Multi-scale Hierarchical Conditional Random Field for Railway Electrification Scene Classification Using Mobile Laser Scanning Data

Leihan Chen

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

GRADUATE PROGRAMME IN EARTH AND SPACE SCIENCE
YORK UNIVERSITY
TORONTO, ONTARIO

AUGUST, 2018

Abstract

With the recent rapid development of high-speed railway in many countries, precise inspection for railway electrification systems has become more significant to ensure safe railway operation. However, this time-consuming manual inspection is not satisfactory for the high-demanding inspection task, thus a safe, fast and automatic inspection method is required. With LiDAR (Light Detection and Ranging) data becoming more available, the accurate railway electrification scene understanding using LiDAR data becomes feasible towards automatic 3D precise inspection.

This thesis presents a supervised learning method to classify railway electrification objects from Mobile Laser Scanning (MLS) data. First, a multi-range Conditional Random Field (CRF), which characterizes not only labeling homogeneity at a short range, but also the layout compatibility between different objects at a middle range in the probabilistic graphical model is implemented and tested. Then, this multi-range CRF model will be extended and improved into a hierarchical CRF model to consider multi-scale layout compatibility at full range. The proposed method is evaluated on a dataset collected in Korea with complex railway electrification systems environment. The experiment shows the effectiveness of proposed model.

Acknowledgements

I would like to extend my sincere gratitude to everyone who supported and helped me during my master program.

At first, I want to give a special appreciation to Dr. Gunho Sohn, my supervisor, for his clear guidance, useful suggestions, and challenging discussions which help me to learn how to study and step into the door of scientific research. His continuous support, mentally and financially, for both my research and life, which created a great environment for my study and research. His endless academic passion impresses and sharpens me to overcome all the difficulties in my research.

I would like to thank Dr. Costas Armenakis, as my supervision committee member, for his continuous scientific suggestions, valuable advice, and warm encouragement. His guidance also helps me to understand my research better from multiple perspectives. I also want to express my gratitude to Dr. Florent Lafarge, who was my supervisor at INRIA Sophia Antipolis, to guide my research-based internship. His plentiful knowledge and rigorous research attitude guides me how to do research independently and effectively.

I also want to show my gratefulness to all members in our lab, Kiin Bang, Heungsik Kim, Jaewook Jung, Connie Ko, Chao Luo, Mojgan Jadidi, Kivanc Babacan, Ali Baligh, Phillip Robbins, Eros Gulo, and Kang Zhao. Especial thanks to Dr. Jaewook Jung, who helped me a lot in my research in data preparation, algorithm implementation and writing guidance. Also, the gratitude to Mr Kivanc Babacan, who gave me so many valuable suggestions both academically and spiritually as a life-long friend.

I wish to thank Thales Canada (Michael Rickard), Teledyne-Optech (Chris Verheggen), Ontario Centres of Excellence (OCE) VIP II Program, Natural Sciences and Engineering Research Council of Canada (NSERC) CREATE Data Analytics and Visualization (DAV) for arranging the financial support for my research. I also appreciate Mitacs Global Link to support my internship in France to enrich my research experience.

Finally, I deeply thank my parents, Yan Xia and Zhifei Chen, for their endless love and continuous encouragement. I especially want to express my miss to my father, Zhifei Chen, who left me forever last year, your love will remain in my heart forever.

Content

Abstract ii

Acknowledgements iii

Content v

List of Tables ix

List of Figures x

List of Abbreviation xiv

Chapter 1 Introduction 1

1.1 Motivation 1

1.2 Research Objectives 5

1.3 Methodology Overview 6

1.4 Contributions 9

1.5 Thesis Outline 10

Chapter 2 Background 12

2.1 Automatic Railway Recognition 13

2.1.1 Railroad Classification and Inspection 13

2.1.2 Railway Object Detection 14

2.1.3 Railway Scene Classification 15

2.2 Context Based Classification 16

2.2.1 Statistical Context Information 16

2.2.2 Probabilistic Graph Model 18

2.2.2.1 Markov Random Field 19

2.2.2.2 Conditional Random Field 22
2.3 Chapter Summary 26
Chapter 3 Multi-Range Conditional Random Field 27
3.1 Graphical Model Design 28
3.2 Definition of the Graph 29
3.3 Unary Term 31
3.4 Pairwise Terms 33
3.4.1 Short-Range Binary Term 33
3.4.2 Middle-Range Binary Term 33
3.4.2.1 CRF Based on Vertical Layout Compatibility 34
3.4.2.2 CRF Based on Horizontal Layout Compatibility 36
3.5 Training and Inference 38
3.5.1 Parameter Estimation 39
3.5.2 Inference 42
3.6 Experiment Results 42
3.6.1 Line Extraction Results 46
3.6.2 Classification Results 48
3.6.2.1 SVM Classification Results 52
3.6.2.2 SrCRF Classification Results 53
3.6.2.3 MrCRF Classification Results 54
3.7 Discussion 56
3.8 Chapter Summary 59

Chapter 4 Multi-Scale Hierarchical Conditional Random Field 61
4.1 Multi-scale Data Representation 62
4.2 Graph Model Design 64
4.3 Bottom and Middle Layers 66
4.3.1 Short-Range Binary Term 67
4.3.2 Middle-Range Binary Term 68
4.4 Fully Connected Top Layer 70
4.5 Training and Inference 73
4.5.1 L-BFGS Based Training 74
4.5.2 Maximum Posterior Marginal Inference 75
4.6 Experiment and Results 75
4.6.1 Multi-Scale Line Extraction 76
4.6.2 Classification Results 77
4.6.3 Discussion 84
4.6.3.1 Improvement due to local smoothness and relative spatial location 85
4.6.3.2 Classification performances at different scales 87
4.6.3.3 Improvement due to fully-connected top layer 88
4.7 Chapter Summary 91
Chapter 5 Conclusions and Future Works 93
5.1 Conclusions 93
5.2 Future Works 95

References.....98

List of Tables

Table 3.1 Specifications of the Trimble MX8 system.....43

Table 3.2 Extracted lines for each sub-region. 47

Table 3.3 Edge number of different models in different sub-regions. 50

Table 3.4 Confusion matrix for SVM results: electricity feeder (EF), catenary wire (CAW), contact wire (COW), current return wire (CRW), connecting wire (CNW), suspension insulator (SI), movable bracket (MB), dropper (Dro), pole (Pole) and ground (Gro). 53

Table 3.5 Confusion matrix for SrCRF results: electricity feeder (EF), catenary wire (CAW), contact wire (COW), current return wire (CRW), connecting wire (CNW), suspension insulator (SI), movable bracket (MB), dropper (Dro), pole (Pole) and ground (Gro). 54

Table 3.6 Confusion matrix for MrCRF results: electricity feeder (EF), catenary wire (CAW), contact wire (COW), current return wire (CRW), connecting wire (CNW), suspension insulator (SI), movable bracket (MB), dropper (Dro), pole (Pole) and ground (Gro). 56

Table 3.7 A summary of the classification performance achieved by three classifiers (unit: %). 57

Table 4.1 Line number in three scales for each sub-region.....77

Table 4.2 Confusion matrix of HiCRF model 79

Table 4.3 Classification result of proposed model and SVM.....83

Table 4.4 Comparison between BlCRF and TlCRF.....87

Table 4.5 Comparison between TlCRF and BlCRF.....89

List of Figures

Figure 1.1 Overhead line's inspection in Hungary. (Innoteka, 2012).....2

Figure 3.1 Neighboring systems: (a) for short-range graph; (b) for middle-range graph; and (c) combined neighboring systems.....30

Figure 3.2 Prior and likelihood estimation of the vertical middle-range pairwise potential: (a) look-up table where row *i* is below column *j*; and (b) probability density distribution of the height difference (red curve: estimated Gaussian distribution learned from the training data).....36

Figure 3.3 Prior and likelihood estimation of the horizontal middle-range pairwise potential: (a) look-up table where row *i* is closer to the railway than column *j*; and (b) probability density distribution of the verticality difference (red curve: estimated Gaussian distribution learned from training data).....38

Figure 3.4 Trimble MX8 mounted on a train.....43

Figure 3.5 Electrification system configuration and 10 object classes of the Honam high-speed railway: (a) a photograph; and (b) Mobile Laser Scanning (MLS) data.....45

Figure 3.6 Manually-classified reference data: (a) entire region; and (b–g) six sub-regions (different colors represent different classes).....46

Figure 3.7 Example of voxelization and line extraction: (a) input MLS data and rail vectors, (b) voxelization; and (c) extracted lines.....47

Figure 3.8 Classification results produced over the entire railway test scene: (a) SVM; (b) short-range Conditional Random Field (CRF); and (c) multi-range CRF; electricity feeder (black), catenary wire (blue), contact wire (red), current return wire (sky blue), connecting

wire (dark green), suspension insulator (brown), movable bracket (magenta), dropper (green), pole (grey) and ground (orange). 48

Figure 3.9 Examples of the adjacency graph: (a) short-range graph; and (b) middle-range graph. 50

Figure 3.10. Weight learning of MrCRF using the SGD algorithm. 55

Figure 3.11 The comparison of classification result: (a) the dropper in the SVM results (left) was classified to movable bracket in the integrated CRF model (right); (b) the ground in the SrCRF (left) was classified to pole in MrCRF model (right); (c) the suspension insulators were well rectified in MrCRF model. Black circle represents the misclassified line primitives in SVM corrected by MrCRF. 59

Figure 4.1 Flow chart of line merging algorithm to generate multi-scale representation. 63

Figure 4.2 Hierarchical three-layer CRF (HiCRF) model. The circle represents node as line primitives and the size of the circle depicts the line primitive scale. Blue lines demonstrate intra-layer edges while red dotted lines show inter-layer edges. All edges are pairwise (without high-order cliques). 66

Figure 4.3 Example of relative displacement vector. (a) shows the relative displacement vector distribution for the electricity feeder and catenary wire in horizontal and vertical direction. (b) shows movable bracket 'self' spatial distribution. 70

Figure 4.4 Learnt displacement map. (a) 3D relative displacement probability map for the class pair pole and movable bracket. (b) a table of all conditional probabilities at a certain relative displacement. Darker means higher probability. 73

Figure 4.5 Three-scale line representation. (a) is the original pint cloud in subset 1 and (b) is the extracted line primitives using RANSAC algorithm. (c) shows the merged line primitives and (d) represents final merged line primitives with relaxer thresholds. (black: electricity feeder, blue: catenary wire, red: contact wire, sky blue: current return wire , dark green: connecting wire, brown: suspension insulator, magenta: movable bracket, green: dropper, grey: pole and yellow: ground). 77

Figure 4.6 Multi-layer hierarchical classification result: (a) is the visualization of the classification result of subset 1 and (b) shows the misclassified line primitives in subset 1 with red and correct classification with black color. (c) and (d) presented similar results for subset 5. 79

Figure 4.7 The average quantitative classification result of different models..... 81

Figure 4.8 Label transition from SVM classification result to HiCRF model. EF (electricity feeder), CAW (catenary wire), COW (contact wire), CRW (current return wire), CNW (connecting wire), SI (suspension insulator), MB (movable bracket), Dro (dropper), Pol (pole) and Gro (ground)..... 84

Figure 4.9 Comparison of different classification model result in subset 4. The classification results were produced by (a) SVM, (b) CRF with short range and (c) CRF with short and middle range: Each colored circle shows classification errors produced for different object classes. 86

Figure 4.10 The classification result of subset 3: (a) is the result of BlCRF model while (b) shows the result of TlCRF model. Red rectangle means misclassification in BlCRF but corrected in TlCRF and red circle is opposite. 88

Figure 4.11 Examples of misclassification occurred in TlCRF corrected by HiCRF in subset 5: The left column is the classification of TlCRF while the right column is the corresponding result of HiCRF. (a), (b) are left view and the rest are front view. (sky blue: current return wire; black: electricity feeder; blue: catenary wire; green: dropper; brown: suspension insulator; magenta: movable bracket; grey: pole). 90

List of Abbreviation

ALS	Airborne Laser Scanning
AMN	Associative Markov network
CRF	Conditional Random Field
DT	Decision Tree
EM	Expectation Maximization
GLCM	Gray-Level Co-occurrence Matrix
HMM	Hidden Markov Model
L-BFGS	Limited Memory Broyden–Fletcher–Goldfarb–Shanno
LBP	Loopy Belief Propagation
LiDAR	Light Detection and Ranging
MAP	Maximum A Posterior
ML	Maximum Likelihood
MLS	Mobile Laser Scanning
MPM	Maximum Posterior Marginal
MRF	Markov Random Field
RANSAC	Random sample consensus
RBF	Radial Basis Function
RF	Random Forest
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
TLS	Terrestrial Laser Scanning

Chapter 1
Introduction

1.1 Motivation

Since the first industrial revolution, railway network gradually became a major component of transportation system and the dominant transportation tool in many countries. Nowadays, railroad transportation is still one most significant transportation tools both domestically and regionally. In United States, approximately 4,758 million railway trips occurred annually, and the average trip length is around 6.5 miles in 2012 (APTA, 2014). In Canada alone, the railway systems cover 50,000 km of main track-line and transport 75 million people, as well as $250 billion worth of goods each year. The high-speed railway systems have been rapidly developed in some countries, such as France, Japan and China in last three decades. Japan is the first country in the world with more than 20 billion trips annually and China will construct up to 38,000 km high-speed railway until 2025 (Railway Gazette, 2016). Therefore, railway can still play a key role in public transportation in the future. To provide electric power for railway trains, a railway electrification system is indispensable. In fact, these systems already comprise one third of the global rail track systems. A railway electrification system supplies electric power to railway trains without an on-board prime mover or local fuel supply, which is advantageous over diesel-based systems for faster acceleration and regenerative braking. (Gómez-Expósito et al., 2014). In these systems, overhead transmission lines are designed to transmit electricity from feeder station and contains several types of transmission lines. A utility pole is a post-like object

to support overhead wires for electric power distribution and are associated with objects such as suspension insulators and movable brackets. These elements usually have some evident designs and distributions, which cause the spatial regularity of railway electrification system. Regular maintenance and inspection is indispensable to ensure a constant supply of electricity power to all trains on track. However, this task is traditionally labor-intensive, time consuming and dangerous for human operators, which impedes frequent large-scale inspection. Conventionally, human inspectors usually use crane to manually check whether the overhead lines satisfy the regulation requirement as shown in Figure 1.1.

Figure 1.1 Overhead line's inspection in Hungary. (Innoteka, 2012)

Recently, various techniques using remotely-sensed data, such as images, Airborne Laser Scanning (ALS) data and Mobile Laser Scanning (MLS) data, have been introduced to supplement or replace humans' visual inspection. The remote-sensed data can be obtained at regular basis to provide updated information and it is much more accessible for inspectors to check the system without physical presentence. Particularly, MLS data provide very accurate and high-density point clouds over the railway scene scanned by laser scanners mounted on a train or inspection cart. Compared with other data sources, MLS data collection is faster, cheaper, more replicable and it will not be influenced by extreme weather conditions. Therefore, MLS data is suitable for railway electrification inspection tasks.

As MLS is a relative young technology for infrastructure surveying, there are both hardware and software problems need to be solved. Different research topics related with MLS data processing are conducted, including point cloud registration (Gálai et al., 2016; Gressin et al., 2013; Men et al., 2011), point cloud to image registration (Cui et al., 2016; Li et al., 2018) , geo-referencing (Gao et al., 2015), segmentation (Nurunnabi et al., 2016; Zhou et al., 2014), classification (Golovinskiy et al., 2009; Najafi et al., 2014; Xiong et al., 2011; Zhu and Hyyppa, 2014), and change detection (Qin and Gruen, 2014; Xiao et al., 2015; Xiao et al., 2016).

For automatic railway electrification system inspection from MLS data, the first step is to classify point clouds into meaningful railway objects. After the accurate classification of railway scene objects, important objects can be extracted and modeled. By comparing the modeled objects with existed models, automatic inspection can be achieved

to determine whether the objects still satisfy design requirement. Therefore, accurate classification of railway electrification objects is the first important step towards automatic inspection. Semantic segmentation or classification of mobile laser point cloud is always a popular research topic due to the importance of semantic information for all different tasks such as modeling, detection and scene understanding. However, as indicated in previous research (Grilli et al., 2017), the large-scale MLS classification for railway electrification system has to overcome several problems:

- ***Scene complexity***: Railway scene is a complex scene with different objects varying from their sizes and shapes. Meanwhile, railway electrification system is a complicated structure with specific design requirements. Although there are spatial relations among objects existed in the scene, this information is also hard to be encoded and be applied for classification system.
- ***Large dataset***: MLS data has high point density compared with data obtained from other LiDAR platforms. Majority of mobile laser equipment can measure more than 600,000 points per second which causes the volume of dataset reach millions of points within hundred meters on the railway.
- ***High accuracy requirement***: To avoid potential operation risk, all railway electrification system objects should rigorously follow the accuracy requirements of the designs. Therefore, the inspection should distinguish any minor displacements and deformations, also requires an extremely accurate classification result for all objects.

To overcome these limitations, a solid classification algorithm should be proposed to handle large-scale dataset and complex scene structures. Fortunately, even though railway infrastructure consists of complex objects, it has strong spatial regularities among railway elements (Jung et al., 2016). For instance, two consecutive poles have an almost fixed distance along railway track direction; contact wire is above the center of the rail with a certain height; the catenary wire is also above the contact wire; dropper connects between catenary wire and contact wire; the current return wire and pole are located outside rails. This relative spatial information is vital in railway scene to help the classification system understand the scene and overcome the geometric ambiguities among different objects.

1.2 Research Objectives

The main objective of this thesis is to develop an automatic, accurate and robust classification algorithm to classify railway electrification system objects in MLS data. This objective can be clarified in two points:

1. **Classify railway electrification scene with relative spatial regularities**: Considering some strong relative spatial regularities both vertically and horizontally among railway electrification scene, a supervised method should be appropriate to learn this information automatically and then the learnt information should be utilized to impose this relative spatial regularity restriction to enforce the classification result follow the similar spatial distribution.
2. **Enhance learnt relative spatial regularities**: The learnt spatial regularities (Gould et al., 2008; Jung et al., 2016; Luo and Sohn, 2014; Roig et al., 2011) is usually

primitive-based and this spatial regularity can be imprecise because spatial regularity should be instance-based in the railway scene. Due to the difficulty of obtaining instance information, a hierarchical spatial relation is introduced to overcome the limitation of local primitives. Also, the spatial regularities should be learnt in the full scene but not limited to specific ranges. Especially the very long-range relation in railway electrification scene (For instance, the distance between adjacent poles on the railway direction) is usually hard to encode. The objective is improving and correcting learnt spatial regularities constrained within fixed-size primitives and expanding it to a full-range spatial regularity.

1.3 Methodology Overview

This thesis proposed a classification algorithm for targeting railway electrification systems using mobile laser scanning data and input data is a 3D point cloud with (X, Y, Z) coordinates. To efficiently handle with large volume of mobile laser scanning data, lines were extracted to represent original point cloud and it will be used as classification unit. This line or sketch based recognition process has also been widely studied in image data (Eitz et al., 2011; Yang et al., 2005; Yoon et al., 2010) and it has shown that this line based representation can also be effective for object description and recognition. Considering many key objects such as overhead lines have linear properties, line-based representation can significantly reduce the complexity of the data and preserve objects' geometrical properties.

First, the entire MLS point cloud is transformed into 3D voxels. Within each voxel, a RANSAC algorithm was applied to extract multiple lines. Several geometric features were extracted which include orientation, point density and distance. Then a general discriminative classifier Support Vector Machine (Samadzadegan et al., 2010) was selected to classify line segments based on its geometrical features.

To overcome the problem of feature ambiguities in local classifiers, a discriminative graph probabilistic framework Conditional Random Field (Lafferty et al., 2001) was applied to encode contextual information to enhance the SVM classification result. Specifically, a multi-range Conditional Random Field (MrCRF), which considered layout compatibility between railway elements, as well as local smoothness, was introduced. SVM was used as the unary potential of the MrCRF model. Two different graphs were designed to consider local smoothness and long-range spatial regularities in the horizontal and vertical directions. The MrCRF integrated both local smoothness and layout compatibilities in the vertical and horizontal directions.

There are two types of parameters in this MrCRF model: the first type is the parameters in the long-range term, while the other is the weights between different the sub-terms. Considering the challenge of simultaneously learning all parameters, a piecewise training strategy (Shotton et al., 2009) was utilized to train these two types of parameters. Parameters in unary term and each pairwise terms were learned individually at first while the weights of all CRF potential terms were learned using Stochastic Gradient Descent (SGD). Once the training is done, an approximate inference algorithm, Loopy Belief

Propagation (Murphy et al., 1999) was used for inference; and the final class label was selected by maximizing the node belief.

However, the spatial regularities learnt in MrCRF model are calculated based on the local line primitives and it is usually not accurate and the spatial regularities are restricted to certain ranges. To overcome these limitations, a three-layer multi-scale hierarchical CRF (HiCRF) classification framework was investigated. A seed-filling line merging strategy was proposed to merge line primitives based on its geometry properties. With different merging criteria, a three-scale line representation was generated, which will serve as nodes in proposed HiCRF respectively. A multi-range CRF model is used for first two finer line primitives' classification within the bottom and middle layer. A fully connected CRF model with relative displacement prior potential will be applied for the classification of final coarsest line primitives within the top layer.

The simplified MrCRF model with the combination of vertical and horizontal terms in a Radial Basis Function (RBF) kernel similarity function. To accelerate training, I replaced SGD optimization with L-BFGS optimization, this can automatically determine gradient descent direction and mean field approximation inference was utilized because it had a faster convergence (Toyoda and Hasegawa, 2008).

The proposed model was evaluated on a MLS data of railway electrification system scene data collected in Korean high-speed railway. A variety of classification indices including confusion matrix, overall accuracy, completeness, correctness, quality and kappa index to evaluate the classification performance. To track how the specific class changes the classification result, a label transition was also conducted to analyze the influence of

different contextual information. With the comparison of different classification models, I demonstrated the effectiveness of the designed contextual information for correcting misclassification caused by local geometry ambiguities.

1.4 Contributions

The main contribution of this thesis is the introduction of a probabilistic framework to encode a multi-scale context information at full range of the scene. The specific contribution of this thesis can be summarized as follows:

- Proposing a multi-scale line representation of original mobile laser scanning data to reduce the computational burden of original massive point cloud data produced by mobile laser scanner. Inspired by Luo and Sohn (2014), a line representation can convert large volume of point cloud to several thousand lines but also well preserve the geometric detail of railway electrification scene objects. Meanwhile, this single scale line representation is extended to a multi-scale line representation which can also implicitly utilize scale information of objects in the scene.
- Learning spatial regularities automatically from training data in the full scene from training data. Also inspired from Chao and Sohn (2014), a multivariate gaussian distribution is assumed to express the likelihood of every two class pair distribution in the railway scene. A look-up table is also formulated to model prior distribution for every two class pair. All these distributions can be obtained from training data via Maximum Likelihood (ML) criterion and statistical calculation. However, this automatic spatial regularities learning is only restricted in a certain range. A

displacement probability map is later proposed to learning relative spatial displacement of every two line primitives at all ranges in the scene from training data. Test data will be encouraged to have same spatial regularities at all ranges in its scene. This probability displacement map also avoids this multivariate gaussian distribution assumption, which is not valid for all class pairs in the railway scene. Then, all pairwise spatial regularities can be learnt effectively and completely in the training data.

In this thesis, the single scale line extraction from mobile laser scanning data and feature extraction from line primitives are kindly provided by Jung (Jung et al., 2016). Meanwhile, Luo's work (Luo and Sohn, 2014) inspired this work from the idea of learning spatial regularities from multiple directions in the training data.

1.5 Thesis Outline

This thesis is organized into five chapters. An overview of the chapters is as follows:

Chapter 1 introduces the motivation and problem domain of this thesis, as well as the overview of proposed method.

Chapter 2 describes the background information and literature review of previous study. The comprehensive literature reviews will involve with the topic of railway tracking, railway scene understanding and CRF based contextual classification.

Chapter 3 proposes a multi-range CRF model for the railway electrification system classification. It firstly describes how line primitives are extracted from original mobile

laser scanning data and how to calculate geometric features to represent individual line primitives. Then, the formulation of both unary term and three pairwise terms in CRF model is described in the next step. Finally, the training and inference of the model will be discussed. The classification result of this multi-range CRF model is presented and analyzed.

Chapter 4 proposes a three-layer hierarchical CRF (HiCRF) model to overcome the limitations of multi-range CRF model. A line merging strategy is introduced to generate multi-scale line representation firstly and then the graph structure of HiCRF model is subsequently demonstrated. Formulation of individual layer in hierarchical CRF model is described later with the training and inference strategy. A comparison of different classification model is finally conducted and then a detailed analyse of the result reveals the rationality of proposed model.

Chapter 5 gives the conclusion of this study and potential future works.

Chapter 2
Background

Since the development of railway electrification system for power supply, more than one third of the railroads have been electrified globally. Thus, the inspection of railway electrification system at a regular basis became important for railway operation. However, traditional human visual inspection is dangerous and label-intensive. Therefore, remote-sensed data, especially mobile laser scanning data, is recently introduced to supplement or replace humans' visual inspection. As the first step of automatic railway electrification system inspection, the classification of railway electrification related objects is vital for the accuracy of automatic inspection. Although there exists the scarcity of the research about automatic classification of railway electrification objects, there are fundamental research within the topic of automatic railway environment understanding, which includes but not limits to railroad tracking, railway pole detection, transmission wire modeling and railway scene classification. Considering the strong contextual information existed in railway scenes, contextual-based classification method is a reasonable solution for railway scene classification. In this chapter, previous research works related to automatic railway environment understanding and contextual-based classification method will be reviewed. The first part of this chapter reviews previous research on a variety of topics related to railway recognition. The second part discusses existing works focusing on contextual-based classification but not limited to railway scene. Through this review, a conclusion can

be made that there is a research gap between the existing work and a fully automatic railway electrification scene classification solution.

2.1 Automatic Railway Recognition

Automatic railway recognition is an active research topic in the last few years within the field of photogrammetry, computer vision and civil engineering. This topic involves railroad classification and inspection, railway side object detection and classification, railway scene understanding and modeling, which will be reviewed respectively in next three subsections.

2.1.1 Railroad Classification and Inspection

Railroad classification and tracking is actively investigated with the use of LiDAR data. An automatic rail extraction method (Mohamad et al., 2013) combined terrestrial laser points and ALS data to model railway tracks as a dynamic system of local pairs of parallel line segments. Another algorithm (Beger et al., 2011) also used ALS data with airborne image to detect railroad. A railroad track mask was firstly generated from image, which assisted the classification of laser points combined with its height and spatial information. Railroad centerline was finally extracted from those classified points using an adopted random consensus algorithm. A similar strategy was also applied on the extraction of railroad centerline using only mobile laser scanning data (Jwa and Sohn, 2015). Naïve Bayes classifiers was applied for finding railway track points and then a Kalman filter was estimated to progressively model railroad centerline. These algorithms always relied on a previous classification of railway track points.

After the classification is the detection and tracking of railroad. There are some initial research works about railway inspection. A neural network based (Sacchi et al., 2001) algorithm was proposed to detect vandal acts, but the method is too restricted only in metro station with abnormal human behavior. Another neural network based (Marino et al., 2007) method relied on two multilayer perceptron neural classifiers (MLPNCs) to specifically classify and locate fastening bolts. The input is image set from a digital line-scan camera. The limitation of this method is that it can only detect bolts to determine the quality of railway tracks and it requires a lot of railway track images. Another fastener detection algorithm (Feng et al., 2014) detected and classify fasteners damage degree from images. The method detected track and sleeper first, then the extracted information was utilized to iteratively localize and model fasteners using a probabilistic structure topic model (STM). Their method is more robust and general for different railway tracks, but the result is significantly influenced by line detection quality for track and sleeper detection.

2.1.2 Railway Object Detection

Railway object detection was attracted a lot of attention recently as many specific objects along railway track are interesting for users. Power line is an important part of railway electrification system and a region growing based algorithm (Zhang et al., 2016) was proposed to detect them in MLS data. The powerline was classified according to the assumption that it should be parallel to railway tracks. The classified power line points were modeled by fitting the points to a polynomial model. Guo et al. (2016) proposed a power line reconstruction method based on the Random Sample Consensus (RANSAC) rule. Before the reconstruction of power lines, ALS data were classified into five categories

(power line, vegetation, building, ground and pylon) by applying the JointBoost classifier. Admitted that those results of power line detection algorithms were promising, but it did not distinguish different power line types, which is not satisfactory for precise further inspection task.

2.1.3 Railway Scene Classification

Railway scene understanding is a more challenging task for understanding the railway environment. Pastucha (2016) presented a method to detect and locate the overhead catenary system using mobile laser scanning data. Both point density and location information were utilized to detect the overhead catenary system related objects and then a modified DBSCAN algorithm was applied to classify original point cloud. Nevertheless, the overhead catenary system is composed of multiple objects and several types of electricity wires, which was not distinguished in their method. A more well-defined classification system (Arastounia, 2015) was presented to classify the whole scene into railway tracks, cantilevers and electricity cables. The method recognized those objects progressively with their geometric features and they achieved a good both object and point based accuracy. This method was completely unsupervised, and it needed a lot of parameter tuning to adjust different datasets. Another similar railway scene understanding algorithm (Zhu and Hyyppa, 2014) was introduced with different railway scene types. The approach utilized mobile laser scanning data for pole and building façade classification while ALS data was applied to get tree, ground and building roofs. Building roofs will combine with building facades to generate building models. The ground will refine through existing DTM

model. Their approach also met the generalization problem and required large amount of input data from different data sources.

2.2 Context Based Classification

Contextual information (Strat and Fischler, 1991) is defined as any information which influence the perception of objects or scenes within its domain. It has been proven effective to improve the classification results that consider only the unary appearance features. Usually context can be interpreted in visual perception level or objective statistical level. Recently, due to the technology of several graph probabilistic models, objective statistical level context dominates the context information research in the community of computer vision. In this review, only the classification methods related with objective statistical level context is reviewed. Meanwhile, there is a thorough review on how those graph probabilistic models is used for classification task.

2.2.1 Statistical Context Information

There are a variety of different context information types. Galleguillos and Belongie (2010) categorized the statistical context into three major types: semantic (probability), spatial (position) and scale (size).

Semantic context shows the compatibility of certain objects and certain scenes or other objects. For instance, cows are more likely to appear together with grassland but not airplane. Compared to early manual designed compatibility rules, current research prefers to extract semantic context automatically from labeled training data due to the difficulty of defining a lot semantic context rules. The symmetric, co-occurrence matrix is the most

common type of semantic context. In occurrence matrix, each element indicates the number of times that a given class occurs in a relation to another class. For example, the widely applied gray-level co-occurrence matrix (GLCM) defined the gray-level spatial dependence over pixels, which is widely used as a image texture descriptor. An extension of co-occurrence (Ladicky et al., 2010) was proposed under a probabilistic graph framework with computational-efficient inference to encode this semantic context generally for all class-pairs within the same scene.

Spatial context describes the likelihood of locating an object at certain positions. Spatial layout in a graph probabilistic framework (Shotton et al., 2009;Winn and Shotton, 2006) was encoded as the probability of pairwise class occurrence at the specific location in the image in a look-up table. Gould et al. (2008) encoded non-parametric relative location maps over super-pixels as a unary global feature, which can efficiently combine local and global features within a simple local framework. Also, their method is capable of modeling complex 2D spatial relationships, not only those simple relative location relations (above, beside, or enclosed), but also complex relationships among multiple objects at multiple directions. This multiple object layout (Desai et al., 2011) was also encoded as spatial histogram feature in several directions for object detection task.

Scale context refers to the information for object sizes in the scene. Metadata (e.g. position, orientation, geometric horizon, and calibration) of cameras (Strat and Fischler, 1991) can give some implicit information about object size through certain camera physical model. Nevertheless, scale context is still very hard to express due to the complexity of the object's geometry and potential occlusion. The recent object detectors may give some clues

in scale context and it has been explored (Li et al., 2009;Yao et al., 2012) to combine with other context information. These methods both use a pre-designed object detector to provide objects scale information.

2.2.2 Probabilistic Graph Model

Instead of encoding contextual information into a local unary feature, probabilistic graph model is a general framework to encode the contextual information directly as an interaction among multiple primitives in different data sources. Probabilistic graph model expresses multivariate probability statistics based on both the graph theory and probability theory (Bishop, 2006). By considering dependency of variables using potential functions, the graphical model successfully simplifies the design of a complex probabilistic system. The key idea of the graphical model is that it parameterizes the probability distributions into a graph factorization. Due to its great flexibility and universality, it has been widely applied to many fields, such as computer vision, bioinformatics, social network analysis and natural language processing.

In a probabilistic graph, nodes and edges are indispensable two elements. The nodes in the graph are random variables. As the random variable can be discrete or continuous, the probabilistic graph model can be discrete or continuous. But classification problem is predicting the variable hidden state from a discrete set, so all the graph probabilistic models applied in classification problem are discrete. Edge demonstrates the statistical dependencies between random variables. Due to the different properties of dependencies, the graph can be also directed or undirected. Directed graph only has directed edges to correlate nodes as a "parent-child" relation, which can be modeled using conditional

probabilities. Typical directed graph models are Bayesian Network and Hidden Markov Model (HMM). However, directed probabilistic graph model will have difficulty to model the random variables in which their causality cannot be fully observed (Pearl, 2014). In contrast, the undirected graphical model is much more flexible with this causality structure, which makes it more general to different problems. For undirected probabilistic graph model, clique is the concept that the subset of undirected graph in which nodes inside a clique are fully-connected. The key of undirected probabilistic graph model is expressing this probabilistic dependency among nodes factor multiplication inside of the clique. According to the size of maximum clique inside the graph, undirected probabilistic graph can be categorized as pairwise or high-order graph. There are two typical undirected probabilistic graph models, Markov Random Field (MRF) and Conditional Random Field (CRF), which will be respectively reviewed on their applications for image classification task in next two subsections.

2.2.2.1 Markov Random Field

Markov Random Field is an undirected probability graph which follows local Markov property among its nodes connected with edges within a clique. Specifically, in a Markov Random Field, if there is an absence of edge between two nodes, which indicates that they are conditionally independent given all other random variables in the graph. By contrast, an edge means a dependency relation between two nodes. A Markov random field Y, the probability distribution $p(Y)$ can be written as a Gibbs distribution according to the Hammersley-Clifford theorem (Hammersley and Clifford, 1971) as follows:

$$p(Y) = \frac{1}{Z}\prod\nolimits_{c\in C}\varphi_c\,(y_c) \tag{2.1}$$

where C is the clique sets in the graph and φ_c is the potential function of a clique c. Z is a normalization term summing up the product of the potential function over the collection of cliques C.

MRF provides a general framework to design potential function to encode context information for classification problems, which is widely applied in both image and LiDAR point cloud classification. A MRF based spectral–spatial hyperspectral image classification approach (Sun et al., 2015) was introduced to apply a MRF based prior as a hidden field for spatial smoothness. MRF prior was served as a spatial smoothness term to generate a smoother classification result. Schindler (2012) made a comparison between several smoothing methods for land-cover classification. It reported that MRF based smoothing model can significantly improve image classification quality but also caused over-smoothing problem. Another similar algorithm (Grinias et al., 2016) improved this spatial MRF smoothness by incorporating a multilevel "fuzzy no-border/border" map as weight coefficients in a MRF-based multilevel logistic model to overcome potential over-smoothing problem caused in previous proposed MRF model. Cao et al.(2018) also combined MRF with Convolutional Neural Network (CNN) to classify hyperspectral image. In the paper, MRF served as a smoothing post-processing to refine the classification result of CNN model. Besides smoothing based MRF model, a MRF based unsupervised segmentation approach Grinias et al., (2016) was proposed to combine all region appearance clues for segmenting roads and building and then a typical random forest

supervised classification coupled with segments to detect roads and buildings in very high-resolution satellite image.

Besides these applications on image classification, MRF model was also introduced for point cloud classification in various ways. Lafarge and Mallet (2012) applied MRF to classify airborne LiDAR point clouds into building, vegetation and ground, where the Potts model was used for the pairwise potentials, maximizing local smoothness. Another smoothness MRF for airborne LiDAR point cloud classification (Zhu et al., 2017) was coupled with high-order semantic constraints. These methods only considered smoothness and they also met over-smooth problem. A variant of MRF, Associative Markov network (AMN) has been introduced (Triebel et al., 2006) for applying pairwise potential as a weighted Potts model to reduce the risk of over-smoothness. Correspondingly, non-associative Markov Network was also introduced (Shapovalov and Velizhev, 2011) for 3D point cloud classification. The advantage of the model is that it can impose more flexible constraints in contrast to AMN model simply for smoothness purpose. Also, a higher-order non-AMN model (Najafi et al., 2014) was introduced to encode a variety of high-order potential functions to model complex smoothness and geometric constraints.

As seen from the previous review, the most popular application of MRF based classification model is by encoding a smoothness context to assume neighboring primitives that belongs to the same semantic objects. However, these MRF based smoothness terms cannot capture the conditional dependencies between labels and features, which causes the over-smoothing problem. Fortunately, Conditional Random Field (Lafferty et al., 2001)

gives a more flexible solution for allowing the potential function conditioned on class label as well as observations data (Kumar and Hebert, 2006).

2.2.2.2 Conditional Random Field

A Conditional Random Field (CRF) is an undirected probabilistic graph G which is composed of a set of nodes n and edges e which connect nodes. The nodes can be image pixel, individual point or in my case line primitive. Edges which connect nodes model the statistical relation between the connected nodes. As a discriminative classifier, CRF directly computes posterior probability $p(Y|X)$ between hidden variable $Y$and observed data X into a multiplication of potential factor as follows:

$$p(Y|X) = \frac{1}{Z(X)} \prod_{i \in S} \varphi_i(y_i, X) \prod_{h \in C} \varphi_c(y_c, X) \tag{2.2}$$

Where $\varphi_i(y_i, X)$ represents unary potential, which encodes the importance of each node with observed data X. The clique potential $\varphi_h(y_h, X)$ models the relation between hidden variable y_i and nodes x_i within a clique c. Dependent on the nodes involved in each clique, the clique potential $\varphi_h(y_h, X)$ can be pairwise or high-order. Due to the exponential computational increase with the clique order, high-order potential is not widely applied even though some specific formulated high-order clique potentials (Kohli and Torr, 2009) can be computational trackable. S and C are the number of nodes and cliques within the graph G respectively. $Z(X)$ is a normalization constant to normalize the multiplication into a probability output ranging from 0 to 1 and is not related to any class label. In the context of classification problem, hidden variable Y is a label vector which should be determined for each node x_i and observed data X is usually a feature vector to represent each node in

feature space. The classification goal is assigning the most probable class label y_i to each node x_i which can maximize the posterior probability.

Compared with those MRF based classification methods, CRF was more heavily applied for classification tasks to encode more complex context information as well. He et al. (2004) introduced a multi-scale CRF to label the image which firstly integrated local, regional and global features in different spatial scales for image semantic labeling. A hierarchical Conditional Random Field (Yang and Förstner, 2011) was proposed to model contextual information via both spatial and hierarchical structures. The image was first hierarchically segmented into 3 scales and a 3-layer Conditional Random Field was established, edges can represent the neighborhoods at one scale or at different scales. Through this way, the CRF model can express the influence on different region scales which is not modeled by one-local conditional random field. Another two-layer Conditional Random Field (Albert et al., 2014) was introduced for the classification of land cover and land use. One layer in the model is land cover layer while the other layer is land use layer. The node in land cover layer represents super-voxel from aerial image while the node in land use layer represents geometry of objects from geospatial database. Inter-layer edge was constructed if the super voxel has the overlapping with objects in the database. Intra-layer edge is the same as standard CRF. To consider more sophisticated relation among multiple objects, a high-order smoothness CRF framework (Kohli and Torr, 2009) was proposed to extend conventional Potts model into a more general P^N Potts model for high-order label consistency in image semantic segmentation. Another high-order CRF (Wegner et al., 2013) was later introduced to form a high-order clique for road extraction.

They connected super pixels together along straight lines or 3-junctions and used a truncated linear function to penalize background pixels in road-dominant cliques, which is a special case for the P^N Potts Model. These methods usually can only encode contextual information within some certain ranges. Koltun (2011) represented a fully-connected CRF model to expand traditional grid-like CRF model to connect every node pair and a Gaussian pairwise potential was used to model the similarity between connected nodes. A mean field inference algorithm was proposed to solve the large-scale inference problem. The biggest contribution of this method is extending the range of contextual information to the full scene. Recently, CRF successfully combined multiple types of context information (semantic, spatial and object scale) together into a unified framework. Yao et al. (2012) proposed a unified CRF framework to impose several context information, which included object-scene compatibilities, object and semantic compatibilities, smoothness and relative spatial location among objects. An interesting trend is that CRF and CNN can be tightly coupled for image semantic task (Arnab et al., 2016;Arnab et al., 2018;Chen et al., 2018;Zheng et al., 2015), which becomes a standard technology for image semantic task nowadays.

Several different CRF frameworks were also proposed for 3D point cloud or RGBD image classification. Lim et Suter (Lim and Suter, 2007;Lim and Suter, 2008;Lim and Suter, 2009) used both standard and multi-scale CRF to incorporate both spatial relationship and neighbor smoothness to classify terrestrial LiDAR point cloud. The node in the graph model is super-voxel to reduce the computational difficulty. Within each super-voxel, there are also edges related to the points belong to the same super-voxel. This

algorithm enlarged the neighbor scope to capture more contextual information. However, the down sampling operation deteriorated the distinction of minor objects and the algorithm was only tested on small and simple scene. Koppula et al. (2011) demonstrated a discriminant function to consider visual and geometry unary information and geometry spatial context (vertical and horizontal displacement) as a pairwise term to classify point cloud obtained from RGBD data. Niemeyer et al. (2012, 2014) introduced a point-based CRF framework for semantic labeling to classifying complex urban scene from airborne laser scanning data. Random Forest was used to train both association and interaction potentials in the CRF model. The same strategy was also used by (Vosselman et al., 2017) on a segment level for the same task. Niemeyera et al. (2016) continued their work on proposing a two-layer CRF to preserve the small object in a point-based layer and formulate long range relationships in segment-based layer. Two different CRF models on point and segments were established individually and interacted with each other through a high-order smooth term with an extra long-range energy cost in point-layer and segments formulation in segment-layer. This method considered context information at two scales individually and the relation between two layers was just formulated as a high-order homogeneity constraint. Similar hybrid CRF (Lu et al., 2009) was introduced to extract DTM from DSM from airborne LiDAR data, in which both height estimation and ground classification are within the proposed framework. The approach calculated some geometric feature to predict point cloud as ground/non-ground and then a DTM will be extracted through a random field based on point classification. However, this hybrid representation did not encode contextual information with different classification primitives. Spatial

layout is also incorporated into a multi-range asymmetric CRF for terrestrial LiDAR data classification (Luo and Sohn, 2014). In the short range, the algorithm applied Potts model for local smoothness. Layout information is obtained from a prior table which represents layout information obtained from training data. A multivariate Gaussian distribution is assumed to represent long-range pairwise terms.

2.3 Chapter Summary

Railway electrification system classification is still not well researched albeit that there are quite few researches on railway object and scene recognition. But these methods still commonly meet several problems unsolved. Firstly, they usually set very complex ad-hoc rules to detect and classify objects, which is not general for different scenes. Another limitation is that these methods all sequentially processed the data and both detection and classification results were influenced by previous operation result. Finally, these methods do not have a detailed definition of objects in the railway electrification scene and they also did not consider the spatial regularities in the scene. From the review of context-based classification methods, it is noticeable that probabilistic graph model, especially Conditional Random Field, provides quite solid framework which can incorporate multiple types of context information to improve classification result. Therefore, CRF based model shows the potential to solve railway electrification scene classification problem. But the context information imposed in CRF model also have the problem of ambiguities, inaccuracies and incompleteness.

Chapter 3
Multi-Range Conditional Random Field

This chapter proposes a multi-range conditional random filed (MrCRF) framework to classify railway electrification system objects in mobile laser scanning data. Railway tracks, which were firstly extracted using a method proposed by Jwa and Sohn (2015), with MLS data are used as input data. The railway track contains the coordinates of several railway points and every two adjacent railway points will compose a railway line segment. Therefore, every railway track will consist of several railway line segments. Line-based classification is applied where linear segments are used as unit entities. After generating the voxel structure from MLS data, linear segments for each voxel are extracted by applying the RANSAC algorithm where points are considered as consensus if the distance between the point and a candidate line is smaller than a certain user-defined distance. Note that multiple linear segments can be extracted in each voxel. After applying the SVM classifier, the proposed MrCRF, which considers short range and middle-range horizontal and long-range vertical relations, is applied to classify linear segments. In MrCRF, two different graphs, which represent short-range and middle-range relations, respectively, are generated to define adjacent relationships. Based on the generated graphs, integrated CRF is conducted to refine the SVM results. First, a brief description about model design in Section 3.1 will be introduced, then the design of two graphs used in the model will be shown in Section 3.2. The unary term will be demonstrated in Section 3.3 while three pairwise terms in the model will be explained in Section 3.4. The training and the inference

of the proposed model is illustrated in Section 3.5. the classification model is evaluated on a high-speed dataset in South Korea in Section 3.6. Some discussion and comparison of my methods with several different classification models will be presented in Section 3.7. A summary is provided in Section 3.8.

3.1 Graphical Model Design

CRF is used to encode known relations between observations and to construct consistent interpretations. It usually consists of a unary term, which represents the importance of each node, and a pairwise term, which represents the contextual information with a graph. In this thesis, the contextual information is expressed by the local smoothness with a short-range graph and spatial layouts in both vertical and horizontal directions with a middle-range graph. Thus, the proposed multi-range CRF consists of a combination of a unary term and three different pairwise terms. The unary term is designed to encode the likelihood of each node to be assigned with each label given the node features. Three pairwise terms formulate the local smoothness, vertical spatial layout and horizontal spatial layout through edges in the graphs given the observation of edge feature X.

In a pairwise CRF, the posterior probability $p(Y|X)$ of the label vector Y based on the observed data X is expressed as follows:

$$p(Y|X) = \frac{1}{Z(X)} \prod_{i \in S} \varphi_i(y_i, X) \prod_{i \in S} \prod_{j \epsilon N_i} \varphi_{ij}(y_i, y_j, X) \tag{3.1}$$

where $\varphi_i(y_i, x)$ is the unary potential and $\varphi_{ij}(y_i, y_j, x)$ represents the pairwise potential. $Z(x)$ is the normalization constant (partition function) to ensure the probabilities p sum up

to 1. S is a set of nodes in the graph, and N_i represents the neighbors of node i connected via edges in the graph. Due to the monotonic property of the logarithm, Eq. (3.1) can also be expressed as follows:

$$p(Y|X) = \frac{1}{Z(X)} exp\left(\lambda \sum_{i \in S} \varphi_i(y_i, X) + \alpha \sum_{i \in S} \sum_{j \epsilon N_i} \varphi_{ij}(y_i, y_j, X) \right) \quad (3.2)$$

where λ and α are the weight parameters to balance the unary term and pairwise term, respectively. In the model, the pairwise term in Eq. (3.2) is expanded to three pairwise terms as follows:

$$p(Y|X) = \frac{1}{Z(X)} exp\left[\lambda \sum_{i \in S} \varphi_i(y_i, X) + \alpha \sum_{i \in S} \sum_{j \epsilon N_i^S} \varphi_{ij}^S(y_i, y_j, X) + \beta \sum_{i \in S} \sum_{j \epsilon N_i^L} \varphi_{ij}^{MV}(y_i, y_j, X) + \gamma \sum_{i \in S} \sum_{j \epsilon N_i^L} \varphi_{ij}^{MH}(y_i, y_j, X) \right] \quad (3.3)$$

where $\varphi_i(y_i, X)$, $\varphi_{ij}^S(y_i, y_j, X)$, $\varphi_{ij}^{MV}(y_i, y_j, X)$ and $\varphi_{ij}^{MH}(y_i, y_j, X)$ represent the unary potential, short-range pairwise potential, vertical middle-range pairwise potential and horizontal middle-range pairwise potential, respectively. λ, α, β and γ are the weight parameters for the four sub-terms, respectively.

3.2 Definition of the Graph

In a CRF model, dependent relations between nodes are defined by an adjacent graph. In image space, the adjacent relation is normally determined by adjacent pixels using the standard four-connected neighborhood (Gould et al., 2008;Shotton et al., 2009) or using the eight-connected neighborhood (He et al., 2004). However, in laser scanning points that are

irregularly distributed, the definition of adjacent relations is not straightforward. In previous studies using point data, neighborhood relations are defined by Delaunay Triangulation (DT) (Douillard et al., 2008), k nearest neighbors (Niemeyer et al., 2014) and super-voxels (Lim and Suter, 2009). Chao and Sohn (2014) also projected 3D points into a 2D plane to model both horizontal and vertical relations within the plane. In this thesis, two different neighboring systems are defined for establishing the short-range and long-range relation. A sphere with a user-defined radius (1.5 m in this thesis to connect adjacent lines within the adjacent voxel) is used to define the short-range relation (Figure 3.1a), while a cylinder with a hole is used for middle-range relation (Figure 3.1b). The height and radius of the cylinder and the radius of the hole are heuristically chosen as 5 m, 1.5 m and 1.5 m, respectively, based on a priori knowledge of the railway electrification system design used for the current site. those parameters are set for fully constructing edges with the lines on one side of the railway to discover all layout information, while separating them from

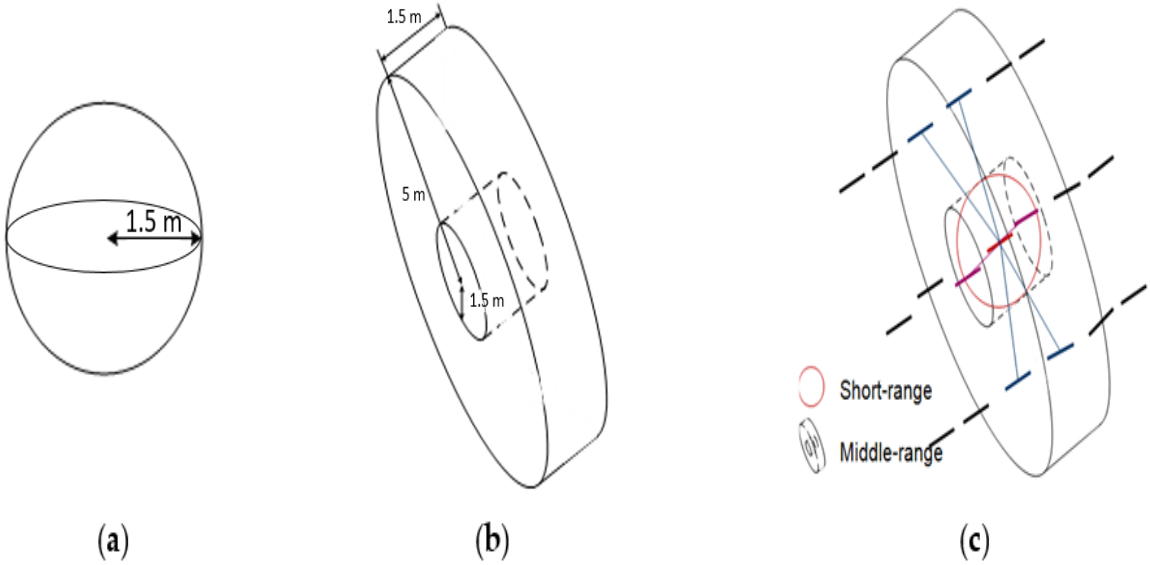


Figure 3.1 Neighboring systems: (a) for short-range graph; (b) for middle-range graph; and (c) combined neighboring systems.

Once two types of neighboring systems are defined, short-range and middle-range graphs are generated. In graphs $G = (V, E)$, a node $v \in V$ represents a line segment extracted from MLS data, and an edge $e \in E$ is constructed if a line is found in each neighboring system. As mentioned above, three different graphs, the short-range graph, middle-range vertical graph and middle-range horizontal graph, are generated for the proposed CRF model $\mathrm{G} = \{\mathrm{G_S}, \mathrm{G}_{MV,}\mathrm{G}_{MH,}\}$.In the short-range graph $G_S = (V, E_S)$, a line is considered as a neighbor node if the center point of the line is found in the sphere generated from the other line. Note that an edge is excluded if the angle difference between two lines is significantly different from a user-defined threshold (manually set in this thesis). This exclusion is applied because that two lines, which have a larger angle difference, are likely to belong to different classes. Middle-range vertical graph $G_{LV} = (V, E_{MV})$ is generated by applying the cylinder with a hole so that short-range relations are excluded. The middle-range horizontal graph $G_{LH} = (V, E_{MH})$ is the same as the middle -range vertical graph, but with different edge features. Furthermore, a line whose center is below the corresponding rail vector is excluded from the middle range graph. It can largely reduce the number of middle-range edges, so that the inference speed can be significantly accelerated. In both graphs, multiple edges for one line can be generated. The implementation of two neighboring systems were conducted together with Dr. Jung (Jung et al., 2016).

3.3 Unary Term

The unary term in Eq. (3.3) corresponds to the log posterior probability of any label y_i given observation x_i. Because the unary term only considers node features, the posterior

probability of any local classifiers can be used. SVM is a very typical discriminative classifier that maps the data into a high-dimensional feature space and finds a hyperplane that separates the feature space with the maximum margin (Lodha et al., 2006). The SVM classifier shows success in multiple class classification problems. Thus, the SVM classifier is firstly apply to classify the railway scene. In the SVM setting, a common Radius Basis Function (RBF) kernel is selected and the gamma value in RBF kernel is 0.1. There is also an equal weight for all classes for SVM training. The six-dimensional features is used to represent the property of a line segment as follows:

- Point density: the density of points that support a line segment.
- Residuals: the standard deviation calculated from the line segment and its supporting points.
- Verticality: the vertical angle of the line.
- Horizontal angle: the angle between the line segment and its corresponding rail vector in the XY plane.
- Height: the height difference between a line segment and its corresponding railway line segments.
- Distance: the horizontal distance between a line segment and its corresponding railway line segments.

The SVM log posterior probability results are used as the unary term in the CRF model as follows:

$$\varphi_i(y_i, X) = log\big(p(y_i|x_i)\big) \tag{3.4}$$

3.4 Pairwise Terms

There are three pairwise terms in the model, which express the local smoothness in a short range and spatial regularities in both vertical and horizontal direction in a middle range.

3.4.1 Short-Range Binary Term

The second term in Eq. (3.3) represents the short-range pairwise term, which is designed to enforce local smoothness. Local smoothness is a universal assumption that things in the physical world are spatially smooth (Schindler, 2012), which means that the neighboring line segments are more likely to have the same label. This term is designed by the Potts model favoring neighboring entities i and j to have the same label and penalizing the configuration of different labels. The Potts model is simple, but quite effective for many smoothness applications. In my research, the short-range pairwise potential $\varphi_{ij}^{S}(y_i, y_j, X)$ can be expressed as follows:

$$\varphi_{ij}^{S}(y_i = l, y_j = k, X) = \begin{cases} 1, l = k \\ 0, l \neq k \end{cases} \tag{3.5}$$

3.4.2 Middle-Range Binary Term

The scene layout illustrates the relative location of objects in the scene. For the railway scene, obvious regularities in terms of the relative location are evident in both the vertical and horizontal directions. For instance, the suspension insulator is always higher than the transmission wires, while the catenary wire is always closer to the rail tracks compared to the current return wire. This layout information can be automatically learned from the training data. Co-occurrence statistics recently have attracted more attention in

representing spatial layout. This can reflect relative locations for all objects in a map, and then, the map intensity represents how possible it is that two objects co-occur in a certain pattern. For the middle-range terms, the co-occurrence statistic is adopted to define "above-below" and "near-far" relations in both the vertical and horizontal directions, which are described in next two subsections, respectively.

3.4.2.1 CRF Based on Vertical Layout Compatibility

To embed the vertical layout compatibility in the CRF model, the "above-below" relationship is modeled for middle-range neighbors. The Bayes rule is used to calculate the posterior probability as follows:

$$p(y_{above} = l, y_{below} = k | e_{ij}) = \frac{p(e_{ij} | y_{above} = l, y_{below} = k) p(y_{above} = l, y_{below} = k)}{\sum_{y_{above} \in L, y_{below} \in L} p(e_{ij} | y_{above} = l, y_{below} = k) p(y_{above} = l, y_{below} = k)} \quad (3.6)$$

where (y_i, y_j) is a pair of lines consisting of the edge e_{ij} in graph G_{LV}. y_{above} indicates the node above the other in the edge e_{ij}, while y_{below} indicates the node below the other.

$p(y_{above} = l, y_{below} = k)$ is the prior probability that class type l is above class type k. The prior probability is represented by the co-occurrence rate, which is statistically obtained from the training data. In this thesis, the co-occurrence rate is formulated from a look-up table, as shown in Figure 3.2(a). The likelihood function in Eq. (3.6) is the probability distribution function of edge e_{ij} given a configuration that class l is above class k, which quantitatively measures how likely class l can be found above class k. Here, it uses three-dimensional feature vector u_{ij} to represent edge e_{ij}. The feature vector

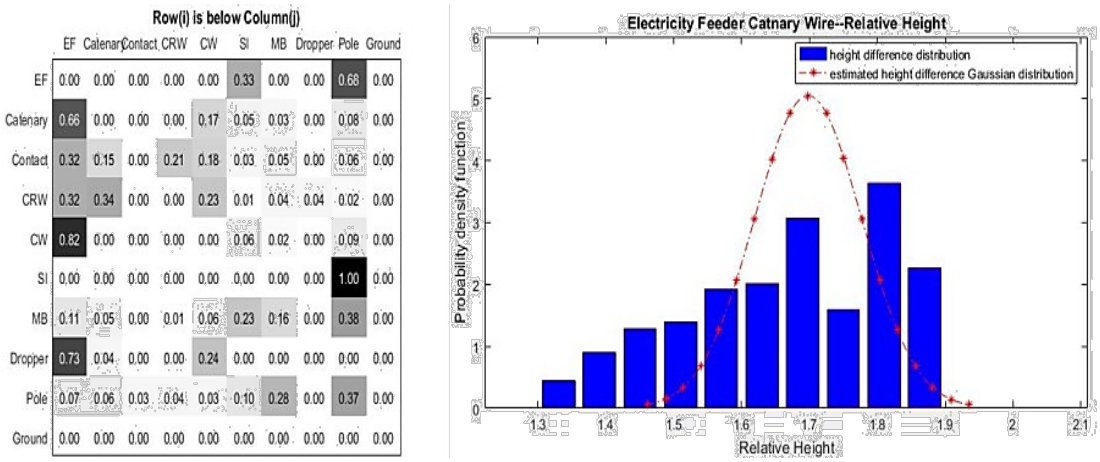
consists of the height difference, horizontal angle difference and verticality difference between two line segments. It is assumed that the edge feature distribution follows a multivariate Gaussian distribution as follows:

$$p(u_{ij}|y_{above} = l, y_{below} = k) = \frac{exp\left(-\frac{1}{2}(u_{ij} - \mu_{l,k})^T {\Sigma_{l,k}}^{-1}(u_{ij} - \mu_{l,k})\right)}{2\pi\sqrt{|\Sigma_{l,k}|}} \quad (3.7)$$

where $\mu_{l,k}$ and $\Sigma_{l,k}$ are the mean vector and covariance matrix, respectively. In the study, the parameters are trained from the training data through the Maximum Likelihood (ML) algorithm. Figure 3.2(b) shows the estimated probability distribution of the height difference between the electricity feeder and catenary wires. In the figure, the estimated probability distribution from the training data fits the test data feature distribution well. This indicates that the multivariate Gaussian distribution is applicable to the railway scene. Then, the vertical long-range pairwise term can be expressed as follows:

$$\varphi_{ij}^{LV}(y_i, y_j, X) = log\left(p(y_{above} = l, y_{below} = k|e_{ij})\right) \quad (3.8)$$

With ten classes (introduced in Section 3.6), 100 types of pairwise potentials are learned from the training data, generating different multivariate Gaussian distributions. The designed long-range potentials are not asymmetric because both the prior and likelihood are asymmetric, which makes potential $\varphi_{ij}^{LV}(y_i, y_j, X) \neq \varphi_{ji}^{LV}(y_j, y_i, X)$. This configuration will encourage the right vertical layout and penalize the opposite vertical layout.



(a) **(b)**

Figure 3.2 Prior and likelihood estimation of the vertical middle-range pairwise potential: (**a**) look-up table where row *i* is below column *j*; and (**b**) probability density distribution of the height difference (red curve: estimated Gaussian distribution learned from the training data).

3.4.2.2 CRF Based on Horizontal Layout Compatibility

Similar to the vertical middle-range pairwise term, a "near-far" relationship in this middle-range horizontal pairwise term is modeled. The same middle-range graph is used, but a different feature property is applied to represent the near-far relationship. Three-dimensional feature vector δ_{ij} which consists of the horizontal angle difference, vertical angle difference and horizontal distance difference, is formulated between two line segments. The Bayes rule is also used to calculate the posterior probability as follows:

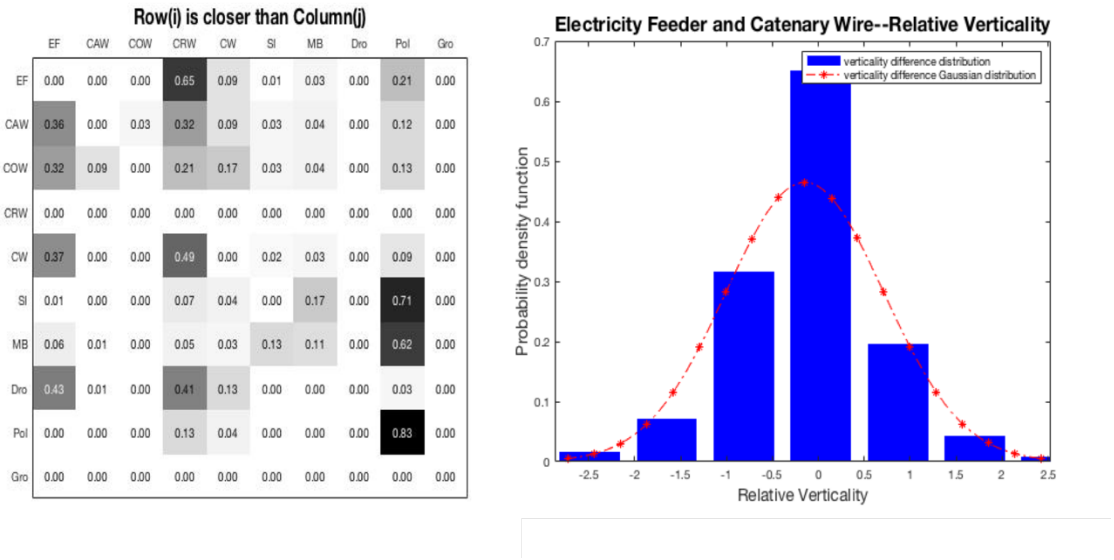
$$p(y_{near} = l, y_{far} = k|\delta_{ij}) = \frac{p(\delta_{ij}|y_{near} = l, y_{far} = k)p(y_{near} = l, y_{far} = k)}{\sum_{y_{near}\in L, y_{far}\in L} p(\delta_{ij}|y_{near} = l, y_{far} = k)p(y_{near} = l, y_{far} = k)} \quad (3.9)$$

where (y_i, y_j) is a pair of lines consisting of the edge e_{ij} in graph G_{LH}. y_{near} and y_{far} represent the horizontal relations between nodes.

In Eq. (3.9), $p(y_{near} = l, y_{far} = k)$ is the prior probability that class type l is closer to the railway than class type k. The prior probability is expressed by a look-up table for the near-far relation as shown in Figure 3.3(a). Similarly, distributions for the edge features in horizontal relation distributions are formulated as a multivariate Gaussian distribution as follows:

$$p(u_{ij}|y_{near} = l, y_{far} = k) = \frac{exp\left(-\frac{1}{2}(u_{ij} - \mu_{l,k})^T {\Sigma_{l,k}}^{-1}(u_{ij} - \mu_{l,k})\right)}{2\pi\sqrt{|\Sigma_{l,k}|}} \quad (3.10)$$

where $\mu_{l,k}$ is the mean vector and $\Sigma_{l,k}$ represents the covariance matrix. The estimated probability distribution of the horizontal angle difference is shown for the electricity feeder and catenary wire in Figure 3.3(b). Like the vertical middle-range pairwise potential, the horizontal middle -range pairwise potential is also asymmetric, and it encourages the right horizontal layout configuration.



(a) **(b)**

Figure 3.3 Prior and likelihood estimation of the horizontal middle-range pairwise potential: (a) look-up table where row *i* is closer to the railway than column *j*; and (b) probability density distribution of the verticality difference (red curve: estimated Gaussian distribution learned from training data).

3.5 Training and Inference

As mentioned above, there are two types of parameters to be trained in the integrated CRF model. The first type is the parameters in the middle-range term, while the other is the weights between different sub-terms (λ, α, β and γ in Eq. (3.3)). The parameters in the middle-range term include the prior term and the parameters (μ, $\sum$) in multivariate Gaussian distributions for estimating the likelihood function. Generally, parameters in CRF can be learned by maximizing the posterior probabilities of true labels given the training data (Luo and Sohn, 2014). The partial derivative needs to be calculated to find the best parameters that maximizes the posterior probability of true labels. However, because the partial derivative is a nonlinear function with respect to each term, it is challenging to

directly calculate the partial derivative, which makes it very difficult to train all parameters at once. Some previous research works (Gould et al., 2008;Kohli and Torr, 2009;Shotton et al., 2009) simplified the training through assigning the same weight value to the unary term and pairwise term. However, this simplification cannot reflect the relative importance of each term in the final decision-making. Alternatively, a two-step training strategy is set to train all parameters. Firstly, the parameters in the middle-range term are trained individually. Second, the relative weights for sub-terms are subsequently learned through the Stochastic Gradient Decent (SGD) algorithm. The inference is applied when all parameters are trained. Section 3.5.1 introduces how these parameters in the CRF model are trained, while Section 3.5.2 demonstrates the inference operation to the final decision-making.

3.5.1 Parameter Estimation

For the unary term in the integrated CRF model, it directly uses the SVM confidence value as the unary term, which is learned from the same training data as the CRF model. Pairwise potential is implemented as the Potts model that each edge potential is the exponent of an identity matrix. Thus, no parameter needs to be trained. In two middle-range pairwise terms, the prior is obtained from relative location probability maps (look-up tables, L_v and L_h), which statistically calculate the co-occurrence rate over all class pairs. If a line primitive i with class label c is higher than a line primitive j with class label c', the corresponding element $L_v(c, c')$ in the look-up table gets a vote. Once all vertical relations are recorded in the look-up table, elements in the look-up table are normalized, satisfying

$\sum_{c=1}^{K} L_v(c, c') = 1$. In a similar way, L_h is calculated by considering the near-far relation. The multivariate Gaussian distribution parameters are estimated by the traditional maximum likelihood algorithm (Gauvain and Lee, 1994), which calculates the mean vector and covariance matrix from training data.

Once all terms are estimated, the weights for sub-terms using the SGD algorithm (Vishwanathan et al., 2006) are learnt. SGD is a stochastic approximation of gradient descent optimization technology to find the global minimum of the objective function. Different from traditional gradient descent (GD), which uses whole training data to calculate the partial derivate, SGD randomly picks a subset of training samples and then updates the parameters according to the gradient calculated by the subset of training data. Although it is not the exact gradient that moves to the optimal solution directly, the parameter updating process using the subset of training data can be much simplified. In the CRF model proposed in this thesis, the marginal probability of training data is required to compute the partial derivative, so the inference process should be applied at every iteration to update the partial derivative.

The objective function to be maximized is the logarithm form of the estimated posterior probability as follows:

$$\mathrm{LP}(\theta) = \lambda \sum_{i \in V} \varphi_i(X, y_i) + \alpha \sum_{i \in V} \sum_{j \in N_{SR}} \varphi_{ij}^{S}(X, y_i, y_j) + \beta \sum_{i \in V} \sum_{j \in N_{LR}} \varphi_{ij}^{LV}(X, y_i, y_j) + \gamma \sum_{i \in V} \sum_{j \in N_{LR}} \varphi_{ij}^{LH}(X, y_i, y_j) - In[Z(X)] \tag{3.11}$$

In Eq. (3.11), λ is set to one. This is because weights (λ, α, β and γ) can be scaled up or down, which does not affect the result of inference. To update the weight parameters (α, β and γ), the partial derivative regarding each weight term is calculated as follows:

$$g_{\alpha}^{t} = \frac{\partial LP(\theta^t)}{\partial \alpha} = \sum_{i \in V} \sum_{j \in N_{SR}} \varphi_{ij}^{S}(X, y_i, y_j) - \frac{1}{Z(X)} \times \frac{\partial Z(X)}{\partial \alpha} \tag{3.12}$$

$$\frac{1}{Z(X)} \times \frac{\partial Z(X)}{\partial \alpha} = \sum_{i \in V} \sum_{j \in N_{SR}} \varphi_{ij}^{S}(X, y_i, y_j) \times M\left(P\left(Y|X, y_i, y_j, \theta^t\right)\right) \tag{3.13}$$

where g_{α}^{t} is the partial derivative of α after t updates and $M\left(P\left(Y|X, y_i, y_j, \theta^t\right)\right)$ is the edge margin of the short-range pairwise term, which is obtained from the inference operation given the current weight parameters. The weight parameter of short-range updates uses the following equation:

$$\alpha^{t+1} = \alpha^{t} - \varepsilon \times g_{\alpha}^{t} \tag{3.14}$$

In Eq. (3.14), a learning rate ε needs to be properly determined to make the function converge stably and control the converging speed. However, it is not an easy task to determine a proper learning rate. The common strategy is to set a relatively larger learning rate at the beginning to accelerate convergence and then reduce the learning rate gradually

to ensure a stable convergence (Bowling and Veloso, 2002). A similar strategy is applied here to determine the learning rate.

3.5.2 Inference

Inference is the operation to find the best possible label configuration in the graphical model given the observation X. Usually the inference operation can be divided into exact inference and approximate inference. Exact inference is applicable to certain special graphs, such as chain-structure or tree-structure graphs. However, the exact inference cannot be applied in this case where a graph has loops. Thus, the Loopy Belief Propagation (LBP) algorithm, which was reported as a satisfactory solution for the inference of graphs with loops, is applied for approximate inference. The final label is decided by maximizing the node belief from the inference results.

3.6 Experiment Results

The proposed MrCRF method was tested on MLS data taken at the Honam high-speed railway in South Korea. The MLS data were acquired in 2014 using the Trimble MX8 system, which was mounted on an inspection train with a speed of 50 km/h to 70 km/h. Trimble MX8 system is a premium mobile laser scanning system which can produce accurate dense point cloud with dual scanners producing more than one million points per second. The advantage of Trimble MX8 system makes it suitable for our task because it can provide higher resolution of railway electrification scene objects for the later classification and modeling task. The average density varies on the position of the laser

scanner ranging from 100 points/m^2 to 800 points/m^2. Figure 3.4 and Table 3.1 show the Trimble MX8 system used for this study and its technical specifications.

Figure 3.4 Trimble MX8 mounted on a train.

Table 3.1 Specifications of the Trimble MX8 system.

Parameter	Values
Accuracy	10 mm
Precision	5 mm
Maximum effective measurement rate	600,000 points/second
Line scan speed	Up to 200 lines/second
Echo signal intensity	High resolution 16-bit intensity
Range	Up to 500 m

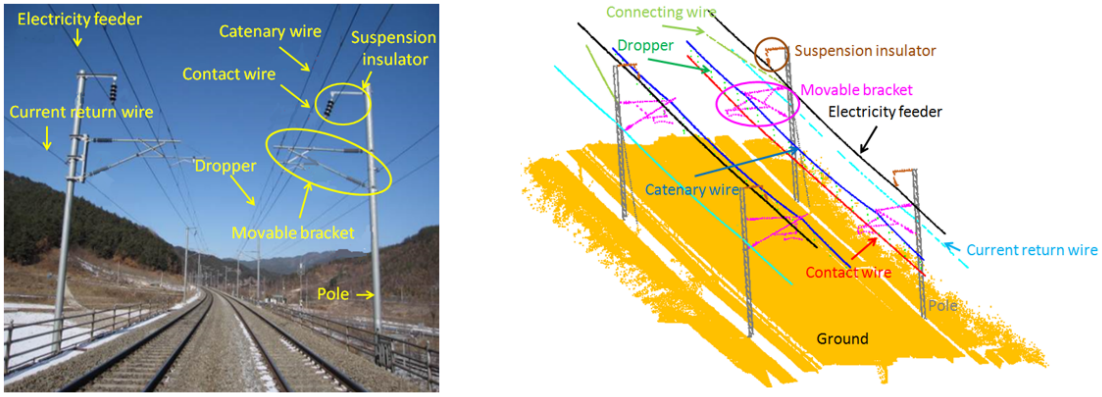
The length of the dataset selected for this thesis is approximately 1 km. There are two pairs of rail tracks and 24 poles at regular intervals. The dataset was divided into six sub-regions for cross-validation purposes, each of which has four poles (two pole-pairs), and its length is approximately 160 m. Under this setting, to evaluate the result of each sub-region, the rest five sub-regions will be used for training. Therefore, usually approximate 800m railway dataset will be used for training and each training set has more than 20,000 line primitives. Each sub-region has a slightly different configuration of key objects comprising the railway electrification system. This thesis is aimed to recognize 10 different classes of the railway electrification system objects, as shown in Figure 3.5. The targeted

objects play important roles for safely supplying the electricity to the trains. A brief description of the targeted object classes is given below:

- Electricity feeder (EF): a set of electric conductors that originate from a primary distribution center and supply power to one or more secondary distribution centers. The electricity feeders are located at the top of the railway scene with an elevation of 8 m above the ground (vertical configuration), while they are horizontally placed between rail tracks and poles (horizontal configuration).
- Catenary wire (CAW): a wire to keep the geometry of contact wires within defined limits. The catenary wires are at an elevation of approximately 6.5 m above the rails and approximately 1.2 m above the contact wire (vertical configuration). In the horizontal configuration, the wire is located just above the rails.
- Contact wire (COW): a wire that transfers electrical energy to the train and directly contacts the train. In the vertical configuration of the railway scene, this wire is at an elevation of approximately 5.3 m above the rails.
- Current return wire (CRW): a wire observed outside the rails in the horizontal configuration and supported by a pole. The current return wires are located between the catenary wire and the contact wire in the vertical configuration.
- Connecting wire (CNW): a wire connecting the catenary or contact wires to the poles. Thus, the characteristics of the horizontal and vertical configurations of the connecting wires vary relatively more compared to the other targeted objects.
- Suspension insulator (SI): a structure connecting between the electricity feeder and pole, which is observed at the top of the railway scene. Note that the suspension

insulator and a structure attached to the pole are defined as the class "suspension insulator".

- Movable bracket (MB): a movable structure attached to the pole, which supports the wires.
- Dropper (Dro): a vertical wire connecting between catenary wire and contact wire.
- Pole (Pole): a pole is located outside of the rail tracks and track beds.
- Ground (Gro): a ground surface placed underneath the overhead wires and rail tracks.



(a) (b)

Figure 3.5 Electrification system configuration and 10 object classes of the Honam high-speed railway: (a) a photograph; and (b) Mobile Laser Scanning (MLS) data.

The reference data labeled with 10 object classes was produced by a manual classification method provided by commercial software, TerraScan. Figure 3.6 shows the results of the manually-labelled reference data. In Figure 3.6, major overhead wires (i.e., contact and catenary wires) and associated structures (i.e., poles, suspension insulators and brackets) have relatively strong regularities of object layout and appearance. However, some scenes, such as Sub-region 5 and Sub-region 6 in Figure 3.6, show more complex

object configurations where the layout regularity is not directly applicable; the sub-regions contain many merging wires and double contact/catenary wires (not single). Furthermore, the contact wire is often not observed at one side of the railway.

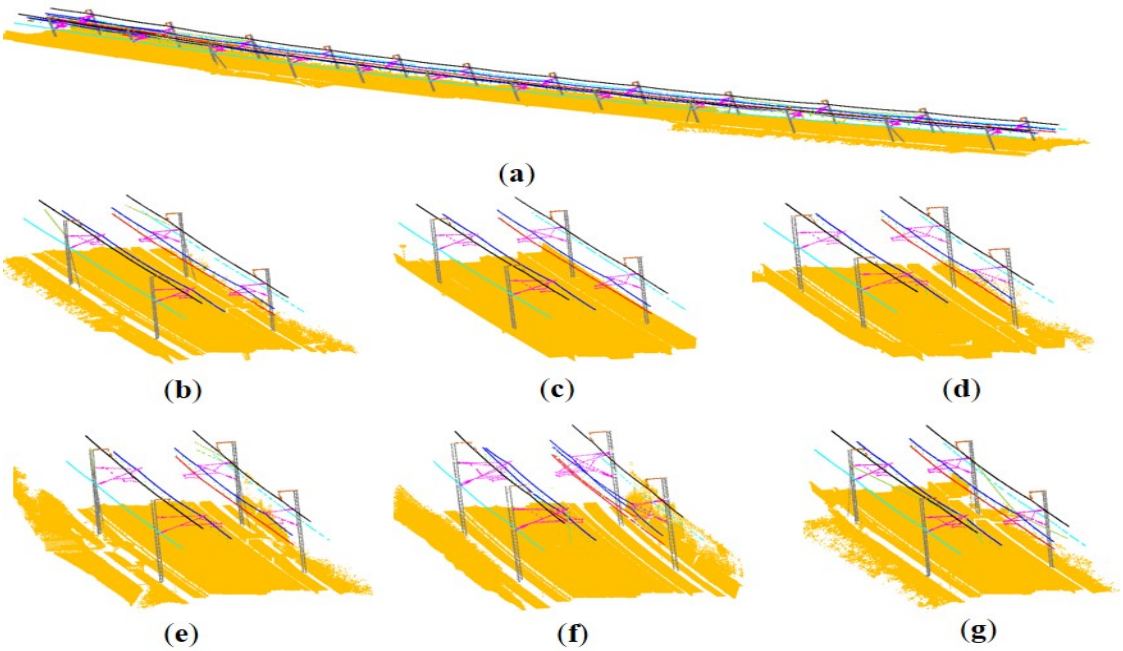


Figure 3.6 Manually-classified reference data: (a) entire region; and (b–g) six sub-regions (different colors represent different classes).

3.6.1 Line Extraction Results

Instead of classifying the entire MLS laser point cloud, our classification process determines object labels to lines where their member points are classified with the same labels. This line-based classification is suitable for classifying railway corridor scenes, as many key objects (i.e., wires and poles) can be well represented with linear primitives. For converting the MLS point clouds (Figure 3.7a) into the line space, the railway corridor scene was represented with voxels with a 1-m bin size (Figure 3.7b), and line segments

were extracted per voxel using a conventional RANSAC algorithm (Figure 3.7c). The inlier threshold (maximum point distance with respect to the corresponding line) used for RANSAC was heuristically determined as 5 cm by considering the positional accuracy of the mobile laser scanner Trimble MX8, the minimum distance between wires and the acceptable tolerance of the noises degrading the performance of the classification. Table 3.2 shows the total number of lines extracted from each sub-region. Note that the RANSAC-based method works iteratively until a termination condition is met, which allows extracting multiple line segments within a voxel. Due to the scene complexity, a relatively larger number of lines were extracted in Sub-region 5 and Sub-region 6.

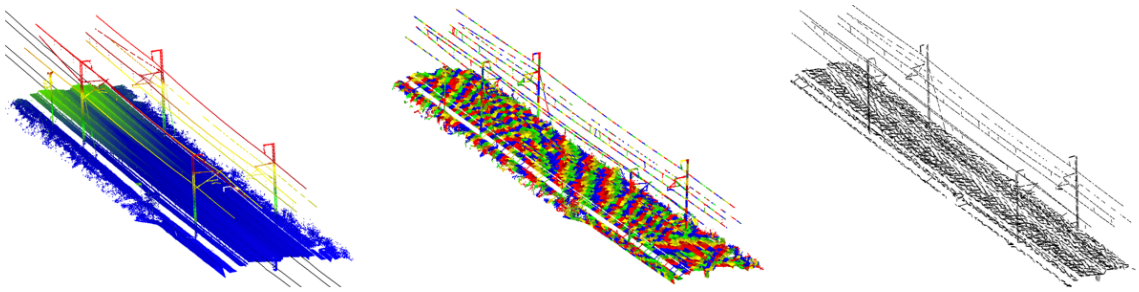


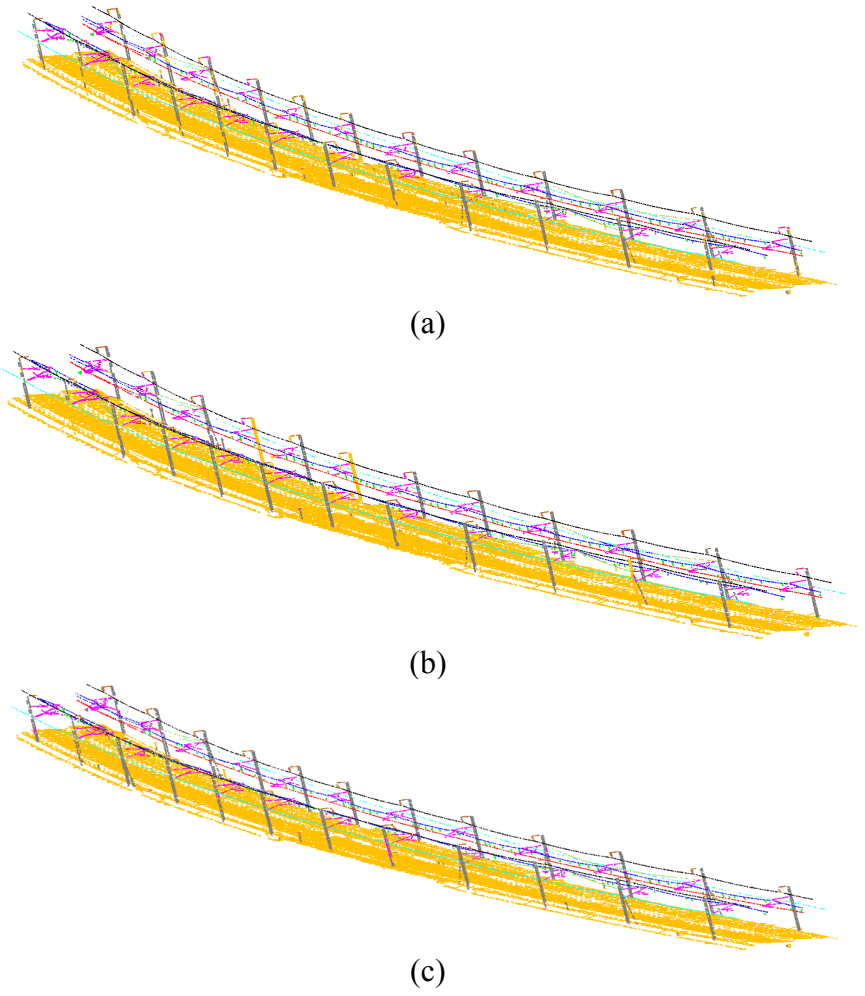
Figure 3.7 Example of voxelization and line extraction: (a) input MLS data and rail vectors, (b) voxelization; and (c) extracted lines.

Table 3.2 Extracted lines for each sub-region.

Sub-Region	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
# of extracted lines	4327	4427	3783	4091	4730	4762

3.6.2 Classification Results

The classification results over the test railway corridor scene were produced by three different classifiers: (1) the local classifier (SVM) without contextual features; (2) the short-range CRF (SrCRF) model with local smoothness; and (3) the multi-range CRF (MrCRF) with local smoothness and layout regularity. The overall classification results are shown in Figure 3.8.



(a)

(b)

(c)

Figure 3.8 Classification results produced over the entire railway test scene: (a) SVM; (b) short-range Conditional Random Field (CRF); and (c) multi-range CRF; electricity feeder (black), catenary wire (blue), contact wire (red), current return wire (sky blue), connecting wire (dark green), suspension insulator (brown), movable bracket (magenta), dropper (green), pole (grey) and ground (orange).

Six-fold cross-validation was applied to evaluate the performance of each classifier to reduce the bias due to training and testing data selection. Under this setting, five sub-regions will be used as training data while the rest one sub-region will be evaluated. Therefore, all six sub-regions can be evaluated finally and the result will be combined together. The posterior probability generated by SVM was used as the input of the unary potential in the short-range and multi-range CRF models. The SVM classifier used several features characterizing the targeted railway electrification system objects, which included density, residuals, verticality, horizontal angle, height and horizontal distance. In CRF models, the line-based graphs were generated with two difference scales, one for the short-range graph with a smaller proximity of associations and the other for the long-range graph with a larger one (Section 3.2). Note that, in the middle-range graph, lines were excluded if the height of a line is below one of its corresponding rail track vectors. This exclusion can reduce the number of middle-range edges and simplify the graph complexity, which can significantly accelerate the inference speed. Table 3.3 shows the number of edges generated in each sub-region. Figure 3.9 (a), (b) show the examples of short-range and middle-range graphs, respectively. In the horizontal CRF model, the horizontal angle difference, vertical angle difference and the difference of horizontal distances between two nodes were used as features, while in the vertical CRF model, the height difference between two nodes was used as the feature. For the multi-range CRF model, multivariate Gaussian parameters in middle-range pairwise terms were estimated by the maximum likelihood algorithm, while the weight parameters for four sub-terms in the CRF model were estimated by the SGD algorithm as described in Section 3.5. To ensure stable convergence,

the learning rate in SGD starts at 0.0001, and it will halve with the increase of the iterations. For the vertical and horizontal middle-range pairwise terms, the learning rate is always half of the short-range term because the gradient is steeper for the long-range term. Under this setting, it can make sure all weights can converge together.

Table 3.3 Edge number of different models in different sub-regions.

Edge Number	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
# of short-range edges	6967	7844	6671	7657	9460	8469
# of long-range edges	7000	7557	6764	7728	11,331	8005

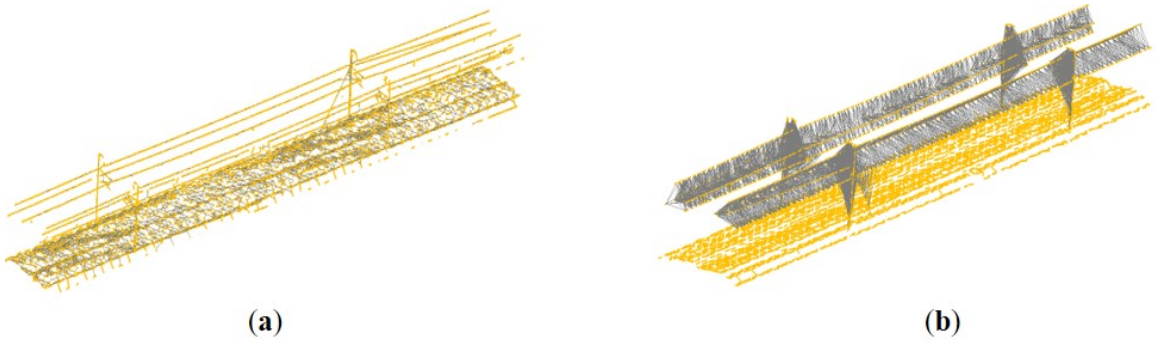


Figure 3.9 Examples of the adjacency graph: (a) short-range graph; and (b) middle-range graph.

The proposed MrCRF classifier was implemented on a desktop computer with 16 GM of memory, an Intel® Core™ i7-4790 CPU with 3.60 GHZ that runs the Windows 10 Professional OS. A total of only 320.45 s was required for classifying the entire datasets. The most of computational gain was obtained by the fact that the proposed algorithm classifies line primitives instead of point clouds. In the training stage, the training of both horizontal and vertical multivariate Gaussian distribution parameters cost 0.94 s, while the training of weight parameters was relatively time consuming, varying from 1.5 h to 3 h in six-fold cross-validation due to inefficient convergence in the LBP algorithm.

To evaluate the performance of the classification methods by comparing results obtained with the reference data, a confusion matrix, also known as an error metric, was used. Each column of the confusion matrix indicates the instances in a classification result, while each row represents the instance in a reference. Based on the confusion matrix, the performance of a classifier for each class is measured with five different scores, overall accuracy (OA), completeness (Comp.), correctness (Corr.), quality (Quality) and kappa index (Kappa) (Rutzinger et al., 2009).

$$\text{Overall Accuracy} = \frac{TP + TN}{TP + TN + FN + FP}$$

$$\text{Correctness} = \frac{TP}{TP + FP} \tag{3.15}$$

$$\text{Completeness} = \frac{TP}{TP + FN}$$

$$Quality = \frac{\text{Corr.} \times \text{Comp.}}{\text{Corr.} + \text{Comp.} - \text{Corr.} \times \text{Comp.}}$$

$$Kappa = \frac{OA - OE}{1 - OE}$$

TP (true positive) represents an instance in the classification result correctly identified by a reference; *FP* (false positive) is an instance incorrectly identified by a reference; *FN* (false negative) represents a missing instance incorrectly identified by a reference. *OA* is overall accuracy while *OE* is hypothetical probability of chance agreement. Overall accuracy shows how many entities are correctly classified. Meanwhile, the correctness, measuring the fraction of the number of true positive prediction of a certain class from the total number of the positive class predicted, can be estimated. The completeness can measure the percentage of the number of true positive prediction of a certain class from the

total number of the class in the reference. The quality is the balance between correctness and completeness, which reflects the classification quality of a certain class. The Kappa Index is an index to illustrate overall classification result with the consideration of the quality of every class.

3.6.2.1 SVM Classification Results

Table 3.4 shows the confusion matrix produced by the SVM classifier. A high rate of the overall accuracy (approximately 98.91%) was achieved by the SVM classifier. As in Table 3.4, per class precision and recall indicate that a group category of "major wire" objects (i.e., electricity feeder, catenary wire, contact wire, current return wire) and ground were significantly well classified with over 99% accuracy. The major wire objects play important roles to transfer the electricity. The shapes of these major objects do not vary much across railway corridor scenes. This strong regularity leads to a small variance in features characterizing the major wire objects used in SVM, which produced highly accurate classification results. However, there are some misclassification errors, which mainly occurred over a group category of "supporting structure" objects (i.e., suspension insulator, movable bracket, dropper and pole) and the other group category of "non-major wire" objects (i.e., dropper and connecting wire). In particular, relatively low completeness for those classes can be observed compared to their corresponding correctness value. The highest classification errors in both correctness (80.22%) and completeness (66.97%) were produced by SVM over suspension insulator objects, which were often confused with movable brackets and poles. Furthermore, it was interesting to observe that movable brackets and poles were mislabeled as various classes, such as suspension insulator,

dropper, pole and ground; where many poles were misclassified to ground in SVM results; while some droppers in the reference were classified to movable bracket or pole in the classified results.

Table 3.4 Confusion matrix for SVM results: electricity feeder (EF), catenary wire (CAW), contact wire (COW), current return wire (CRW), connecting wire (CNW), suspension insulator (SI), movable bracket (MB), dropper (Dro), pole (Pole) and ground (Gro).

		Classified										Comp.
		EF	**CAW**	**COW**	**CRW**	**CNW**	**SI**	**MB**	**Dro**	**Pole**	**Gro**	**(%)**
Reference	EF	1249	0	0	0	2	0	0	0	0	0	99.84
	CAW	0	1365	0	0	4	0	2	0	0	0	99.56
	COW	0	1	687	0	2	0	1	0	0	1	99.28
	CRW	0	0	0	970	0	0	1	0	0	0	99.90
	CNW	0	8	0	0	372	0	12	0	6	6	92.08
	SI	0	0	0	0	0	73	15	0	16	5	66.97
	MB	1	1	0	6	7	12	375	12	8	11	86.61
	Dro	0	0	0	0	0	0	4	134	5	1	93.06
	Pole	0	0	0	0	2	6	9	1	678	69	88.63
	Gro	0	0	0	0	0	0	2	0	46	19,932	99.76
Corr. (%)		99.92	99.27	100	99.38	95.63	80.22	89.07	91.16	89.33	99.54	
Qual. (%)		99.76	98.84	99.28	99.28	88.36	57.48	78.29	85.36	80.15	99.30	
Overall Accuracy (%)												98.91
Kappa Index (%)												97.31

3.6.2.2 SrCRF Classification Results

The classification results of SrCRF are summarized in Table 3.5. As in Table 3.5, the overall accuracy of the SrCRF is 98.76%, which shows similar classification performance compared to the SVM results. It is found that the highest classification errors were produced by SrCRF, over the suspension insulator in precision (89.10%) and the movable bracket in recall (76.15%). Thus, the lowest error bound per object produced by SrCRF is higher than SVM. However, similar to the SVM results, it is observed that a classification tendency of SrCRF, which produced higher accuracy over the “major wire” objects and a

lower accuracy over "non-major wire" objects, while the lowest success rate was obtained over "supporting structure" objects. Furthermore, the pole and movable brackets were misclassified with several types of classes. Compared to the SVM results (Table 3.4), the SrCRF improved the classification performance in both correctness and completeness measures for over the suspension insulator and dropper: +10.99% and +3.28% in correctness; +9.18% and +1.34% in completeness, respectively. While, over the movable bracket, +2.07% completeness was improved by SrCRF, but shows similar performance in precision. Major wires and pole remained at a similar level of accuracy.

Table 3.5 Confusion matrix for SrCRF results: electricity feeder (EF), catenary wire (CAW), contact wire (COW), current return wire (CRW), connecting wire (CNW), suspension insulator (SI), movable bracket (MB), dropper (Dro), pole (Pole) and ground (Gro).

		Classified										Comp. (%)
		EF	CAW	COW	CRW	CNW	SI	MB	Dro	Pole	Gro	
Reference	EF	1249	0	0	0	2	0	0	0	0	0	99.84
	CAW	0	1365	0	0	4	0	2	0	0	0	99.56
	COW	0	0	688	0	2	0	1	0	0	1	99.42
	CRW	0	0	0	970	0	0	1	0	0	0	99.90
	CNW	0	30	0	0	349	0	17	0	4	4	86.39
	SI	0	0	0	0	0	83	14	0	11	1	76.15
	MB	0	2	0	6	7	8	384	8	13	5	88.68
	Dro	0	0	0	0	0	0	3	136	5	0	94.44
	Pole	0	0	0	0	2	0	6	0	612	145	80.00
	Gro	0	0	0	0	0	0	3	0	17	19,960	99.90
Corr. (%)		100	97.71	100	99.39	95.36	91.21	89.10	94.44	92.45	99.22	
Qual. (%)		99.84	97.29	99.42	99.29	82.91	70.94	80.00	89.47	75.09	99.12	
Overall Accuracy (%)												98.76
Kappa Index (%)												96.92

3.6.2.3 MrCRF Classification Results

Different with SVM classifier and SrCRF classifier, MrCRF needs a training stage to train both the middle range potential terms and weight matrix described in Section 3.5.1. The look-up table and multivariate Gaussian distribution parameters were trained from the

training data (Section 3.5.1). The weight parameters for different terms were estimated using the SGD algorithm (Section 3.5.1). The weight of the unary term (λ) was always fixed to one. The maximum number of iterations was fixed at 250. Figure 3.10 shows the transition of the weight values according to the iterations. The weight for short-range term α slightly increased and quickly converged to a little higher value than the unary term (λ=1). The weight values for horizontal and vertical long-range terms (β and γ) rapidly decreased at the first stage, and then, the slope was gradually reduced. The results indicate that short-range potential affects the classification results in the proposed CRF model more.

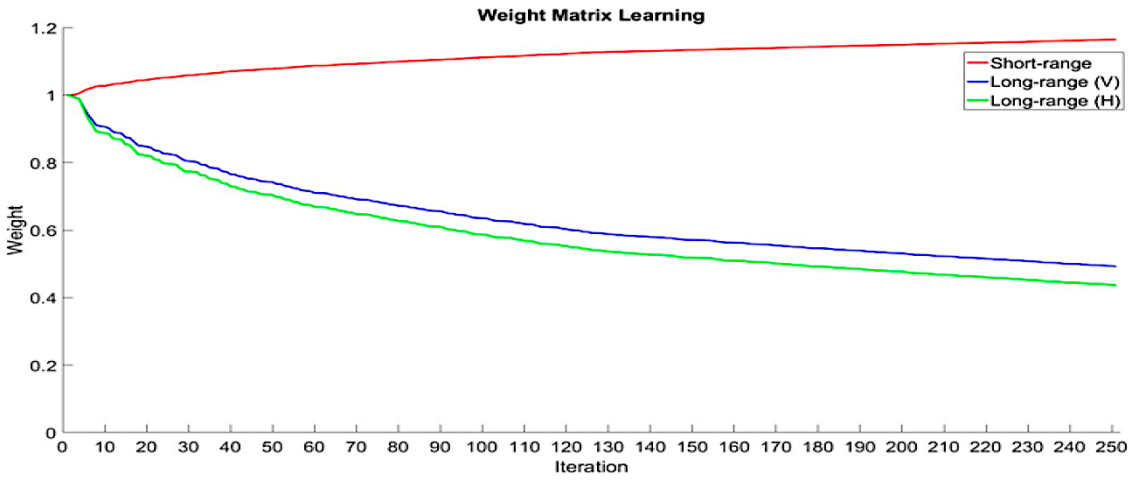


Figure 3.10. Weight learning of MrCRF using the SGD algorithm.

Table 3.6 describes the confusion matrix measuring the classification performance of MrCRF. As shown in Table 3.6, the overall accuracy of the MrCRF is 99.44%. The MrCRF shows that both correctness and completeness values of all classes were higher than 90%. Compared to the SVM results, major wires and ground remained at a similar level of accuracy. Significant improvement was achieved over the suspension insulator (93.58% in completeness and 99.03% in correctness), movable bracket (94.92% in

completeness and 93.84% in correctness) and pole (97.65% in completeness and 93.14% in correctness). However, the classification accuracy for the connecting wire was degenerated, representing 91.58% in completeness and 95.36% in correctness. Completeness for the dropper was also degenerated (90.28%). Overall, the results (Table 3.4) clearly suggest that MrCRF outperformed SVM and SrCRF, by improving not only the overall classification accuracy, but also per-class accuracy.

Table 3.6 Confusion matrix for MrCRF results: electricity feeder (EF), catenary wire (CAW), contact wire (COW), current return wire (CRW), connecting wire (CNW), suspension insulator (SI), movable bracket (MB), dropper (Dro), pole (Pole) and ground (Gro).

		Classified										Comp. (%)
		EF	CAW	COW	CRW	CNW	SI	MB	Dro	Pole	Gro	
Reference	EF	1244	0	0	0	7	0	0	0	0	0	99.44
	CAW	0	1366	0	0	2	0	1	0	2	0	99.64
	COW	0	2	684	0	3	0	1	0	2	0	98.84
	CRW	0	0	0	970	1	0	0	0	0	0	99.90
	CNW	0	8	14	7	370	0	1	0	4	0	91.58
	SI	0	0	0	0	0	102	0	0	7	0	93.58
	MB	0	3	0	11	4	0	411	0	4	0	94.92
	Dro	0	0	0	0	0	0	6	130	8	0	90.28
	Pole	0	0	0	0	1	1	16	0	747	0	97.65
	Gro	0	0	0	0	0	0	2	0	28	19,950	99.85
Corr. (%)		100	99.06	97.99	98.18	95.36	99.03	93.84	100	93.14	100	
Qual. (%)		99.44	98.71	96.88	98.08	87.67	92.73	89.35	90.28	91.10	99.85	
Overall Accuracy (%)												99.44
Kappa Index (%)												98.63

3.7 Discussion

Three different classifiers, including SVM, SrCRF and MrCRF, were developed to classify the railway electrification system objects from MLS data. Table 3.7 summarizes the overall classification performance obtained by three classifiers measured with correctness, completeness and quality score using Eq. (3.15). As in Table 3.7, the SVM classifier produced the lowest classification performance in terms of quality (93.39%) and

correctness (94.35%) compared to SrCRF and MrCRF classifiers, while a similar completeness rate to the one produced by SrCRF. It is found that SVM is the least effective classifier for recognizing the supporting structure objects, including the suspension insulator, movable brackets and poles. However, this result is expected. The supporting structure objects are much more complex compared to the object types. The objects are comprised of multiple parts (e.g., suspension insulator and movable brackets), which cause difficulties to holistically characterize the objects in terms of shape, geometry and spatial relations. Furthermore, the physical size of supporting structure objects (object scale) is large and often attached to the other type of classes. Thus, it can be observed that the lines extracted from the objects were easily fragmented. In this study, line segments are used as features for classification purposes. The fragmented line segments are effective in representing local object characteristics but are not effective to characterize them in their full object scales. Thus, similar feature distributions can be found in different objects.

Table 3.7 A summary of the classification performance achieved by three classifiers (unit: %).

Class	(a) SVM			(b) SrCRF			(c) MrCRF		
	Comp.	**Corr.**	**Qual.**	**Comp.**	**Corr.**	**Qual.**	**Comp.**	**Corr.**	**Qual.**
EF	99.84	99.92	99.76	99.84	100	99.84	99.44	100	99.44
CAW	99.56	99.27	98.84	99.56	97.71	97.29	99.64	99.06	98.71
COW	99.28	100	99.28	99.42	100	99.42	98.84	97.99	96.88
CRW	99.90	99.38	99.28	99.90	99.39	99.29	99.90	98.18	98.08
CNW	92.08	95.63	88.36	86.39	95.36	82.91	91.58	95.36	87.67
SI	66.97	80.22	57.48	76.15	91.21	70.94	93.58	99.03	92.73
MB	86.61	89.07	78.29	88.68	89.10	80.00	94.92	93.84	89.35
Dro	93.06	91.16	85.36	94.44	94.44	89.47	90.28	100	90.28
Pole	88.63	89.33	80.15	80.00	92.45	75.09	97.65	93.14	91.10
Gro	99.76	99.54	99.30	99.90	99.22	99.12	99.85	100	99.85
Average	92.57	94.35	88.61	92.43	95.89	89.34	96.57	97.66	94.41

The most gains achieved by MrCRF come from its discriminative ability improved by enforcing spatial layout regularities (horizontal and vertical layout compatibility) among objects. For instance, all movable brackets, which were misclassified to dropper in the SVM results, were rectified by the middle-range CRF (Figure 3.11a). This is because the horizontal layout term in the middle-range CRF utilizes the placement relations of droppers to the rail track and movable bracket in the horizontal direction, which shows a strong regular pattern (i.e., the dropper is closer to the railway vector than the movable bracket in the horizontal direction). Furthermore, poles, which were misclassified to ground in SrCRF, were well refined (Figure 3.11b). With a similar reason to the dropper case, the misclassification errors over the pole class can be rectified by utilizing the horizontal layout compatibility between the rail track and pole (i.e., the pole is always observed at the farthest position from the rail track in the horizontal direction). In contrast to the horizontal regularity, the suspension insulator and movable brackets were significantly improved in both completeness and correctness by enforcing their vertical regularities in the middle-range CRF (Figure 3.11c).

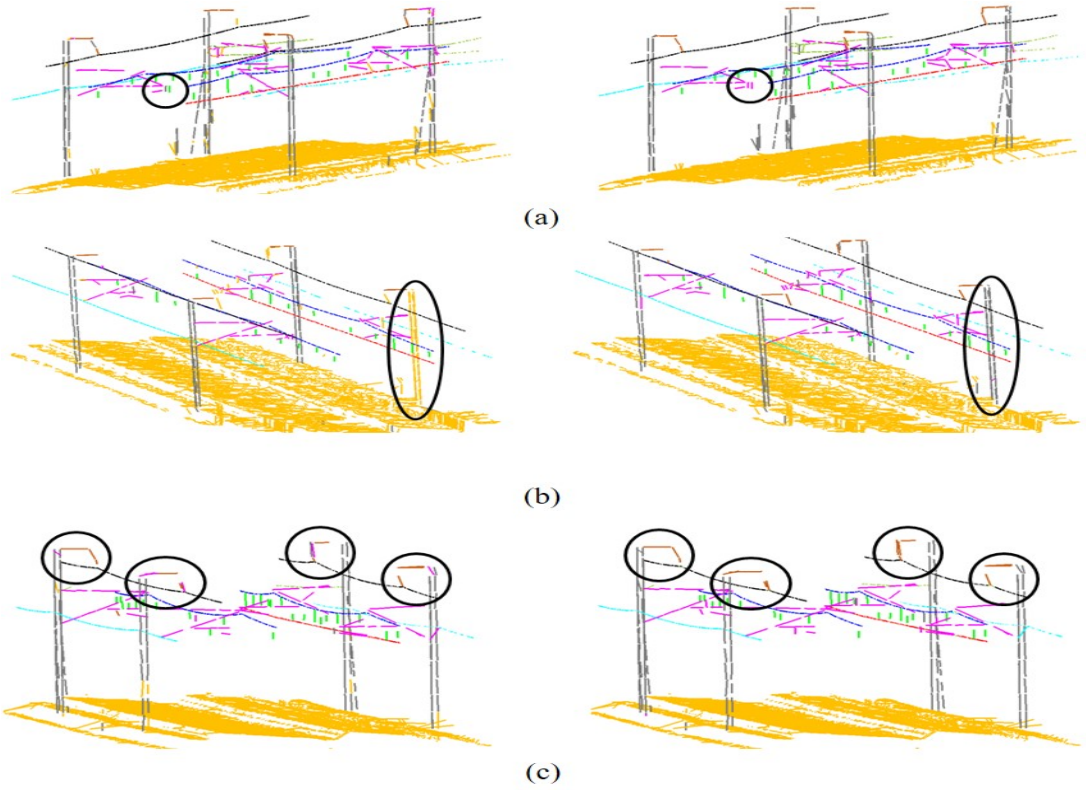


Figure 3.11 The comparison of classification result: (a) the dropper in the SVM results (left) was classified to movable bracket in the integrated CRF model (right); (b) the ground in the SrCRF (left) was classified to pole in MrCRF model (right); (c) the suspension insulators were well rectified in MrCRF model. Black circle represents the misclassified line primitives in SVM corrected by MrCRF.

3.8 Chapter Summary

In this chapter, MrCRF model is described and the classification performance of the model is evaluated on a 1 km dataset in South Korea, which shows the proposed MrCRF achieved significant improvement on the classification results obtained by local classifier SVM. The result clearly reflects how local smoothness and spatial layout encoded in MrCRF model can help the classification decision-making. However, it is found that its performance still

needs to be further improved, especially over the connecting wire and dropper. Both completeness and correctness for the connecting wire were degenerated, as shown in Table 3.7. This degeneracy is caused by a locality of the line segment used for characterizing the spatial layouts. If a set of fragmented line segments is extracted from a single connecting line, their distributions in horizontal locations vary, which leads to the ambiguity of encoding the horizontal layout characteristics between connecting lines and other objects. Furthermore, the completeness of the dropper was lowered. This is due to the fact that the contact wire is missing at certain regions so that the relation between the dropper and contact wire does not follow the defined vertical regularity. These problems can be potentially resolved by encoding the layout regularities with primitives adaptive to object scales and learning more spatial layout at more ranges, which inspires us to propose a multi-layer hierarchical CRF model.

Chapter 4
Multi-Scale Hierarchical Conditional Random Field

In this chapter, a multi-scale hierarchical Conditional Random Filed (HiCRF) framework is proposed to improve the classification results obtained by MrCRF proposed in the previous chapter. The MrCRF introduces multi-scale line representation and full-range spatial regularities. A hierarchical multi-scale data representation such as image pyramid have shown its success in pattern recognition tasks. In this hierarchical framework, spatial location between objects can be learnt on different scales to adopt different object geometry properties. Also, this hierarchical framework can incorporate the information and combine all the spatial cues from different scales. For instance, Yang and Forstner (2011) proposed a hierarchical model to favor region consistency for classifying man-made objects. Albert et al. (2017) used a two-layer CRF to both classify land cover and land use. The success of these methods used multi-layer hierarchical model to encode contextual relations among entities at different scales. Moreover, recently proposed fully-connected CRF model Koltum (2011) extends contextual information for any entity pair, which largely enriches the contextual information formulated in the model. Therefore, to overcome the limitations of proposed MrCRF model in limited ranges and inaccurately learnt relative displacement. A multi-scale hierarchical CRF (HiCRF) model is introduced to understand relative spatial displacement at multiple scales with arbitrary range. To represent objects at different scales, the line primitives will be merged according to line geometric similarity and

consequently the corridor scene will be represented by lines at three different scales. The HiCRF with three different layers will correspondingly encode the contextual information of line primitives individually and mutually. In the bottom and middle layer, a similar idea of MrCRF introduced in the previous chapter will be applied to encode the local smoothness and relative spatial displacement at different scales. The top layer will be a fully-connected CRF model to learn the relative spatial distribution at arbitrary range in the whole scene at an approximate object scale.

First, it introduces a seed-fill based line merging algorithm to generate a multi-scale line representation of mobile laser scanning data in Section 4.1. The structure of a HiCRF to extend MrCRF model will be explained in Section 4.2. Then, the bottom and middle layer design in proposed three-layer hierarchical model will be demonstrated in Section 4.3 while the top layer design will be explained in Section 4.4. The training and inference of this model will be explained in Section 4.5. The evaluation of the results on the same dataset shown in Section 3.6. The comparison and analysis between different methods will be conducted in Section 4.6. A chapter summary will be made in Section 4.7.

4.1 Multi-scale Data Representation

Line primitives restricted in a certain scale (depends on voxel size) is over-segmented and can affect the correctness of spatial relation learnt in a local range. Multi-scale data representation is a common strategy to express data into different levels and it can also propagate longer range dependencies (Yao et al., 2012). For image data and point cloud data format, there are several segmentation methods to segment them into meaningful

patches according to the homogeneity of individual primitives. Similar to this idea, line primitives based on their geometric similarities are aggregated. Specifically, a seed line is randomly selected and its nearest neighbor will be checked. If the differences of horizontal and vertical angle of two lines are within thresholds, the algorithm will consider merging these two lines and this closet neighbor line will become the next seed line. Otherwise, a new seed line will be randomly selected. This process is repeated until all line primitives are processed. Then, the associated point is found with each line primitive and the points of merged lines will be used to generate new lines using the same RANSAC algorithm. Finally, this procedure process will repeat once to generate another line representation with a more relaxed merging threshold. It is worth to note that this line representation will be closest to instance data representation. Therefore, the mobile laser scanning data is represented into three different representations. The flow chart for this line merging is shown is Figure 4.1 as follow:

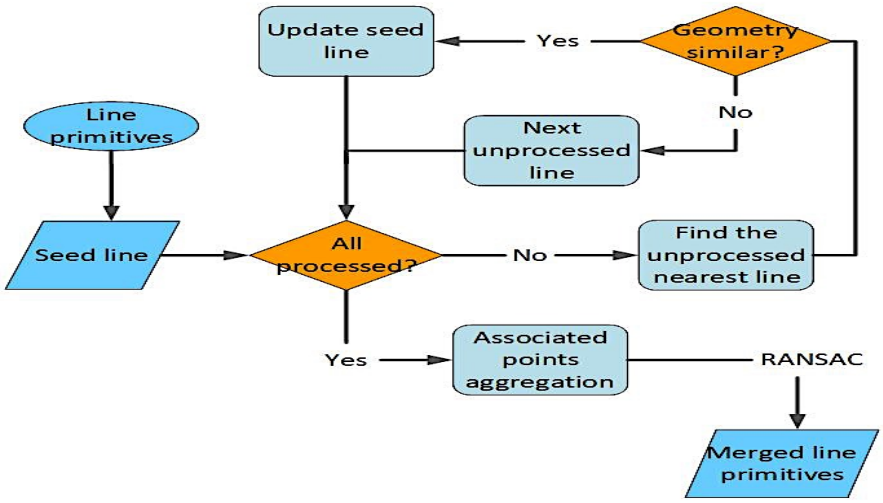


Figure 4.1 Flow chart of line merging algorithm to generate multi-scale representation.

4.2 Graph Model Design

The goal of the HiCRF model is to consider the aforementioned concept of the multi-scale line representation so that a three-layer CRF model is designed to encode the relations among line primitives both at individual scale and different scales. The mathematical model of HiCRF is expressed as follow:

$$p(Y|X) = \frac{1}{Z(X)} exp\left[E_1(y_i, X) + E_2(y_i, y_j, X) + E_3(y_i, y_j, X)\right] \tag{4.1}$$

Where E_1 is the sum of unary potential of all line primitives in three layers and E_2 is the sum of all pairwise potentials within each layer. E_3 is the sum of hierarchical pairwise potential between bottom and middle layers and also between middle and top layers. The specific formulation of each term is defined as:

$$\mathrm{E_1}\left(\mathrm{y_i},\mathrm{X}\right)=\lambda \sum_{\mathrm{i}\in\mathrm{B}} \varphi_{\mathrm{i}}\left(\mathrm{y_i},\mathrm{X}\right) +\eta \sum_{\mathrm{i}\in\mathrm{M}} \varphi_{\mathrm{i}}\left(\mathrm{y_i},\mathrm{X}\right) +\delta \sum_{\mathrm{i}\in\mathrm{T}} \varphi_{\mathrm{i}}\left(\mathrm{y_i},\mathrm{X}\right) \tag{4.2}$$

$$\begin{aligned} E_2(y_i, y_j, X) = {} & \alpha \sum_{i\in B} \sum_{j\epsilon N_i^{BS}} \varphi_{ij}^{BS}(y_i, y_j, X) + \mu \sum_{i\in B} \sum_{j\epsilon N_i^{BM}} \varphi_{ij}^{BM}(y_i, y_j, X) \\ & + \beta \sum_{i\in M} \sum_{j\epsilon N_i^{MS}} \varphi_{ij}^{MS}(y_i, y_j, X) + \nu \sum_{i\in M} \sum_{j\epsilon N_i^{MM}} \varphi_{ij}^{MM}(y_i, y_j, X) \\ & + \gamma \sum_{i\in T} \sum_{j\epsilon N_i^{T}} \varphi_{ij}^{TF}(y_i, y_j, X) \end{aligned} \tag{4.3}$$

$$\begin{aligned} E_3(y_i, y_j, X) = {} & \rho \sum_{i\in BM} \sum_{j\epsilon N_i^{BM}} \varphi_{ij}^{BM}(y_i, y_j, X) \\ & + \omega \sum_{i\in MT} \sum_{j\epsilon N_i^{MT}} \varphi_{ij}^{MT}(y_i, y_j, X) \end{aligned} \tag{4.4}$$

Where superscript set $\{B, M, T\}$ is added to represent variables in bottom, middle and top layer respectively. Second superscript set $\{S, M, F\}$ corresponds to intra-layer edges at short, middle and full range. Variable $\{i, j\}$ indicates nodes in the graph while N_i shows intra-layer or inter-layer edges which connect node i. The variable y represents node label and X is node feature vectors. $\varphi(.)$ is potential function in the model while $Z(X)$ is the partition function to normalize the potential function into probabilities. The weight parameter set $\Omega = \{\lambda, \eta, \delta, \alpha, \mu, \beta, \nu, \gamma, \rho, \omega\}$ express the relative importance of each potential term in the model.

Figure 4.2 illustrates the detail graph structure of this model. The HiCRF model is composed of three layers in which corresponds each scale of line primitives representing original point cloud data. In the bottom and middle layer, a multi-range CRF model will be utilized individually to learn local smoothness at short range and relative spatial regularities at middle range. A fully connected CRF model will be introduced in the top layer to encode relative spatial regularities at arbitrary range. To express the relation among all line primitives in three layers, there are two types of edges: intra-layer edge (blue lines in Figure 4.2) and inter-layer edge (red dotted lines in Figure 4.2). The intra-layer edge represents the line primitives relations at each layer, while the inter-layer edges encode the relation between line primitives between adjacent layers. The intra-layer edge in bottom and middle layer will be determined by adjacency graph defined in Section 3.2 with same setting while it will be fully connected in top layer of the proposed model. The inter-layer edge is dependent on the line merging procedure which means the edge will be established among line primitives if a line primitive is merged to another line primitive.

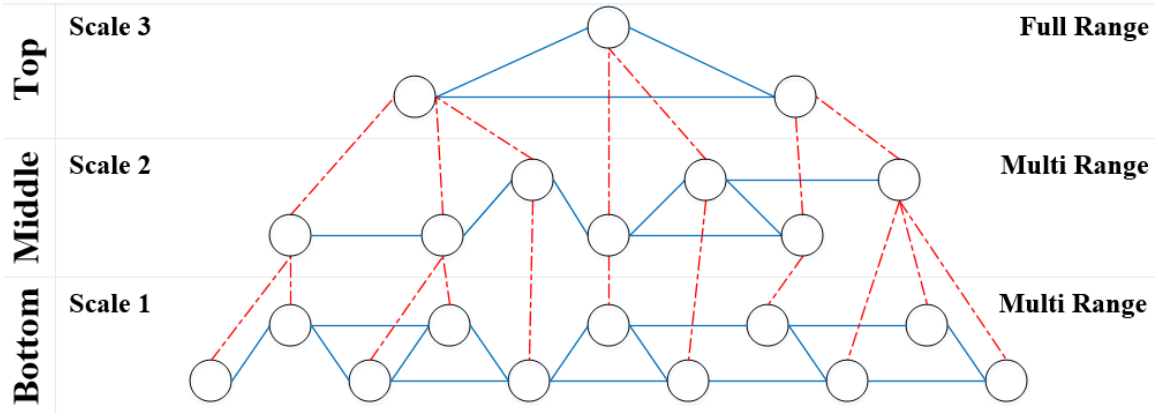


Figure 4.2 Hierarchical three-layer CRF (HiCRF) model. The circle represents node as line primitives and the size of the circle depicts the line primitive scale. Blue lines demonstrate intra-layer edges while red dotted lines show inter-layer edges. All edges are pairwise (without high-order cliques).

4.3 Bottom and Middle Layers

In bottom and middle layers, both local smoothness and multi-scale spatial regularities are introduced into classification. Therefore, a simplified multi-range CRF model is introduced in these two layers. The core idea of this simplified multi-range CRF model is also considering different context information adapted at different ranges. The multi-range graph will combine a short-range graph to apply local smoothness and a middle-range graph to apply relative spatial layout (Jung et al., 2016). However, the vertical and horizontal pairwise terms will be combined into a middle-range pairwise term. Considering the monotony of logarithm, the multi-range CRF model will be expressed as below:

$$p(Y|X) = \frac{1}{Z(X)} exp\left[\lambda \sum_{i \in S} \varphi_i(y_i, X) + \alpha \sum_{i \in S} \sum_{j \in N_i^S} \varphi_{ij}^S(y_i, y_j, X) + \beta \sum_{i \in S} \sum_{j \in N_i^M} \varphi_{ij}^M(y_i, y_j, X)\right] \quad (4.5)$$

As seen from the Eq. 4.5, $\varphi_i(y_i, X)$ is a unary potential representing the importance of individual node. $\varphi_{ij}^S(y_i, y_j, X)$ is short-range pairwise potential encoding local smoothness while $\varphi_{ij}^M(y_i, y_j, X)$ is middle-range pairwise potential modeling relative spatial layout at a certain middle range in the whole scene. $N_i^S$and N_i^M are nodes connected to node i at short-range and middle-range respectively. $\{\lambda, \alpha, \beta\}$ are the weight parameters for corresponding three sub-terms. Keeping consistency with our previous work (Jung et al., 2016) to define the adjacency graph and unary potential, a novel short-range pairwise term and middle-range pairwise term will be explained in the Section 4.3.1 and Section 4.3.2.

4.3.1 Short-Range Binary Term

Local smoothness is a very general assumption that neighboring entities such as pixel or individual point prefer to belong with same object and this is more significant with the increase of sensor resolution (Schindler, 2012). In a short-range domain, this assumption is also applicable to scene interpretation. The Potts model was firstly proposed to encode the smoothness into a CRF framework which simply give penalty if two connected nodes in CRF are assigned different labels, and vice versa. However, this arbitrary penalty does not consider data feature and it sometimes causes over-smoothness. Therefore, contrast-sensitive Potts model was introduced to compromise between data feature and smoothness

degree. It still gives a penalty to nodes having the same label and the penalty assigned to connected nodes with different labels, which will be determined by the similarity between feature vectors of two nodes with its Euclidean distance (Niemeyera et al., 2016):

$$\varphi_{ij}^{S}(y_i, y_j, X) = \begin{cases} 0 & if\ y_i \neq y_j \\ p + (1-p)e^{-\frac{{d_{ij}}^2}{2\sigma^2}} & if\ y_i = y_j \end{cases} \quad (4.6)$$

Where $d_{ij} = \left\| f_i(x) - f_j(x) \right\|$ is the feature vector Euclidean distance. p measures the degree of data-dependent smoothing. The value of p is between from 0 and 1 while larger p means the smoothing is less-dependant on the data, and vice versa. The parameter σ^2 determines how large feature similarity can affect smoothing degree. Usually it corresponds to the mean value of all squared feature distance (Niemeyera et al., 2016) and can be determined in the training stage. This contrast-sensitive Potts model can preserve some small objects while have a smoothing effect on background.

4.3.2 Middle-Range Binary Term

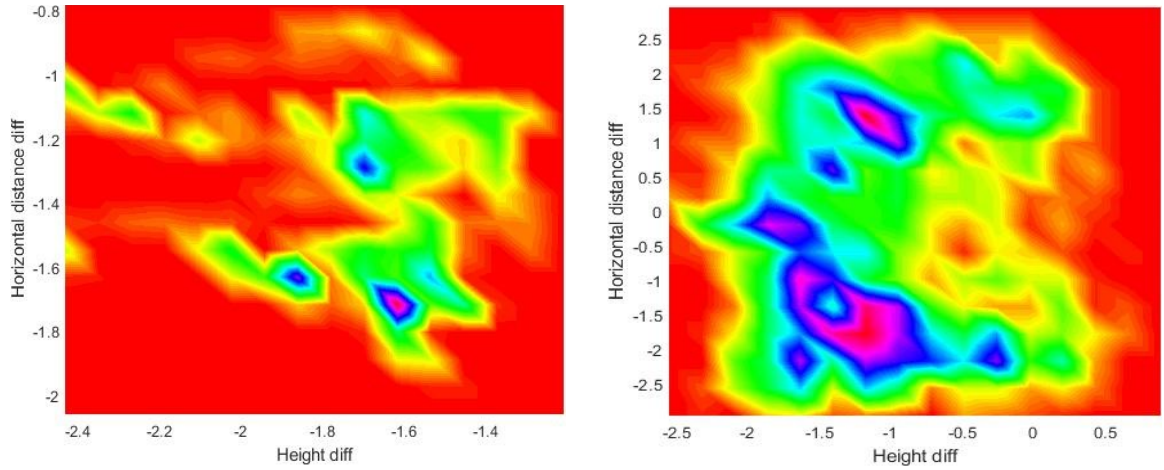
Spatial relationship is an important global feature in different scenes (Gould et al., 2008). This relationship is very general and can be applied to different data sources. In corridor scene, there is obvious spatial regularities existed in both vertical and horizontal direction. For instance, diverse types of transmission lines distribute vertically in different heights and catenary wire is always horizontally closer to railway track compared with the current return wire. The task is learning all this inter-class spatial relation at a certain range from training data and use this relative spatial information to guide classifier to correct misclassification which does not follow trained spatial regularities. Instead of the previous

method (Jung et al., 2016) using two potential terms to encode horizontal and vertical relation from ad-hoc multivariate Gaussian distribution assumption, a Relative Location Prior potential (Volpi and Ferrari, 2015) was applied for middle range pairwise term to encourage similar spatial arrangement found in the training data. The potential is formulated as follows:

$$\varphi_{ij}^{M}(y_i, y_j, X) = \log\left[\frac{1}{m}\sum_{n=1}^{m} f_{l|k}\left(d_{lk}^{n}, d_{ij}\right)\right] \tag{4.7}$$

Where d_{lk} are relative location vectors between all the nodes labeled as l and connected nodes labeled as k in the training data. d_{ij} is the relative location vector between node i and node j. Relative location vector is a two-dimensional vector in which the first element is the height difference of line primitives while the second element is the difference of line primitives' horizontal distance to its corresponding railway track. Noted that specific class-pair relative location vectors can be both "different" displacement and "self" distribution. To be more specific, $l = k$ means the object itself spatial distribution within middle range. Otherwise, it shows relative location between two different objects. $f_{l|k}\left(d_{lk}^{n}, d_{ij}\right)$ is a Radial Basis Function (RBF) kernel to measure the similarity between vector d_{ij} and a relative location vector d_{lk}^{n} in the training displacement vectors for class label l and k. m is the number of all training displacement vector between label l and label k. Figure 4.3 shows a learnt relative location vector distribution for specific class-pair and another class itself spatial distribution. As shown in the Figure 4.3(a), catenary wire is most obviously horizontally 1.6 meter closer to its railway track compared to electricity feeder while

vertically mostly 1.7 meter lower than electricity feeder. Figure 4.3(b) demonstrates how movable bracket line primitives horizontally and vertically distribute in a middle-range.



(a) (b)

Figure 4.3 Example of relative displacement vector. (a) shows the relative displacement vector distribution for the electricity feeder and catenary wire in horizontal and vertical direction. (b) shows movable bracket 'self' spatial distribution.

4.4 Fully Connected Top Layer

In the bottom and middle layer, the relative location information is learnt through different scale. However, this spatial information is driven based on the line primitive captured at a certain range. There are also longer spatial relations, which is not able to be represented with fixed range constraints. For instance, usually two consecutive poles have a certain pre-defined distance on one side of the railway. However, this information cannot be expressed in relative location prior. To encode the spatial relations at any arbitrary range and learn the spatial relation at instance scale, a fully connected CRF model is proposed to connect all line primitives in the top layer. The fully connected CRF model (Arnab et al.,

2016; Chen et al., 2018; Koltun, 2011; Zheng et al., 2015) already became a standard technique for structure prediction, especially image semantic segmentation tasks. However, usually these fully connected CRF models need some specific type of potential function to ensure efficient inference (Campbell et al., 2013; Koltun, 2011; Wang et al., 2015; Zhang and Chen, 2012). To connect all the line primitives, all line primitives in top layer will be connected to any others at one side of the railway. Under this setting, the number of the fully-connected edges is still tractable to apply general inference algorithm because there are fewest line primitives at top layer. To encode relative spatial information at arbitrary range, a relative displacement probability map which inspired from Gould et al. (2008) is constructed. Different with their map in 2D, the map is a 3D map based on the coordinate offset of points. This map $\mathrm{M}_{l|k}(d_{ij})$ encodes the conditional probability that the line primitive j with a label l given a line primitive i with the label k at a relative displacement vector d_{ij}, which is a three-dimensional vector to express the relative displacement between two lines in both x, y and z direction. To be noted that the coordinate of all points is normalized to the range $[0,1]$ so the relative displacement vector $d_{ij} \in [-1,1] \times [-1,1] \times [-1,1]$.

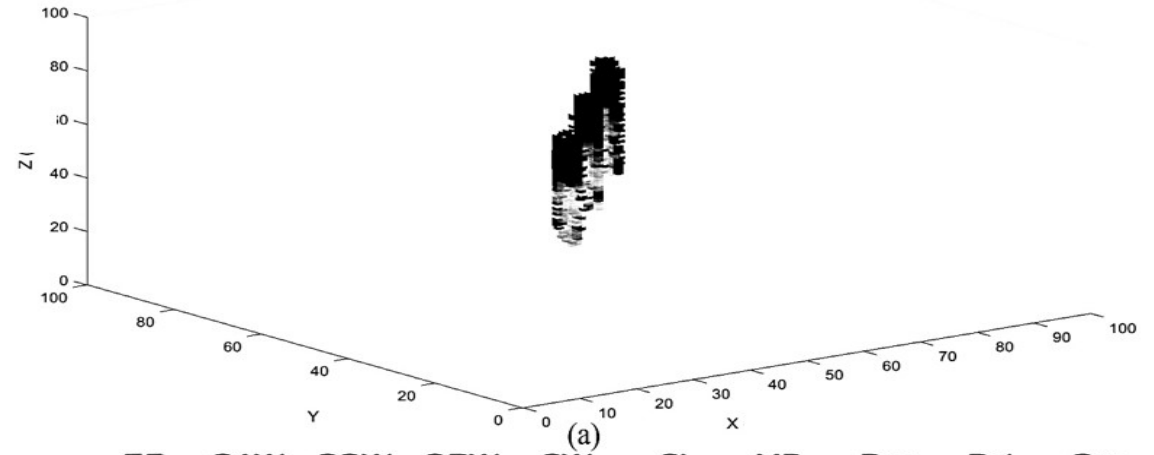
The map is learnt through counting the number of relative displacement of each class-pair from the centroid of each line primitives in the training data and it will quantize into a bin over the size $100 \times 100 \times 100$. If $l \neq k$ in the class-pair lk within the edge, both $\mathrm{M}_{l|k}(d_{ij})$ and $\mathrm{M}_{k|l}(-d_{ij})$ will add a count while $l = k$ indicates that it represents "self" relative displacement for every class label. Finally, the map will normalize to ensure that $\sum_{l=1}^{L} \mathrm{M}_{l|k}(d_{ij}) = 1$ for a given class label k and L is the number of all class labels.

The example of learnt relative displacement probability map is shown in Figure 4.4. It is observable that movable bracket and pole relative displacement almost only occurs at z axis and certain fixed x and y interval. It is natural that movable bracket is a pole-subordinate structure and it usually only has height difference with its belonged pole and fixed x and y interval demonstrates the displacement of two consecutive poles in the scene.

The potential function of fully connected top layer will encourage the configuration which follows the layout from the map, the specific formulation is as follows:

$$\varphi_{ij}^{TF}(y_i, y_j, X) = \log\frac{1}{2}\left(\mathrm{M}_{y_j|y_i}(d_{ij}) + \mathrm{M}_{y_i|y_j}(-d_{ij})\right) \tag{4.8}$$

The designed potential is not symmetric because the relative displacement probability map is asymmetric, which encourages the consistency with the relative spatial regularities from training data and penalizes the opposition.



(a)

	EF	CAW	COW	CRW	CW	SI	MB	Dro	Pol	Gro
EF	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
CAW	0.00	0.50	0.00	0.00	0.00	0.00	0.30	0.20	0.00	0.00
COW	0.00	0.00	0.73	0.00	0.00	0.00	0.00	0.27	0.00	0.00
CRW	0.00	0.00	0.00	0.83	0.00	0.00	0.17	0.00	0.00	0.00
CW	0.00	0.03	0.00	0.00	0.84	0.00	0.10	0.03	0.00	0.00
SI	0.50	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00
MB	0.00	0.08	0.02	0.06	0.08	0.00	0.63	0.02	0.12	0.00
Dro	0.00	0.15	0.08	0.00	0.15	0.00	0.08	0.54	0.00	0.00
Pol	0.05	0.00	0.00	0.03	0.14	0.00	0.05	0.00	0.72	0.00
Gro	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(b)

Figure 4.4 Learnt displacement map. (a) 3D relative displacement probability map for the class pair pole and movable bracket. (b) a table of all conditional probabilities at a certain relative displacement. Darker means higher probability.

4.5 Training and Inference

As discussed in previous sections, there are two types of parameters to be determined in this model, the first type is the parameters within each potential function such as balancing weight in contrast sensitive Potts model and free weight in RBF kernel based similarity measurement. Another type is the set of weight parameters among each potential function. Besides these estimated parameters, unary terms in individual layer should be trained from

discriminative classifiers. the same two-stage training strategy is also utilized to train the multi-layer HiCRF model. After training, an inference algorithm is applied to classify those line primitives in test sites. However, the difference between this training and inference with MrCRF model will be explained in next two subsections.

4.5.1 L-BFGS Based Training

In the two-stage training strategy, the first stage is determining all training parameters except the weight set. For three unary terms in each layer, three independent Support Vector Machine (SVM) classifiers will be trained individually without considering any pairwise potentials. In the multi-range model for the bottom and middle layer, balancing weight in contrast-sensitive Potts model and free weight in RBF kernel function should be determined for each layer individually. Due to the gradient calculation difficulty, balancing weight in contrast-sensitive Potts model will be manually set and the free weight in RBF kernel function will be determined for each class-pair by calculating the mean value of squared distance of all location vectors. In the second stage, a MAP based training strategy is applied to maximize the log conditional likelihood for simultaneously training all weights among all potential term. Instead of using SGD optimization algorithm in previous MrCRF training, the optimization is conducted through a Limited Memory BFGS (L-BFGS) optimization because it can use inverse Hessian matrix to steer its search direction and consequently has a faster convergence. The gradient calculation is same as described in Section 3.5.1.

4.5.2 Maximum Posterior Marginal Inference

The popular MAP inference usually maximizes the posterior probability defined in Eq. 4.1. However, it will have two potential problems in high dimension applications. The estimated posterior probability can be extremely small and also it cannot distinguish the wrong prediction with single error at one component or entirely wrong (Domke, 2013). Another alternative is to maximize the marginal probability of label distribution $P(y_i|x)$ at each node site, which means it minimizes the number of mislabeled node sites (He et al., 2004). This Maximum Posterior Marginal (MPM) inference was reported to have a better performance on image labeling task (Toyoda and Hasegawa, 2008), which will be adopted in this case. The computation of exact MPM inference is usually intractable (He et al., 2004;Toyoda and Hasegawa, 2008) for multi-class classification problem. Therefore, the mean field approximation is applied because it is usually faster than LBP algorithm and can be extended for large scale inference task (Koltun, 2011). Besides, it doesn't require submodular potential function, which is not applicable in this potential function design.

4.6 Experiment and Results

To evaluate the algorithm, the same Korean high-speed railway electrification scene dataset described in Section 3.6 is also applied. The classification performance is also evaluated on six-folder cross validation. Firstly, the line extraction and merging from original massive point cloud will be analyzed to construct a multi-scale representation in Section 4.6.1. Then classification result with several other classifiers will be compared in Section 4.6.2. Finally, an analysis on classification results will be done in Section 4.6.3.

4.6.1 Multi-Scale Line Extraction

A RANSAC algorithm will be firstly applied to generate lines from each voxel after voxelization. The voxel size is 1 meter while RANSAC has a 5 centimeters consensus threshold. Multiple lines can be generated in the same voxel. This setting is same as in the previous multi-range line extraction. Then, the generated over segmented lines were merged continuously to generate another two scales of line representation. These three-scale line representations will be expressed as scale 1 to scale 3 from RANSAC extracted lines to final merged lines in the later paragraph. The goal is trying to get a line representation which is close to the instance level representation of original point cloud. Also, it is also important that the threshold cannot be too loose to merge lines from different objects.

The Figure 4.5 demonstrates the merging result of one subset and it is noticeable that this line representation preserves the detail of object structures in the scene and many objects are much more normalized to instance level in a higher scale. For example, electricity wires are usually represented by many small line primitives with similar length while it is represented as several lengthy line primitives in the last scale. Meanwhile, those more complex structures such as movable bracket is still well preserved with several line primitives. The specific line number of each sub region in three different scales is also shown in Table 4.1. Original point cloud with approximate two hundred thousand points in each sub region was converted to thousands of line primitives but preserves the enough detail of each object instance.

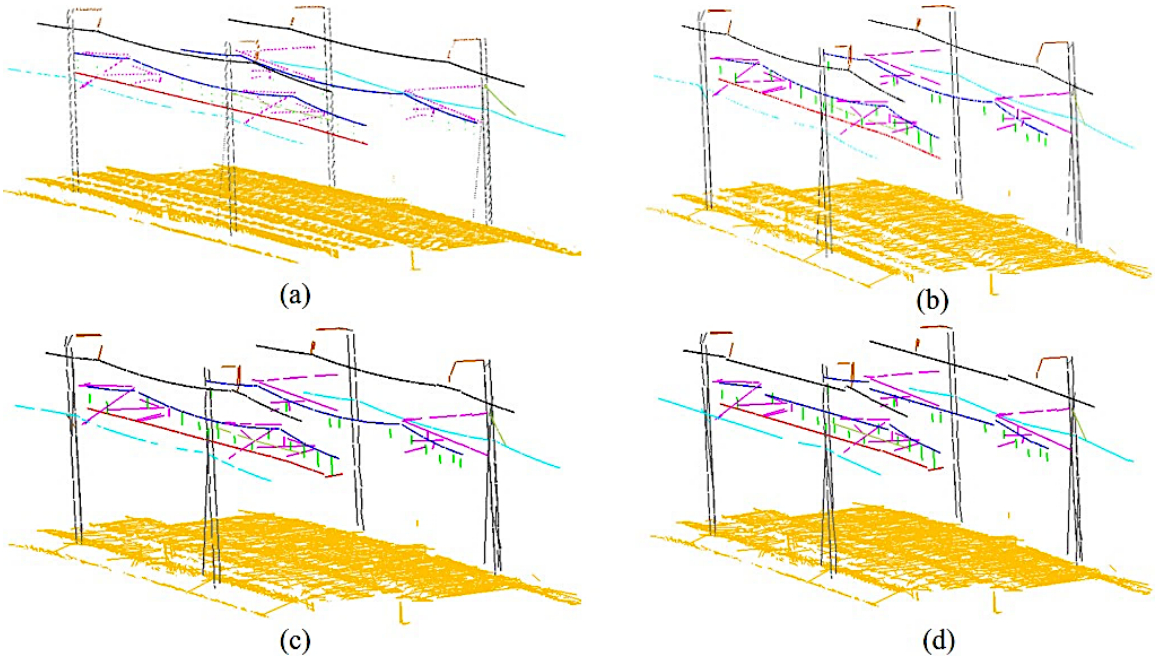


Figure 4.5 Three-scale line representation. (a) is the original pint cloud in subset 1 and (b) is the extracted line primitives using RANSAC algorithm. (c) shows the merged line primitives and (d) represents final merged line primitives with relaxer thresholds. (**black**: electricity feeder, **blue**: catenary wire, **red**: contact wire, **sky blue**: current return wire , **dark green**: connecting wire, **brown**: suspension insulator, **magenta**: movable bracket, **green**: dropper, **grey**: pole and **yellow**: ground).

Table 4.1 Line number in three scales for each sub-region

Line Number	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
Scale 1	4327	4438	3783	4093	4724	4762
Scale 2	2830	2904	2420	2623	2848	3063
Scale 3	1732	1687	1379	1466	1540	1735

4.6.2 Classification Results

The classification of six sub-regions has been done on the same desktop described in section 3.6.2. A total of 268.32s was achieved for classifying all datasets, which was

slightly faster than MrCRF due to the optimization of the algorithm. For the training stage, relative displacement probability map will cost around 30s for each sub-region and the weight training will spend more time, which varies from 3 minutes to 10 minutes. This much faster training contributes from the application of L-BFGS algorithm and also the code optimization with the introduction of parallel computing.

Table 4.2 shows the classification results obtained by proposed HiCRF. For obtaining these results, the same geometric line features is applied to calculate the potential for each term in HiCRF. The training data was used to create relative location prior, relative displacement probability map with bin size 100 and the weight set among all potential functions. An equal weight was initialized and usually the weight training will usually converge after around 25 epochs. Noted that the relative location prior will be generated twice for scale 1 and scale 2 while relative displacement probability map was only generated for scale 3. Three SVM classifiers (Chang and Lin, 2011) were trained independently. Then, the output of three SVM classifiers were served as the unary potential and the graph for each layer was generated subsequently. The MPM based mean field approximation provided by Domke (2013) was finally conducted to determine the final label of all line primitives.

Table 4.2 Confusion matrix of HiCRF model

		Classified										Comp. (%)
		EF	CAW	COW	CRW	CNW	SI	MB	Dro	Pole	Gro	
Reference	EF	1251	0	0	0	0	0	0	0	0	0	100
	CAW	0	1363	0	0	5	0	1	0	0	0	99.56
	COW	0	0	689	0	1	0	1	0	0	0	99.71
	CRW	0	0	0	970	0	0	0	0	1	0	99.90
	CNW	0	7	0	0	385	0	2	0	3	1	96.73
	SI	0	0	0	0	0	106	2	0	3	0	95.50
	MB	0	1	0	1	1	3	405	3	4	0	96.89
	Dro	0	1	0	0	0	0	2	140	0	0	97.90
	Pole	0	0	0	7	6	1	0	0	760	7	97.31
	Gro	0	0	0	0	0	0	3	0	20	19,970	99.88
Corr. (%)		100	99.34	100	99.18	96.73	96.36	97.36	97.90	96.08	99.95	
Qual. (%)		100	98.91	99.71	99.08	93.67	92.24	94.41	95.89	93.46	99.84	
Overall Accuracy (%)												99.67
Kappa Index (%)												99.18

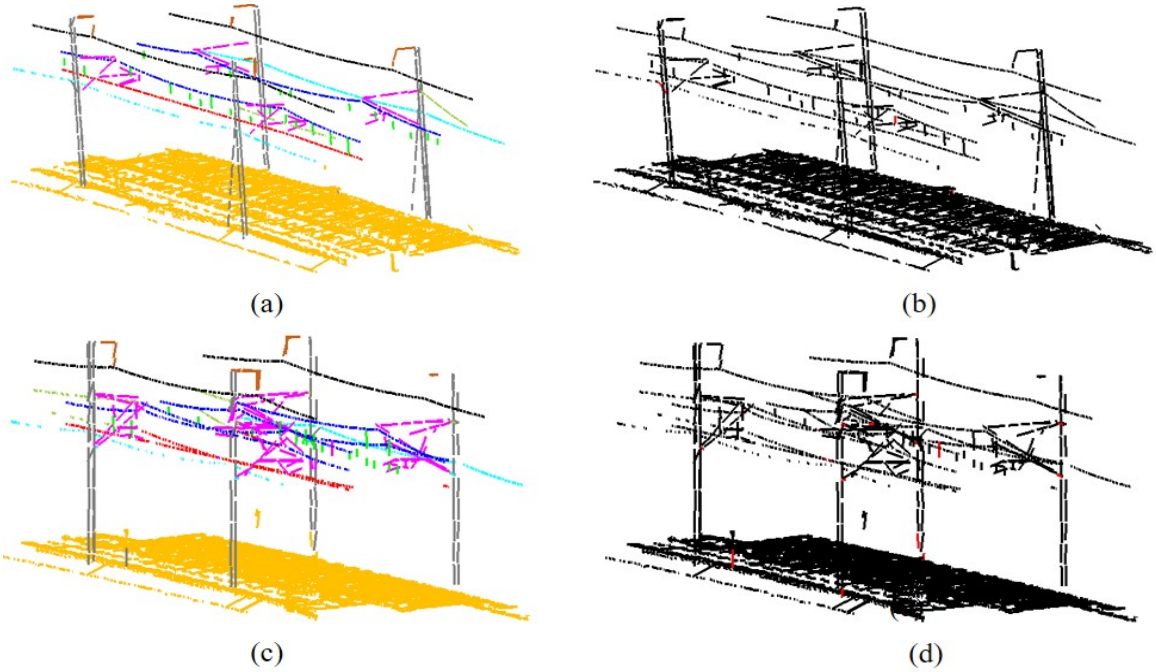


Figure 4.6 Multi-layer hierarchical classification result: (a) is the visualization of the classification result of subset 1 and (b) shows the misclassified line primitives in subset 1 with red and correct classification with black color. (c) and (d) presented similar results for subset 5.

A conducted comparative analysis of HiCRF performance to the results produced by other classifiers, which include SVM, MrCRF and variant of HiCRF (i.e., short-range

bottom layer and two layers) will be conducted. For this performance measure, the same classification evaluation index explained in previous section (Rutzinger et al., 2009) is also used. Table 4.2 demonstrates that the proposed HiCRF can produce an over 99% overall accuracy and kappa index. With respect to the classification performance for each class, it is observed that a performance (almost 99% quality) was achieved on those electricity wires except for connecting wire. The minor objects such as suspension insulator, movable bracket and dropper have relatively lower quality performance, but the quality value of worst suspension insulator was still over 92%. The ground object was detected with a very high classification quality due to its limit on the height. HiCRF produced misclassification error mostly generated between two objects including pole-ground, movable bracket-suspension insulator, catenary wire-connecting wire. It is not a trivial task to discriminate these objects as they are located closely with each other and their appearances are similar. The subset 1 reached the best classification performance, while the subset 5 had the worst. From the visual inspection, the performance variation is mainly caused by the significant difference of spatial distribution between two subsets. The visualization of the classification result of these two subsets were shown in Figure 4.6. In general, the proposed model achieved a very good classification result on those electricity wires, while it also produced slightly lower accuracy but still competitive classification result on those complex and small electrification-related objects in the scene. Also, the challenging poles also had a good classification quality, which was over 93%.

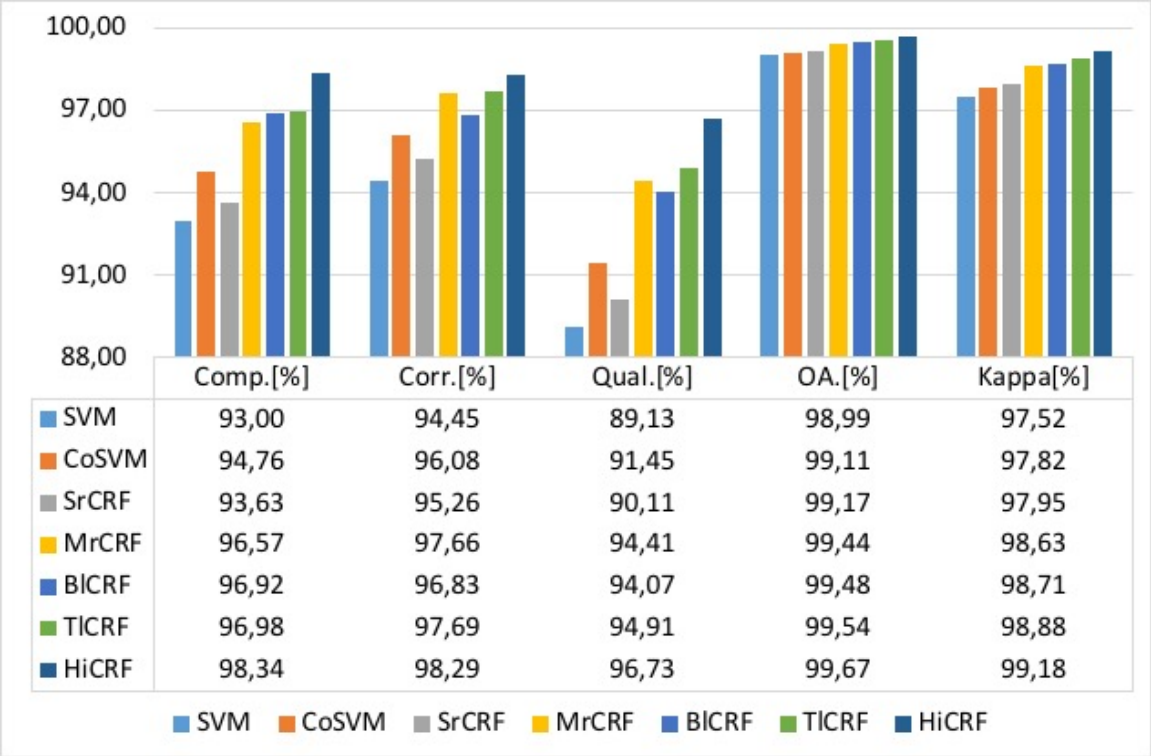


Figure 4.7 The average quantitative classification result of different models.

The HiCRF classification results produced by HiCRF ware compared with following several classifiers: (1) SVM classifier used for the first layer unary term (SVM) (2) SVM with context features (the feature difference of edge features in MrCRF model) (3) CRF with only short-range term (SrCRF) (4) MrCRF model described in Section 3. (5) CRF with short and middle range terms in bottom layer (BlCRF) (6) two-layer CRF model without top layer in proposed model (TlCRF). Figure 4.7 shows the results of comparative performance analysis.

The SVM classifier without any contextual information produced the worst performance while HiCRF outperformed all other classifiers in all classification evaluation indices. With regard to the performance of two SVM classifiers, it is observable that

CoSVM outperforms SVM with only geometry features. It is shown that context features indeed can help the classification. Meanwhile, SVM with context features still had worse performance than MrCRF model which encoded explicit context information into a CRF model, which showed that CRF model is more effective in expressing context information than SVM. Comparing the classification results of all different CRF models, with more introduced contextual information, the performance of the CRF model improved accordingly. For instance, the local smoothness constraint improved the SVM classification result by around 1% in both average completeness and correctness. With learnt spatial relative location in middle range at one scale, the result of quality improved by 4%. Another 2.5% improvement on average classification quality was obtained through this supervised learning of multi-scale relative spatial distribution. Meanwhile, the classification results of BlCRF achieved the similar performance produced by MrCRF reported by Jung et al. (2016) with simpler pairwise functions. The TlCRF showed slightly better performance compared to the results of BlCRF, especially correctness and quality measure. This indicates that introducing an additional graph layer can positively contribute to the classification results. Specifically, the comparison of proposed HiCRF model and SVM classification result was illustrated in Table 4.3. Firstly, it is very observable that HiCRF model improves the classification results of suspension insulator (+36.79%), movable bracket (+11.81%), dropper (+10.08%) and pole (+13.51%) although these objects are quite challenging for classification due to its variant geometry and spatial distribution. All types of electricity wires have limited improvement except for a slight decrease on catenary wire and current return wire. To be noted that the confusing

connecting wire result was significantly improved. This is important because it is very hard to distinguish connecting wire and other electricity wires, especially in overlapped areas. Four typical electricity wires with certain height almost achieved perfect performance both in SVM and HiCRF model. Also, the misclassification between ground and pole which heavily occurred in SVM classification was significantly corrected. The detailed label transition from SVM to HiCRF model shown in Figure 4.8 also demonstrates that proposed model could effectively correct the misclassification in SVM result, especially those lines misclassified to pole due to imbalanced training. For instance, all transitions from pole to ground are positive and more than 70% transitions from ground to pole are positive, which largely improved the overall accuracy.

Table 4.3 Classification result of proposed model and SVM

	SVM (Without context)			HiCRF (With context)		
	Comp.[%]	Corr.[%]	Qual.[%]	Comp.[%]	Corr.[%]	Qual.[%]
EF	99.84	100.00	99.84	100 (+0.16)	100.00 (0.00)	100.00 (+0.16)
CAW	99.63	99.42	99.06	99.56 (-0.07)	99.34 (-0.08)	98.91 (-0.15)
COW	99.42	100.00	99.42	99.71 (+0.29)	100.00 (0.00)	99.71 (+0.29)
CRW	99.90	99.28	99.18	99.90 (0.00)	99.18 (-0.10)	99.08 (-0.10)
CNW	93.97	96.14	90.56	96.73 (+2.76)	96.73 (+0.59)	93.67 (+3.11)
SI	64.86	79.12	55.38	95.50 (+30.64)	96.36 (+17.24)	92.17 (+36.79)
MB	89.71	91.24	82.60	96.89 (+7.18)	97.36 (+6.12)	94.41 (+11.81)
Dro	92.36	92.36	85.81	97.90 (+5.54)	97.90 (+5.54)	95.89 (+10.08)
Pole	90.65	87.30	80.09	97.31 (+6.66)	96.08 (+8.78)	93.60 (+13.51)
Gro	99.69	99.68	99.38	99.88 (+0.19)	99.96 (+0.28)	99.85 (+0.47)
OA (%)	98.99			99.67 (+0.68)		
Kappa (%)	97.52			99.18 (+0.60)		

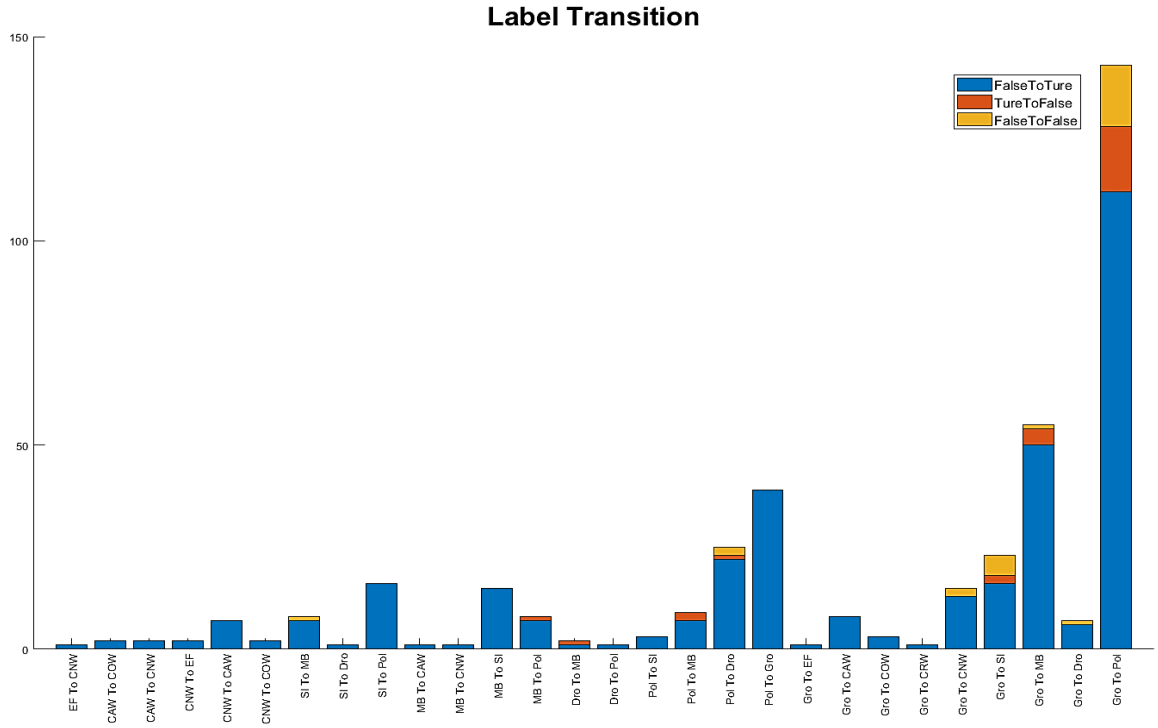


Figure 4.8 Label transition from SVM classification result to HiCRF model. EF (electricity feeder), CAW (catenary wire), COW (contact wire), CRW (current return wire), CNW (connecting wire), SI (suspension insulator), MB (movable bracket), Dro (dropper), Pol (pole) and Gro (ground).

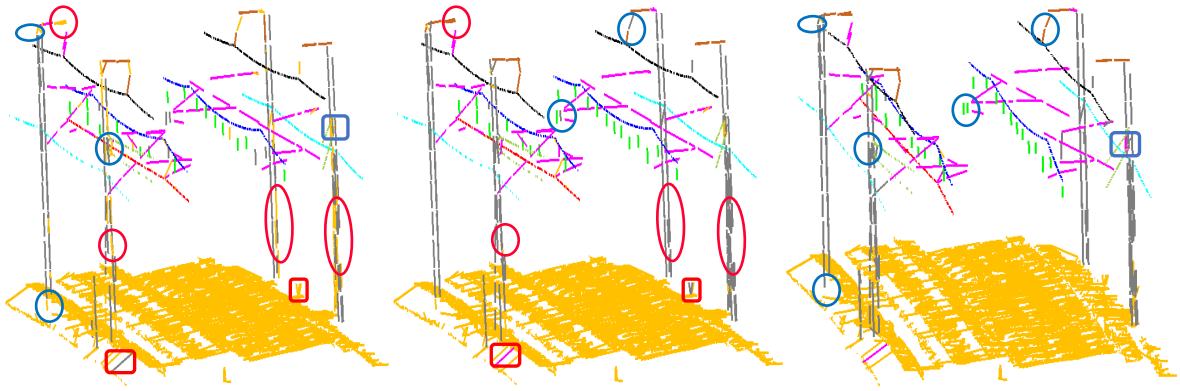
4.6.3 Discussion

This section includes several discussions to investigate how both local smoothness and relative configuration of spatial locations can improve the classification results. Also, this section analyzes the influence of the difference of learnt relative location with different line primitive scales to the classification performance at the bottom and middle layers. Finally, the improvement of classification performance due to relative displacement at arbitrary range in the fully connected top layer will be discussed.

4.6.3.1 Improvement due to local smoothness and relative spatial location

To better understand how the contextual information improves the classification, the visualization of one subset classification result for three different classifiers was shown in Figure 4.9. In the figure, the red circle indicates the misclassification of SVM and corrected by local smoothness in short range, while the blue circle demonstrates the misclassification in SVM result further corrected by learnt relative spatial location in the middle layer. In contrast, the red rectangle shows the wrong classification caused by short range CRF while the blue rectangle means the wrong classification caused by middle range CRF.

From the figure, local smoothness plays a key role to correct the misclassification (pole to ground) which was the most occurred misclassified type in SVM classification result. However, the contrast sensitive Potts model caused the smoothness dependent on the geometric feature of line primitives, which gave rise to some wrong classifications shown as red rectangle. For instance, ground was misclassified as pole because these line primitives were more vertical represented as pole line primitives. This misclassification was lately corrected by middle range spatial location as the height difference between pole and ground.



(a) (b) (c)

Figure 4.9 Comparison of different classification model result in subset 4. The classification results were produced by (a) SVM, (b) CRF with short range and (c) CRF with short and middle range: Each colored circle shows classification errors produced for different object classes.

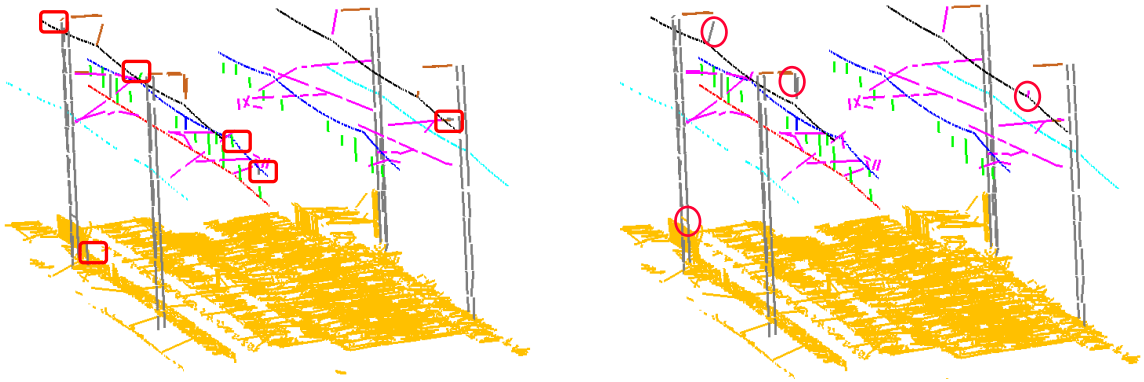
From Figure 4.9 (a) and (c), it is observable that learnt spatial location further improves the classification error which cannot be corrected by short range smoothness pairwise terms. Due to learnt height difference between different objects, suspension insulator and pole misclassified as ground were corrected. Moreover, the misclassification between dropper and pole was corrected due to their difference on horizontal difference to railway. A second example is that suspension insulator misclassified as pole in the short range was also corrected due to horizontal railway distance difference. Nevertheless, connecting wires were misclassified as movable bracket due to their spatial ambiguities between these two objects in local scale. In conclusion, both local smoothness and the learnt relative spatial location can partly correct the misclassification existed in local classifier while it also caused some misclassification due to their geometry and spatial ambiguities.

4.6.3.2 Classification performances at different scales

Table 4.4 presents comparative classification performances produced by BlCRF and TlCRF. Also, a qualitative comparison between BlCRF and TlCRF without fully connected top layer (HiCRF) is illustrated in Figure 4.10. As suggested in Table 4.4, TlCRF produced higher classification performance over most of object classes compared to BlCRF. As shown in this figure, BlCRF shows a tendency to misclassify movable brackets as dropper. It can be observed that the movable bracket-dropper misclassification errors were caused since the spatial relation between these two objects in bottom layer is ambiguous (has a weak discriminative power) because these two objects are smaller and they would be confused in spatial distribution with other smaller line primitives. In Figure 4.10, it can also be noticed that these bracket-dropper misclassification errors were rectified by TlCRF model. However, TlCRF produced other type of misclassification errors with suspension insulators. It is quite understandable because pole and suspension insulator is more horizontally distinguishable with other objects in a smaller scale. These misclassification types made the decrease of the completeness of suspension insulator in the second-layer CRF model, while a significant increase of the dropper's correctness.

Table 4.4 Comparison between BlCRF and TlCRF

	BlCRF			TlCRF		
	Comp.[%]	Corr.[%]	Qual.[%]	Comp.[%]	Corr.[%]	Qual.[%]
EF	100.00	100.00	100.00	100.00 (+0.00)	100.00 (+0.00)	100.00 (+0.00)
CAW	99.63	98.41	98.06	99.78 (+0.15)	98.84 (+0.43)	98.63 (+0.57)
COW	99.71	99.57	99.28	99.71 (+0.00)	99.71 (+0.14)	99.42 (+0.14)
CRW	99.90	99.28	99.18	99.90 (+0.00)	99.28 (+0.00)	99.18 (+0.00)
CNW	92.46	96.84	89.76	94.97 (+2.51)	96.43 (+0.41)	91.75 (+1.99)
SI	95.50	92.98	89.08	91.89 (-3.61)	96.23 (+3.25)	88.70 (-0.38)
MB	91.39	95.50	87.61	93.06 (+1.67)	95.34 (-0.16)	89.02 (+1.41)
Dro	93.71	91.16	85.90	93.71 (+0.00)	96.40 (+5.24)	90.54 (+4.64)
Pole	97.06	94.63	91.99	96.93 (-0.13)	94.74 (+0.11)	91.98 (-0.01)
Gro	99.87	99.94	99.82	99.89 (+0.02)	99.95 (+0.01)	99.84 (+0.02)
OA (%)	99.48			99.54 (+0.06)		
Kappa (%)	98.71			98.88 (+017)		



(a) bottom-layer (b) two-layer

Figure 4.10 The classification result of subset 3: (a) is the result of BlCRF model while (b) shows the result of TlCRF model. Red rectangle means misclassification in BlCRF but corrected in TlCRF and red circle is opposite.

Moreover, all electricity wires improved both the completeness and correctness in TlCRF because their spatial relationship was much more normalized in scale 2 (2nd layer), which caused a more accurate object-based relation. Especially, the completeness of connecting wire was significantly improved. Movable bracket had a better completeness result but produced a slightly worse result in TlCRF due to the correction of misclassification with dropper. Pole almost had same classification results because it has overlapped with many other objects and its vertical structure doesn't change much in these two layers. The observation proved the assumption that same class will have different classification performance at different scales due to disparate spatial distribution.

4.6.3.3 Improvement due to fully-connected top layer

A relative spatial displacement at full range in the top layer should learn more spatial relations at an approximate instance scale. The quantitative evaluation on the improvement

of this layer through the comparison between the classification results obtained by TlCRF and proposed HiCRF model is shown in Table 4.5. Except for a very slight decrease on current return wire, the quality of all other classes improved up to more than 5%. The average results of all evaluation indices increased up to around 2%. The quality of suspension insulator, movable bracket and dropper all clearly improves. This improvement is quite significant considering the classification challenge of these objects because of its variant geometric shape and complex spatial distribution. Also, the classification results of pole and connecting wire were also augmented, especially on the boundary of these two objects. The improvement was credit to the introduced relative displacement at a longer range.

Table 4.5 Comparison between TlCRF and BlCRF

HiCRF - TlCRF		Class										Average (%)
		EF	CAW	COW	CRW	CNW	SI	MB	Dro	Pole	Gro	
Index (%)	Comp.	0.00	-0.22	0.00	0.00	+1.76	+3.60	+3.83	+4.20	+0.38	-0.01	+1.35
	Corr.	0.00	+0.50	+0.29	-0.10	+0.31	+0.14	+2.01	+1.50	+1.34	+0.01	+0.60
	Qual.	0.00	+0.28	+0.29	-0.10	+1.93	+3.48	+5.39	+5.35	+1.62	0.00	+1.82
	OA											+0.12
	Kappa											+0.30

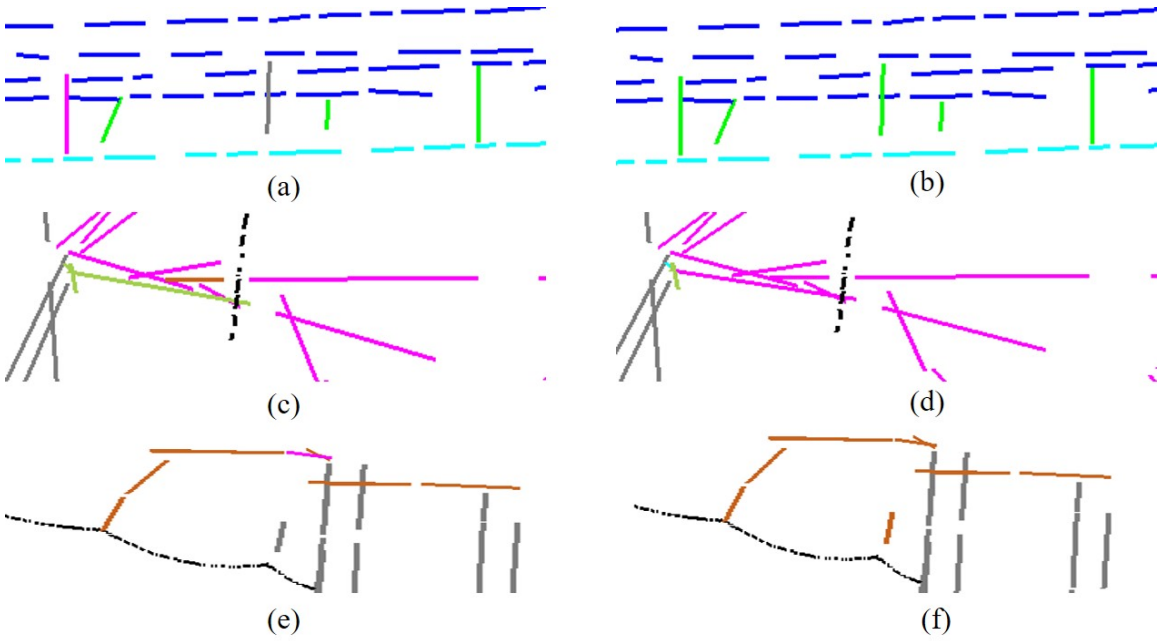


Figure 4.11 Examples of misclassification occurred in TlCRF corrected by HiCRF in subset 5: The left column is the classification of TlCRF while the right column is the corresponding result of HiCRF. (a), (b) are left view and the rest are front view. (**sky blue**: current return wire; **black**: electricity feeder; **blue**: catenary wire; **green**: dropper; **brown**: suspension insulator; **magenta**: movable bracket; **grey**: pole).

The visualized improvement was shown in Figure 4.11. Three pair of improvement illustrated the correctness made by full-connected top layer for dropper, movable bracket and suspension insulator respectively. Dropper usually had some fixed intervals in the whole scene while movable bracket and suspension insulator also appear regularly on the railway direction with its associated poles on the one side of the railway. These spatial regularities can be captured by the encoded relative displacement at full ranges. Combining it with relative location learnt in multi scales, the classification system can better distinguish the label of individual line primitives. The information was particularly

useful to distinguish the line primitives located on the object boundary where most misclassification occurred.

4.7 Chapter Summary

In this chapter, a HiCRF classifier which captures the relative spatial relations among all objects in the full-range of the scene at multiple scales is proposed. To deal with large volume of original point cloud and reduce computational burden, a multi-scale line representation, which was adopted due to a lot linear objects in the scene, is used. The original line representation is generated via fixed voxel generation while a seed-filling based line-merging algorithm was introduced to generate multi-scale line representations considering the geometry of each line primitives.

A local smoothness was applied through a contrast-sensitive Potts model in a short range, which generated a smoother classification result also considered the geometry feature of line primitives. A relative location was considered between two objects based on their horizontal and vertical distance to its corresponding railway vectors in a middle range at different scales. This leaded to a difference spatial relation learning and objects had different classification performance on different scales. Finally, the relative displacement between two line primitives in an approximate instance-level scale was proposed to learn all possible spatial relations at full range in the whole scene. All the information was incorporated into a hierarchical probabilistic graphical model to favor the test data have a similar spatial regularity as training data.

From the classification results, it is noticeable that the proposed contextual information helps the classifier to understand the scene better. The local smoothness and spatial regularities at certain scale significantly improved the classification result. Spatial regularities introduced at multiple scales also produce different classification performance over different objects dependant on the object shape and size. Finally, it is found that the full-range spatial regularities at approximate object-scale can improve the result due to this long-range spatial correction.

Chapter 5
Conclusions and Future Works

This thesis aims to tackle the problem of railway electrification system classification with mobile laser scanning data. Considering the existing strong spatial layout in railway electrification scene, this thesis developed two major methods for this task: (1) proposed a MrCRF model to integrate strong relative and local smoothness in a probabilistic framework to overcome geometry ambiguities problem of conventional supervised local classifier; (2) presented a multi-scale HiCRF model to utilize relative spatial layout at multiple scales and full range in the scene to improve the classification accuracy. In this chapter, I will summarize the research and discuss the conclusions for this problem as well as the potential future work direction.

5.1 Conclusions

In this research, a supervised classification model was proposed to classify railway electrification system objects. The research can be decomposed into two parts, a MrCRF model and a multi-scale HiCRF model.

In Chapter 3, a line-based multi-range Conditional Random Field (MrCRF) model was proposed to combine local smoothness and relative spatial layout in both vertical and horizontal direction. Considering the linearity of electrification system objects and the large volume of mobile laser scanning data, a line representation from fixed size voxel was introduced. Several railway-vector based geometrical features were extracted and a RBF

kernel based SVM classifier was used to produce unary potential in MrCRF model. Local smoothness via Potts model was proposed in short range while one hundred Naïve Bayes classifiers were trained to model spatial layout in middle range. The result showed an average completeness of 97.66% and an average correct of 97.07% for all classes. However, the experiment showed the limitations in representing objects with fixed local scales, which caused problems in characterizing the short-range and middle-range regularities and, thus, led to misclassification errors.

A multi-scale hierarchical Conditional Random Field (HiCRF) model was presented in Chapter 4 to improve the classification result of MrCRF by introducing multi-scale line representation and full range spatial regularities learning in an approximate instance level. A seed-fill based line merging algorithm was proposed to generate a three-scale line representation to consider the scale difference of electrification system objects and generate an approximate instance-level line representation. A three-layer HiCRF model is subsequently described, which the bottom and middle layer was implemented to encode local smoothness and relative location at two different line primitive's scales while the top layer encoded full range displacement vector at an approximate instance level. The experiment showed that proposed model can improve MrCRF classification result on the same railway electrification scene dataset.

The advantage of proposed MrCRF and HiCRF model can both automatically learn the relative spatial regularities from training data. Therefore, they can be mitigated to different datasets with minor modification on some manually set thresholds. In conclusion,

the proposed models achieve a good classification performance which can be extendable for different datasets.

5.2 Future Works

Impressive results on railway electrification system objects classification has been presented in this thesis and the method can be used for automatic inspection task. However, there are some limitations of the current methodology. Firstly, the classification task is separate of railway tracking task, which causes a two-stage processing. Secondly, there are still some complex relative spatial regularities existed in railway electrification scene which cannot be encoded in our model. Finally, we still manually designed some line-based geometric features which can be replaced by automatic feature engineering technologies. To solve these limitations, some innovative ideas can be implemented for the future work.

1. **Unified framework for railway tracking and scene classification**

   Current classification model is based on pre-existing railway tracks extracted from another algorithm (Jwa and Sohn, 2015). All geometrical features are relied on the railway tracking result. However, railway tracking errors will also affect the relative spatial regularities learning, which may degrade classification result. The ideal case is that the railway tracking and the railway electrification scene classification can be conducted simultaneously. Tracked railway vector can establish the relative spatial regularities among electrification system objects while the learnt spatial regulates should also give the cue for railway tracking.

2. **High-order scene regularities**

In the method, all scene layout regularities were encoded as pairwise potential functions in CRF model. In this thesis, only the relations between two line primitives such as “electricity feeder is 1.5 meter higher than catenary wire” or “two adjacent poles are away 80 meters on the railway direction” are modeled. However, there are more complex relative spatial regularities involved more than two objects in railway electrification scene. For example, dropper should connect catenary wire and contact wire. This relation can be modeled as high-order potential functions in the CRF model. This is especially more achievable since the recent introduction of new computational-efficient high-order inference algorithms (Kappes et al., 2013; Kohli and Torr, 2009).

3. **The combination of deep learning and graphical model**

Currently, some hand-crafted features is implemented to encode both local geometric information and contextual information. Since the remarkable success of deep learning algorithms in various computer vision tasks, Convolutional Neural Network (Krizhevsky et al., 2012) became popular for scene understanding task in both image and LiDAR domain. Traditional CNN is limited in learning contextual information which is the advantage of probabilistic graphic model. Therefore, there are some initial exploration (Arnab et al., 2016; Arnab et al., 2018; Zheng et al., 2015) in combining these models. However, these methods are all applied on image semantic segmentation task and it also needs a lot of training data for training which is very hard to get in the practical problem. Therefore, there

are still improvement to combine CNN model and graphical model to learn local and context features automatically for 3D LiDAR data.

4. **Automatic modelling of railway electrification scene objects**

   Classification is the first step towards automatic inspection task. Since the accurate classification result is achieved for all important objects, automatic modeling is the next step to parametrize those objects so the comparison between pre-designed models can be made to determine whether the objects is still effective.

References

Albert L., Rottensteiner F., Heipke C., 2014. A two-layer Conditional Random Field model for simultaneous classification of land cover and land use. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* 40(3), 17.

Albert L., Rottensteiner F., Heipke C., 2017. A higher order conditional random field model for simultaneous classification of land cover and land use. *ISPRS Journal of Photogrammetry and Remote Sensing,* 130, pp 63-80.

Al-Manasi, K., Fraser, C. S., 2006. Registration of terrestrial laser scanner data using imagery. *The Photogrammetric Record*, 21(115), pp. 255-268.

American Public Transportation Association, 2014, 2014 APTA Fact Book. Available online: http://www.apta.com/resources/statistics/Documents/FactBook/2014-APTA-Fact-Book.pdf (accessed on 25th February 2018).

Arastounia M., 2015. Automated recognition of railroad infrastructure in rural areas from LiDARLiDAR data. *Remote Sensing*, 7(11), pp.14916-14938.

Arnab A., Jayasumana S., Zheng S., Torr P. H., 2016. Higher order conditional random fields in deep neural networks. *European Conference on Computer Vision*, pp. 524-540.

Arnab A., Zheng S., Jayasumana S., Romera-Paredes B., Larsson M., Kirillov A., Savchynskyy B., Rother C., Kahl F., Torr P. H., 2018. Conditional Random Fields Meet Deep Neural Networks for Semantic Segmentation: Combining Probabilistic Graphical Models with Deep Learning for Structured Prediction. *IEEE Signal Processing Magazine,* 35(1), pp. 37-52.

Beger R., Gedrange C., Hecht R., Neubert M., 2011. Data fusion of extremely high resolution aerial imagery and LiDARLiDAR data for automated railroad centre line reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6, Supplement), pp. 40-51.

Bishop C. M., 2006. *Pattern recognition and Machine Learning*. Springer.

Bowling M. and Veloso M., 2002. Multiagent learning using a variable learning rate. *Artificial Intelligence,* 136(2), pp.215-250.

Campbell N. D., K. Subr, Kautz J., 2013. Fully-connected CRFs with non-parametric pairwise potential. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, OR, pp. 1658-1665.

Cao X., Zhou F., Xu L., Meng D., Xu Z., Paisley J., 2018. Hyperspectral Image Classification With Markov Random Fields and a Convolutional Neural Network. *IEEE Transactions on Image Processing,* 27(5), pp. 2354-2367.

Chang C.-C. and Lin C.-J., 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST),* 2(3), 27.

Chen L.-C., Papandreou G., Kokkinos I., Murphy K., Yuille A. L., 2018. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence,* 40(4), pp. 834-848.

Cui T., Ji S., Shan J., Gong J., Liu K., 2017. Line-Based Registration of Panoramic Images and LiDARLiDAR Point Clouds for Mobile Mapping. *Sensors*, 17(1), 70.

Desai C., Ramanan D., Fowlkes C. C., 2011. Discriminative models for multi-class object layout. *International journal of computer vision (IJCV),* 95(1), pp. 1-12.

Domke J., 2013. Learning graphical model parameters with approximate marginal inference. *IEEE transactions on pattern analysis and machine intelligence,* 35(10), pp. 2454-2467.

Douillard B., Fox D., Ramos F., 2008. Laser and Vision Based Outdoor Object Mapping. *Proceedings of Robotics: Science and Systems IV*, Zurich, Switzerland.

Eitz, M., Hildebrand, K., Boubekeur, T., Alexa, M., 2011. Sketch-based image retrieval: Benchmark and bag-of-features descriptors. *IEEE transactions on visualization and computer graphics*, 17(11), pp. 1624-1636.

Feng H., Jiang Z., Xie F., Yang P., Shi J., Chen L., 2014. Automatic Fastener Classification and Defect Detection in Vision-Based Railway Inspection Systems. *IEEE Transactions on Instrumentation and Measurement,* 63(4), pp. 877-888.

Gálai B., Nagy B., Benedek C., 2016. Crossmodal point cloud registration in the Hough space for mobile laser scanning data. *23rd International Conference on Pattern Recognition (ICPR)*, Cancun, Mexico, pp. 3374-3379.

Galleguillos C. and Belongie S., 2010. Context based object categorization: A critical survey. *Computer vision and image understanding,* 114(6), pp. 712-722.

Gao Y., Huang X., Zhang F., Fu Z., Yang C., 2015. Automatic Geo-referencing Mobile Laser Scanning Data to UAV images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* 40(1), pp. 41-46.

Gauvain J.-L. and Lee C.-H., 1994. Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. *IEEE transactions on speech and audio processing,* 2(2), pp. 291-298.

Golovinskiy A., Kim V. G., Funkhouser T., 2009. Shape-based recognition of 3D point clouds in urban environments. *12th IEEE International Conference on Computer Vision*, Kyoto, Japan, pp. 2154-2161.

Gómez-Expósito A., Mauricio J. M., Maza-Ortega J. M., 2014. VSC-based MVDC railway electrification system. *IEEE Transactions on Power Delivery,* 29(1), pp. 422-431.

Gould S., Rodgers J., Cohen D., Elidan G., Koller D., 2008. Multi-Class Segmentation with Relative Location Prior. *International Journal of Computer Vision,* 80(3), pp. 300-316.

Gressin A., Mallet C., Demantké J., David N., 2013. Towards 3D lidarLiDAR point cloud registration improvement using optimal neighborhood knowledge. *ISPRS Journal of Photogrammetry and Remote Sensing*, 79(5), pp. 240-251.

Grilli E., Menna F., Remondino F., 2017. A review of point clouds segmentation and classification algorithms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* 42(2), pp. 339-344.

Grinias I., Panagiotakis C., Tziritas G., 2016. MRF-based segmentation and unsupervised classification for building and road detection in peri-urban areas of high-resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing,* 122, pp. 145-166.

Guo B., Huang X., Li Q., Zhang F., Zhu J., Wang C., 2016. A Stochastic Geometry Method for Pylon Reconstruction from Airborne LiDARLiDAR Data. *Remote Sensing,* 8(3), 243.

Hammersley J. M. and Clifford P., 1971. *Markov fields on finite graphs and lattices*. Unknown Publisher.

He X., Zemel R. S., Carreira-Perpiñán M. Á., 2004. Multiscale conditional random fields for image labeling. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, pp. 695-702.

Innoteka, 2012. The Doyen of Railway Electrification in Hungary. Available online: http://www.innoteka.hu/cikk/the_doyen_of_railway_electrification_in_hungary.487.html (accessed on 3rd March 2018).

Jung J., Chen L., Sohn G., Luo C., Won J.-U., 2016. Multi-Range Conditional Random Field for Classifying Railway Electrification System Objects Using Mobile Laser Scanning Data. *Remote Sensing,* 8(12), 1008.

Jwa Y. and Sohn G., 2015. Kalman filter based railway tracking from mobile LidarLiDAR data. *Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, La Grande Motte, France, pp. 159-164.

Kappes J. H., Andres B., Hamprecht F. A., Schnörr C., Nowozin S., Batra D., Kim S., Kausler B. X., Lellmann J., Komodakis N., 2013. A comparative study of modern inference techniques for discrete energy minimization problems. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, pp. 1328-1335.

Kohli P. and Torr P.H., 2009. Robust higher order potentials for enforcing label consistency. *International Journal of Computer Vision,* 82(3), pp. 302-324.

Koltun V., 2011. Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in Neural Information Processing Systems*, Granada, Spain, pp. 109-117.

Koppula H. S., Anand A., Joachims T., Saxena A., 2011. Semantic labeling of 3d point clouds for indoor scenes. *Advances in Neural Information Processing Systems*, Granada, Spain, pp. 244-252.

Krizhevsky A., Sutskever I., Hinton G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, Lake Tahoe, NV, pp. 1097-1105.

Kumar S. and Hebert M., 2006. Discriminative random fields. *International Journal of Computer Vision,* 68(2), pp. 179-201.

Ladicky L., Russell C., Kohli P., Torr P.H., 2010. Graph cut based inference with co-occurrence statistics. *European Conference on Computer Vision (ECCV)*, Berlin, Heidelberg, pp. 239-253.

Lafarge F. and Mallet C., 2012. Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation. *International journal of computer vision,* 99(1), pp. 69-85.

Lafferty J., McCallum A., Pereira F. C., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *Proceeding of International Conference on Machine Learning,* Williamstown, MA, pp. 282-289.

Li J., Yang B., Chen C., Huang R., Dong Z., Xiao W., 2018. Automatic registration of panoramic image sequence and mobile laser scanning data using semantic features. *ISPRS Journal of Photogrammetry and Remote Sensing,* 136, pp. 41-57.

Li L., Socher -J.,R., Fei-Fei L., 2009. Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, pp. 1328-1335.

Lim E. H. and Suter D., 2007. Conditional random field for 3D point clouds with adaptive data reduction. *International Conference on Cyberworlds (CW'07)*, Hannover, Germany, pp.404-408.

Lim E. H. and Suter D., 2008. Multi-scale conditional random fields for over-segmented irregular 3D point clouds classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, pp.1-7.

Lim E. H. and Suter D., 2009. 3D terrestrial LIDARLIDAR classifications with super-voxels and multi-scale Conditional Random Fields. *Computer-Aided Design*, 41(10), pp. 701-710.

Lodha S. K., Kreps E. J., Helmbold D. P., Fitzpatrick D. N., 2006. Aerial LiDARLiDAR Data Classification Using Support Vector Machines (SVM). *Proceeding of the 4th International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT)*, Chapel Hill, NC, pp. 567-574.

Lu W.-L., Murphy K. P., Little J. J., Sheffer A., Fu H., 2009. A hybrid conditional random field for estimating the underlying ground surface from airborne lidarLiDAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8), pp. 2913-2922.

Luo C. and Sohn G., 2014. Scene-layout compatible conditional random field for classifying terrestrial laser point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 2(3), pp. 79-86.

Marino F., Distante A., Mazzeo P. L., Stella E., 2007. A Real-Time Visual Inspection System for Railway Maintenance: Automatic Hexagonal-Headed Bolts Detection. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* 37(3), pp. 418-428.

Men H., Gebre B., Pochiraju K., 2011. Color point cloud registration with 4D ICP algorithm. *IEEE International Conference on Robotics and Automation*. Shanghai, China, pp. 1511-1516.

Mohamad M., Kusevic K., Mrstik P., Greenspan M., 2013. Automatic rail extraction in terrestrial and airborne LiDARLiDAR data. *2013 International Conference on 3D vision*, Seattle, WA, pp. 303-309.

Murphy K. P., Weiss Y.and Jordan M. I., 1999. Loopy belief propagation for approximate inference: An empirical study. *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, Stockholm, Sweden, pp. 467-475.

Najafi M., Namin S. T., Salzmann M., Petersson L., 2014. Non-associative higher-order markov networks for point cloud classification. *European Conference on Computer Vision*, Amsterdam, Netherlands, pp. 500-515.

Niemeyer J., Rottensteiner F., Soergel U., 2012. Conditional random fields for lidarLiDAR point cloud classification in complex urban areas. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3, pp. 263-268.

Niemeyer J., Rottensteiner F., Soergel U., 2014. Contextual classification of lidarLiDAR data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing,* 87, pp. 152-165.

Niemeyer J., Rottensteiner F., Soergel U., Heipkea C., 2016. Hierarchical Higher Order Crf for the Classification of Airborne LIDARLIDAR Point Clouds in Urban Areas. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 41, pp. 655-662.

Nurunnabi A., Belton D., West G., 2016. Robust Segmentation for Large Volumes of Laser Scanning Three-Dimensional Point Cloud Data. *IEEE Transactions on Geoscience and Remote Sensing,* 54(8), pp. 4790-4805.

Pastucha E., 2016. Catenary System Detection, Localization and Classification Using Mobile Scanning Data. *Remote Sensing,* 8(10), 801.

Pearl J., 2014. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.

Qin R. and Gruen A., 2014. 3D change detection at street level using mobile laser scanning point clouds and terrestrial images. ISPRS Journal of Photogrammetry and Remote Sensing, 90, pp. 23-35.

Railway Gazette, 2016. Chinese high speed network to double in latest master plan. Available online: http://www.railwaygazette.com/news/infrastructure/single-view/view/chinese-high-speed-network-to-double-in-latest-master-plan.html (accessed on 25th February 2018).

Roig G., Boix X., Shitrit H. B., Fua P., 2011. Conditional random fields for multi-camera object detection. *IEEE International Conference on Computer Vision (ICCV)*, Barcelona, Spain, pp. 563-570.

Rutzinger M., Rottensteiner F., Pfeifer N., 2009. A Comparison of Evaluation Techniques for Building Extraction From Airborne Laser Scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 2(1), pp. 11-20.

Sacchi C., Regazzoni C., Vernazza G., 2001. A neural network-based image processing system for detection of vandal acts in unmanned railway environments. *Proceedings 11th International Conference on Image Analysis and Processing*, Palermo, Italy, pp. 529-534.

Samadzadegan F., Bigdeli B., Ramzi P., 2010. A multiple classifier system for classification of LIDARLIDAR remote sensing data using multi-class SVM. *International Workshop on Multiple Classifier Systems*, Berlin, Heidelberg, pp. 254-263.

Schindler K., 2012. An Overview and Comparison of Smooth Labeling Methods for Land-Cover Classification. *IEEE Transactions on Geoscience and Remote Sensing,* 50(11), pp. 4534-4545.

Shapovalov R. and Velizhev A., 2011. Cutting-plane training of non-associative markov network for 3d point cloud segmentation. *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, Hangzhou, China, pp. 1-8.

Shotton J., Winn J., Rother C., Criminisi A., 2009. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, 81(1), pp. 2-23.

Strat T. M. and Fischler M. A., 1991. Context-based vision: recognizing objects using information from both 2 d and 3 d imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 13(10), pp. 1050-1065.

Sun L., Wu Z., Liu J., Xiao L., Wei Z., 2015. Supervised spectral–spatial hyperspectral image classification with weighted Markov random fields. *IEEE Transactions on Geoscience and Remote Sensing,* 53(3), pp. 1490-1503.

Toyoda T. and Hasegawa O., 2008. Random field model for integration of local information and global information. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 30(8), pp. 1483-1489.

Triebel R., Kersting K., Burgard W., 2006. Robust 3D scan point classification using associative Markov networks. Robotics and Automation. *Proceedings of IEEE International Conference on Robotic Automation*. Orlando, FL, pp. 2603-2608.

Vishwanathan S., Schraudolph N. N., Schmidt M. W., Murphy K. P., 2006. Accelerated training of conditional random fields with stochastic gradient methods. *Proceedings of the 23rd international conference on Machine learning*, Pittsburgh, PA, pp. 969-976.

Volpi M. and Ferrari V., 2015. Structured prediction for urban scene semantic segmentation with geographic context. *Urban Remote Sensing Event (JURSE)*, Lausanne, Switzerland, pp. 1-4.

Vosselman G., Coenen M., Rottensteiner F., 2017. Contextual segment-based classification of airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing,* 128, pp. 354-371.

Wang P., Shen C., van den Hengel A., 2015. Efficient SDP inference for fully-connected CRFs based on low-rank decomposition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, pp. 3222-3231.

Wegner J. D., Montoya-Zegarra J. A., Schindler K., 2013. A higher-order CRF model for road network extraction. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, pp. 1698-1705.

Winn J. and Shotton J., 2006. The layout consistent random field for recognizing and segmenting partially occluded objects. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, New York, NY, pp. 37-44.

Xiao W., Vallet B., Brédif M., Paparoditis N., 2015. Street environment change detection from mobile laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing,* 107, pp. 38-49.

Xiao W., Vallet B., Schindler K., Paparoditis N., 2016. Street-side vehicle detection, classification and change detection using mobile laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing,* 114, pp. 166-178.

Xiong X., Munoz D., Bagnell J. A., Hebert M., 2011. 3-d scene analysis via sequenced predictions over points and regions. *2011 IEEE International Conference on Robotics and Automation (ICRA)*, Orlando, FL, pp. 2609-2616.

Yang, C., Sharon, D., van de Panne, M. 2005. Sketch-based Modeling of Parameterized Objects. *2nd Eurographics Workshop on Sketch-Based Interfaces and Modeling,* Dublin, Ireland, pp. 1-10.

Yang M. Y. and Förstner W., 2011. A hierarchical conditional random field model for labeling and classifying images of man-made scenes. *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Barcelona, Spain, pp. 196-203.

Yao J., Fidler S., Urtasun R., 2012. Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, pp. 702-709.

Yoon, S. M., Scherer, M., Schreck, T., Kuijper, A., 2010. Sketch-based 3D model retrieval using diffusion tensor fields of suggestive contours. *Proceedings of the 18th ACM international conference on Multimedia*, Firenze, Italy, pp. 193-200.

Zhang S., Wang C., Yang Z., Chen Y., Li J., 2016. Automatic railway power line extraction using mobile laser scanning data. *Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Prague, Czech Republic, pp. 615-619.

Zhang Y. and Chen T., 2012. Efficient inference for fully-connected CRFs with stationarity. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, pp. 582-589.

Zheng S., Jayasumana S., Romera-Paredes B., Vineet V., Su Z., Du D., Huang C., Torr P.H., 2015. Conditional random fields as recurrent neural networks. *Proceedings of the IEEE International Conference on Computer Vision*, Las Condes, Chile, pp. 1529-1537.

Zhou Y., Wang D., Xie X., Ren Y., Li G., Deng Y., Wang Z., 2014. A Fast and Accurate Segmentation Method for Ordered LiDARLiDAR Point Cloud of Large-Scale Scenes. IEEE Geoscience and Remote Sensing Letters, 11(11), pp. 1981-1985.

Zhu L. and Hyyppa J., 2014. The use of airborne and mobile laser scanning for modeling railway environments in 3D. *Remote Sensing,* 6(4), pp. 3075-3100.

Zhu Q., Li Y., Hu H., Wu B., 2017. Robust point cloud classification based on multi-level semantic relationships for urban scenes. *ISPRS Journal of Photogrammetry and Remote Sensing,* 129, pp. 86-102.