolution SAR imagery using Hough transform. IEEE International Geoscience and Remote Sensing Symposium **1**, 336-339.

LACOSTE, C., DESCOMBES, X. & ZERUBIA, J., 2005: Point processes for unsupervised line network extraction in remote sensing. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (10), 1568-1579.

LAFARGE, F., GIMEL'FARB, G. & DESCOMBES, X., 2010: Geometric Feature Extraction by a Multimarked Point Process. IEEE Transactions on Pattern Analysis and Machine Intelligence **32** (9), 1597-1609.

LAFFERTY, J., MCCALLUM, A. & PEREIRA, F. C., 2001: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data.

LIU, J., SUI, H., TAO, M., SUN, K. & MEI, X., 2013: Road extraction from SAR imagery based on an improved particle filtering and snake model. International Journal of Remote Sensing **22**, 34.

LU, P., DU, K., WANG, R., DENG, Y. & BALZ, T., 2014: A New Region Growing-Based Method for Road Network Extraction and Its Application on Different Resolution SAR Images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **7** (12), 4772-4783.

MONTOYA-ZEGARRA, J.A., WEGNER, J.D., LADICKY, L. & SCHINDLER, K., 2014: Mind the gap: modelling local and global context in (road) networks. 36th German Conference on Pattern Recognition, GCPR Proceedings, 212-223.

SHI, W., MIAO, Z. & DEBAYLE, J., 2014: An Integrated Method for Urban Main-Road Centerline Extraction From Optical Remotely Sensed Imagery. IEEE Transactions on Geoscience and Remote Sensing, 3359 - 3372.

SINGH, P.P. & GARG, R., 2013: Automatic Road Extraction from High Resolution Satellite Image using Adaptive Global Thresholding and Morphological Operations. Journal of the Indian Society of Remote Sensing **41** (3), 631-640.

SONG, M. & CIVCO, D., 2004: Road Extraction Using SVM and Image Segmentation. Photogrammetric Engineering & Remote Sensing **12**, 1365-1371.

STOICA, R., DESCOMBES, X. & ZERUBIA, J., 2004: A gibbs point process for road extraction from remotely sensed images. International journal of computer vision **57** (2), 121-136.

VALERO, S., CHANUSSOT, J., BENEDIKTSSON, J., TALBOT, H. & WASKE, B., 2010: Advanced directional mathematical morphology for the detection of the road network in very high resolution remote sensing images. Pattern Recognition Letters **31** (10), 1120-1127.

VAN DEN BERGH, M., BOIX, X., ROIG, G., DE CAPITANI, B. & VAN GOOL, L., 2012: SEEDS: Superpixels Extracted via Energy-Driven Sampling. Computer Vision-ECCV 2012, 13-26.

VERDIE, Y. & LAFARGE, F., 2012: Efficient Monte Carlo Sampler for Detecting Parametric Objects in Large Scenes. Computer Vision – ECCV 2012, Firenze, Italy, Springer, 539-552.

WANG, H., CAI, Z., LUO, H. & WANG, C., 2012: Automatic road extraction from mobile laser scanning data. International Conference on Computer Vision in Remote Sensing (CVRS), 136-139.

WEGNER, J.D., MONTOYA-ZEGARRA, J.A. & SCHINDLER, K., 2013: A Higher-Order CRF Model for Road Network Extraction. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1698-1705.

WEGNER, J.D., MONTOYA-ZEGARRA, J.A. & SCHINDLER, K., October 2015: Road networks as collections of minimum cost paths. ISPRS Journal of Photogrammetry and Remote Sensing **108**, 128-137.

WIEDEMANN, C., HEIPKE, C., MAYER, H. & JAMET, O., 1998: Empirical evaluation of automatically extracted road axes. Empirical Evaluation Techniques in Computer Vision, 172-187.

YUAN, J., WANG, D., WU, B., YAN, L. & LI, R., 2011: LEGION-Based Automatic Road Extraction From Satellite Imagery. IEEE Transactions on Geoscience and Remote Sensing **49** (11), 4528-4538.

ZHANG, G. & JIA, X., 2012: Simplified Conditional Random Fields With Class Boundary Constraint for Spectral-Spatial Based Remote Sensing Image Classification. IEEE Geoscience and Remote Sensing Letters **9** (5), 856-860.

ZHAO, J. & YOU, S., 2012: Road network extraction from airborne LiDAR data using scene context. Providence: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.

ZHONG, Y., ZHAO, J. & ZHANG, L., 2014: A Hybrid Object-Oriented Conditional Random Field Classification Framework for High Spatial Resolution Remote Sensing Imagery. IEEE Transactions on Geoscience and Remote Sensing **52** (11), 7023-7037.