

## SCENE-LAYOUT COMPATIBLE CONDITIONAL RANDOM FIELD FOR CLASSIFYING TERRESTRIAL LASER POINT CLOUDS

Chao Luo<sup>a,\*</sup>, Gunho Sohn<sup>a</sup>

<sup>a</sup>GeoICT Laboratory, Department of Earth and Space Science, York University, 4700 Keele Street,  
Toronto, Ontario, Canada M3J 1P3 – (cluo, gsohn)@yorku.ca

Commission III, WG III/2

**KEY WORDS:** Conditional Random Field, Scene Layout, Asymmetric, Terrestrial Laser Scanning, Classification

### ABSTRACT:

Terrestrial Laser Scanning (TLS) rapidly becomes a primary surveying tool due to its fast acquisition of highly dense three-dimensional point clouds. For fully utilizing its benefits, developing a robust method to classify many objects of interests from huge amounts of laser point clouds is urgently required. Conditional Random Field (CRF) is a well-known discriminative classifier, which integrates local appearance of the observation (laser point) with spatial interactions among its neighbouring points in classification process. Typical CRFs employ generic label consistency using short-range dependency only, which often causes locality problem. In this paper, we present a multi-range and asymmetric Conditional Random Field (CRF) (maCRF), which adopts *a priori* information of scene-layout compatibility addressing long-range dependency. The proposed CRF constructs two graphical models, one for enhancing a local labelling smoothness within short-range (srCRF) and the other for favouring a global and asymmetric regularity of spatial arrangement between different object classes within long-range (lrCRF). This maCRF classifier assumes two graphical models (srCRF and lrCRF) are independent of each other. Final labelling decision was accomplished by probabilistically combining prediction results obtained from two CRF models. We validated maCRF's performance with TLS point clouds acquired from RIEGL LMS-Z390i scanner using cross validation. Experiment results demonstrate that synergetic classification improvement can be achievable by incorporating two CRF models.

### 1. INTRODUCTION

Terrestrial Laser Scanning (TLS) is a relatively new surveying tool, but has been rapidly adopted for modelling urban street scenes. TLS captures a scene by directing laser pulses in various directions at a constant scanning rate. Thus, resulting laser point clouds are collected as a horizontal or vertical line profile per scan angle. One of the main applications of TLS is to assess “as-is” condition of key infrastructures in urban environment and identify their potential risks by comparing it with “as-built” condition. For rapidly responding to risks, it is urgently required to develop fully automated and robust classification algorithm. However, classifying such complex urban scenes from huge amount of TLS point clouds still remains as a challenging vision task.

Classification is the problem of identifying corresponding object class that belongs to an “entity” (e.g., point, line and plane in laser point space) with given observations (“features”). A typical approach to the classification problem is to identify a target object by only relying on local apparent features differentiating the object from the others. However, this naïve method often causes classification errors due to ambiguities in appearance among classes in varying vision conditions. To address this problem, it has been well adopted to consider imposing contextual constraint between adjacent entities on the classification problem. Conditional Random Field (CRF) is a well-known statistical modelling method that enables directly modelling of the posterior probability of object classes for each entity with given global observations. CRF formulates not only dependencies between the entities and their features, but also relational context between adjacent entities in an associated graphical model. CRF often uses the smoothness prior as the

relational context to maximize the label homogeneity between adjacent entities. However, this “smoothness-only” relational encoding scheme is prone to produce classification errors, especially over heterogeneous objects with similar appearances.

In this study, we exploit another type of the relational constraint, that is, the regularity of spatial arrangement between adjacent objects in CRF model. The regularity of spatial arrangement is a global prior knowledge on scene-layout among key urban objects. For instance, the pedestrian roads are usually placed below building façades or trees, while trees or lamp post are placed closer to vehicle roads than building façades. Our CRF model (maCRF) incorporates such scene-layout compatible priors into interaction potential for regulating a configuration of spatial arrangement in pairwise object relation. In particular, we adopt an asymmetric dependency capturing directional spatial arrangement between pairwise objects (e.g. it allows ground is lower than building, not vice-versa).

Typical CRF only works locally as it emphasizes on local label consistency, not capturing its global properties. To overcome this locality problem, we construct two graphical models, one for enhancing a local consistency over labels connected within short-range dense graphical model (srCRF) and the other for favouring a global regularity of spatial arrangement between different object classes that is structured within a sparse long-range graphical model (lrCRF). A label prediction was conducted using two graphical models separately. However, final classification was accomplished by combining their label prediction results. This ensemble classification system synergistically compensates the limitations caused by srCRF (“over-smoothing”) and lrCRF (“under-smoothing”) respectively.

\* Corresponding author.

An ultimate goal of the proposed method is to process laser point clouds scan-by-scan following TLS's natural scanning pattern. We selected the line primitive as the entity for constructing maCRF. Thus, a set of line segments and its feature vectors were extracted from each scan profile acquired by TLS. The graphical models (srCRF and lrCRF) were constructed on extracted line segments per each scan profile. The maCRF combined the label decisions from two graphical models and finally identifies corresponding labels to individual line segments and thus to their member points.

The paper outlines following sections: Section 2 introduces previous works related to context-based classification. In Section 3, we describe details of data pre-processing, such as line segmentation per scan profile and feature extraction. Then, we present the methodology for implementing maCRF in Section 4 and discuss our experimental results in Section 5. Finally, we draw our conclusions and give an outlook of future works in Section 6.

## 2. RELATED WORK

Classification methods can be categorised into three different groups according to the entity used for constructing the spatial structure: point-based (Triebel, et al, 2006; Munoz et al, 2008), line-based (Zhao et al., 2010; Hu and Ye, 2013) and surface-based (Belton and Lichti, 2007; Pu and Vosselman, 2009). Point based method operates classification directly on individual laser points using feature vector extracted from its local neighbours. While, surface-based approach firstly segments the laser point clouds into homogeneous surface segments and then label them based on their feature vectors. Both point-based and surface-based classification methods are typically implemented in 3D volumetric space. This typically requires computationally expensive process for constructing relational network from massive individual points or segmenting large amount of laser point clouds into surfaces.

Besides the points and surfaces, the appearance of objects can be also characterized using lines. It is rather straightforward to extract lines and construct relational graph per scan line. This advantage brings the benefits to improve computational efficiency (Jiang and Bunke, 1994). Much research efforts have been made to use the line primitives for classification purpose. Sithole and Vosselman (2003) detected urban objects by segmenting laser point clouds into two families of orthogonal profiles running along x and y direction and interpreting their properties. Similarly Hu and Ye (2013) used Douglas-Peucker algorithm to segment laser scan profile into a set of line segments and classified them into buildings and vegetation based on a rule-based local analysis of line attributes. Zhao et al. (2010) collected data using single-row laser scanner segment and extracted line segments within each scan line for characterizing building and roads. Then these line segments were classified using Markov network. In this study, we used a line extraction algorithm similar to (Manandhar and Shibasaki, 2001). Firstly, entire TLS data was split into a set of vertical scan profiles. Then each scan profile was decomposed into a set of line segments based on range analysis. The classification task finally labels these line segments.

Local classifiers label individual entity by only considering its apparent features. This approach is prone to produce classification errors due to the ambiguities in the appearance of the laser data. Contextual classifier can compensate the limitation of local classifiers by introducing label interaction among the neighboring entities. Markov Random Fields (MRF)

is one of the most popular graphical models to incorporate local contextual constraints into labeling problems (Cross and Jain, 1983). It has been used to exploit relationships among laser points and been proved to be able to make smoother and more consistent classification result (Angelov et al., 2005). However, with the assumption of conditional independence of the data, the interaction term between each pair of neighboring nodes in MRF is restricted to only the class labels and thus it is likely to produce over-smooth classification result. By capturing spatial dependencies in labels and observed data simultaneously, MRF's limitation can be overcome by Conditional Random Fields (Kumar and Hebert, 2006). CRFs provide discriminative framework that allows interaction potential term conditioned on class label as well as global observations data.

Recently, many works on classifying laser point clouds using CRFs have been reported. Munoz et al. (2008) improved the results of associate MRF by utilizing accurate pairwise potentials. Shapovalov et al. (2010) and Niemeyer et al. (2014) demonstrated that CRF can improve classification results over urban areas by augmenting pairwise label consistency between adjacent data using airborne laser point clouds. However, those CRFs still could mislead over-smoothed labelling configuration since they rely only on short-range context. Targeting on this problem, He et al. (2004) proposed a multi-scale CRF model, which encodes long-range dependency by applying pairwise relationships at multi-scale image data. They combined local classifier with contextual information retrieved from local region to global scale (entire image) into a single probabilistic model. Lim and Suter (2009) presented CRF for classifying laser point clouds using multi-scale super-voxels by enhancing label consistency not only within each voxel (local), but also across neighbouring voxels (global). Yang and Förstner (2011) presented a hierarchical CRF, in which label inference was conducted across segmented imagery generated at different scales using multi-scale mean shift algorithm.

In addition to the role of context, spatial arrangement among object components adds a significant contribution to the perception of the multi-object classification (Bar and Ullman, 1996). With information about the probable location of objects in the scene and relative location relations among objects, the classifier with spatial arrangement helps decrease the risk of misclassification. Gould, et al. (2008) directly modelled spatial relationship by learning the relative locations between classes from a set of labelled images. The inter-class spatial relationship was modelled using non-parametric relative location maps. In our study, we exploited a vertical spatial arrangement between adjacent objects (i.e., "above-and-below" relation) captured in long-range context. This implicit regularity in spatial arrangement was modelled as asymmetric pairwise interaction in a probabilistic manner.

## 3. LINE-BASED FEATURE EXTRACTION

### 3.1 Line Segment Extraction

Our ultimate goal is to apply the proposed method to a real-time application, in which classification result is generated per scanned line profile. Thus, prior to the primitive extraction, we split the entire TLS data into a set of vertical scan profiles. This task was simply achieved using *a priori* knowledge on TLS's scanning angle precision (i.e., 0.05 degree). The interval between vertical scan profiles is 0.05 degree at azimuth angle.

Most of urban objects have apparent shapes which are well characterized with line primitives. Once the scan profiles were

detected, line segments were extracted per scan profile as the spatial entity used for classification process. Similar to the work suggested by Manandhar and Shibasaki (2001), line segments were extracted per scan profile based on range analysis as the spatial entity used for classification process. One can find more detailed description of scan profile generation and line segmentation algorithm used in our previous study (Luo and Sohn, 2013). Figure 1(a) shows laser points in a vertical scan profile and Figure 1(b) shows the result of line segments extraction.

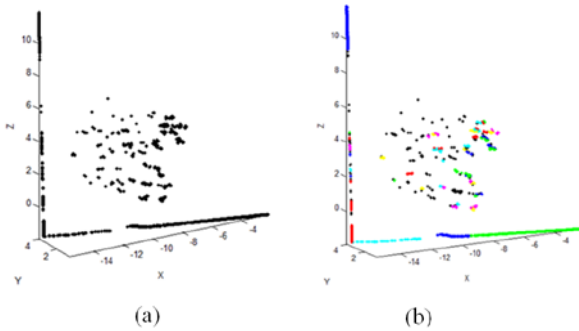


Figure 1. Line segment extraction: (a) laser point in scan profile (b) line segment extraction result.

### 3.2 Feature Extraction

Once the line segments extraction was conducted, seven local features were extracted for each line segment. The features include: 1) maximum height ( $z$ ); 2) minimum height ( $z$ ); 3) mean height ( $z$ ). In addition, all the points belong to the same line segment were fit into one straight line using linear regression. And the following four features are extracted based on the fitted line. 4) length (maximum extension in the major direction); 5) mean absolute residual; 6) standard deviation of residual; and 7) orientation (angle between the fitted line and  $z$  axis). The extracted features characterize local appearance of a line segment.

While, context features characterize grouping property of a line segment of interest and its surrounding neighbours. In our study, two neighbouring systems were used for extracting context features; circle-based (Figure 2(a)) and vertical column-based neighbouring system (Figure 2(b)). As illustrated in Figure 2(a), a circle with 1m radius was created at the centre of a line centroid (black dot). Lines which centroids fall inside the circle (both red and pink dots in Figure 2(a)) are considered as neighbours of the current line of interest. Moreover, further constraint on orientation similarity makes another type of neighbour: neighbouring lines with similar orientation (red dots in Figure 2(a)). And ones with dissimilar orientation (pink dots in Figure 2(a)) were excluded. Here we set the orientation similarity constraint as that included angle between two lines is smaller than 30 degree.

Figure 2(b) shows our vertical column-based neighbouring system. A scan profile was discretized into a set of non-overlapping vertical subspaces (rectangle area between dotted blue lines) with 0.5m width in Figure 2(b). Then, neighbouring lines were searched within a vertical column (blue filled area) where current line falls into. Similar to circle-based neighbouring system, context features with or without considering orientation similarity were extracted.

Once two neighbouring systems were generated for a line segment, seven context features were obtained. These include:

1) maximum  $z$ ; 2) sum of line length. Points belong to the line and its surrounding neighbours were fitted into one straight line and the following three features are extracted from the fitted line: 3) mean residual; 4) standard deviation; 5) orientation (angle between the fitted line and  $z$  axis). The other two features are: 6) point density (point number in the line group) and 7) line density (line number in the line group). Therefore, four types of neighbourhood systems produced total 28 context features.

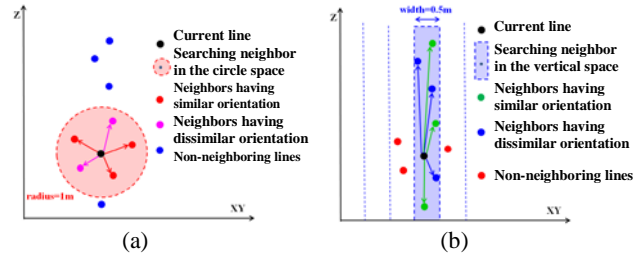


Figure 2. Neighborhood selection for context feature.

### 3.3 Feature Dimension Reduction

Total thirty-five features (7 appearance features and 28 context features) were extracted as described in the previous section. However, high-dimensional feature often causes negative effects for classification such as producing highly correlated features and high computational complexity. To avoid these problems, we applied principle component analysis (PCA) to reduce the feature dimensionality. By adopting the cumulative energy (90%) criteria proposed by Krzanowski (2000), we finally we transformed features with the eight largest eigenvalues selected for classification.

## 4. MULTISCALE ASYMMETRIC CONDITIONAL RANDOM FIELD

We propose a multi-range and asymmetric Conditional Random Field (maCRF) model for classifying TLS point clouds. maCRF is developed to incorporate three primary aspects of contextual inference in a graphical model: 1) maximising the smoothness between short-range nodes, 2) maximizing the regularity of spatial dependency between objects in long range nodes and 3) considering asymmetric properties of spatial arrangement regularity. In this study, it is assumed that two graphical models (short-range and long-range) are independent of each other. Thus, label predictions from two CRF models were separately performed and final conditional probability was accomplished by combining their prediction results.

In this research, it is noted that the classification was implemented per scan profile and so relationship across scale profiles was not considered yet. Supposing that a certain scan line has  $N$  line segments and  $X = \{x_1, x_2, \dots, x_N\}$  are feature vectors extracted for each line segment. We aim to classify seven object classes, which are defined as hidden variable  $Y = \{\text{roof, building, vehicle road (VR), pedestrian road (PR), tree, low vegetation (LV), low man-made objects (LMO)}\}$ . maCRF model is constructed by combining short range CRF(srCRF) and long range CRF(lrCRF), as shown in Figure3. The mathematical model is given in Equation1.

$$P_M(Y|X) = P_S(Y|X)P_L(Y|X) \quad (1)$$

where  $P_M(Y|X)$ ,  $P_S(Y|X)$  and  $P_L(Y|X)$ , represent the conditional probability of maCRF, srCRF and lrCRF respectively.

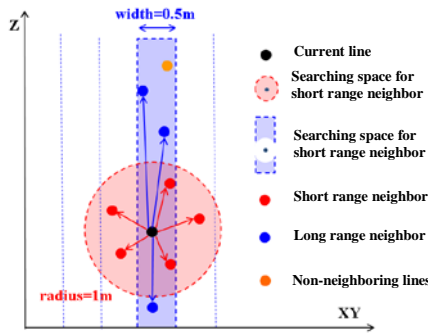


Figure 3. maCRF with short (red) and long (blue) edges.

## 4.1 Short-range CRF model

### 4.1.1 Graphical Model

Let  $G_S = (V, E_S)$  be a short range graph, each of which node,  $v \in V$  represents a line segment (centroid of a line) extracted from one scan profile. Its node adjacency relation,  $e_S \in E_S$  was constructed between two nodes if their distance is smaller than certain distance threshold (e.g., 1 m). A red edge shown in Figure 3 is considered as short-range edge. It is noted that, contrast to a graph model represented in image space, our line-based graph does not follow a regular grid pattern.

Given the fundamental theorem of random fields, the conditional distribution over the labels  $Y$  given observed data  $X$  in the graph  $G_S$  is defined in Equation 2.

$$P_S(Y | X) = \frac{1}{Z_S(X)} \exp\left(\sum_{i \in V} A_i(X, y_i) + \sum_{(i,j) \in E_S} S_{ij}(X, y_i, y_j)\right) \quad (2)$$

Where  $A_i(X, y_i)$  is the association potential, which measures how likely class label  $y_i$  is assigned to a single node  $i$  given global observations  $X$  without considering a relational regularity (interaction) with other nodes.  $S_{ij}(y_i, y_j)$  is the short range potential and measures how the labels at neighbouring nodes ( $y_i, y_j$ ) interact given the observation  $X$ .  $Z_S$  is the normalization term (partition function), which is always computed using a forward-backward algorithm.

### 4.1.2 Association Term

The associate potential term in Equation 2 corresponds to a log posterior probability. Theoretically, the posterior probability of any local classifier can be used, and here we use Gaussian Mixture Model-Expectation Maximization (GMM-EM) classifier:

$$A_i(X, y_i) = \log(P(y_i | x_i)) \propto \log(P(x_i | y_i)) \quad (3)$$

where  $i$  indicates a line segment. Since we assume that each class has equal prior probability, posterior probability  $P(y_i | x_i)$  is proportional with class-conditional probability  $P(x_i | y_i)$ . Due to the complexity of urban objects, the actual probability density function is a multimodal and fails to be fit well with single Gaussian model. Therefore, the mixture Gaussian approximation is a quite reasonable method to model likelihood, which is expressed as follows:

$$P(x_i | y_i) = \sum_{m=1}^M \alpha_m N(x_i; \mu_m, \Sigma_m) \quad (4)$$

where  $N(x_i; \mu_m, \Sigma_m)$  is the Gaussian mixture component,  $\alpha_m$  is the corresponding weight, and  $M$  indicates the number of mixture components. The value of  $\alpha_m$  ranges from 0 to 1 for all components, and the sum of  $\alpha_m$  equals 1. The parameters  $\{\alpha_1, \dots, \alpha_m, \mu_1, \dots, \mu_m, \Sigma_1, \dots, \Sigma_m\}$  define the Gaussian mixture probability density function. To estimate the parameters of the Gaussians Mixture model, the classic Expectation Maximization (EM) algorithm was used. In this research, we use a uniform component value of three for all class-conditional probabilities.

### 4.1.3 Short-range Interaction Term

The short-range interaction term enforces a smoothness constraint to local neighbourhood. The homogeneous Ising model (Potts model for multiclass labelling problem) proposed by (Cross and Jain, 1983) is a popular way to model this interaction term. However, it only enforces a constant smoothing of labels and fails to capture dependencies based on observations. Thus, we adopted a generalized linear model (GLM) to formulate the short-range interaction potential. For each edge connecting two nodes  $i$  and  $j$ , the interaction potential is expressed with a generalized linear function over edge feature vector  $\mu_{ij}(x)$  as below:

$$S_{ij}(x, y_i = l, y_j = k) = w_{l,k}^T \mu_{ij}(x) \quad (5)$$

In Equation 5, the interaction energy was designed to encourage right configuration of ( $y_i, y_j$ ) and penalize the wrong label configuration. The degree of penalization depends on the edge feature vector  $\mu_{ij}$  and weight vector  $w$ , which is learned from training sample. It is assumed that objects have similar node features are more likely to have the same label. Here, the edge features  $\mu_{ij}(x)$  is generated by subtracting node features of two neighbouring lines ( $\mu_{ij}(x) = x_i - x_j$ ).

Standard CRFs assume that clique potential functions are symmetric and do not change with the position of a node on the edge. Therefore, it often fails to model directional compatibility between labels. For example, symmetric potential can model the relationship that roof and building are adjacent but fail to model the directional relationship that roof is on the top of building when they are adjacent. The asymmetric interaction potential used here can model this kind of directional relationship.

We can see from Equation 5 that interaction potential design makes the potential flow from  $i$  to  $j$  different with that following from opposite direction  $S_{ij}(x, y_i = l, y_j = k) \neq S_{ji}(x, y_j = l, y_i = k)$ , because  $w_{l,k}^T (x_i - x_j) \neq w_{l,k}^T (x_j - x_i)$ . This asymmetric potential design corresponds to directed edges in the graph. For example, given the condition that  $x$  represents height information,  $l$  is roof and  $k$  is building, when  $x_i - x_j$  is negative, the potential design will encourage  $y_i = k, y_j = l$  but penalize  $y_i = l, y_j = k$ .

## 4.2 Long-range CRF model

### 4.2.1 Graph Model

Let  $G_L = (V, E_L)$  be a long-range graph over line segments. Each line segment is regards as one node in  $G_L$ . The method we use to construct a sparse long range graph is similar with (Li and Huttenlocher, 2008). In current study, we consider a spatial arrangement between adjacent objects only in vertical direction, “above-and-below” relation. Thus a line segment finds its neighbours upward and downward in relation to current node as shown in Figure 3. A scan profile was discretized into a set of non-overlapping vertical subspaces (rectangle area between

dotted blue lines) with 0.5m width in Figure 3. Long range neighbouring nodes were searched in a vertical subspace (blue filled area) where current node falls into. After excluding those nodes captured within 1 meter away, nodes with the first and second nearest (both upward and downward) were selected as its long range neighbours. Thus, the maximum number of long range neighbours corresponds to four (2 upward and 2 downward). Please note that some nodes may not have any long range neighbours. The conditional distribution over labels  $Y$  given observed data  $X$  in  $G_L$  can be now defined as below:

$$P_L(Y | X) = \frac{1}{Z_L(X)} \exp \left( \sum_{i \in V} A_i(X, y_i) + \sum_{(i,j) \in E_{L_i}} L_{ij}(X, y_i, y_j) \right) \quad (6)$$

where  $A_i(X, y_i)$  is the association potential, similar with short range potential.  $L_{ij}(X, y_i, y_j)$  is the long range potential that penalizes compatibility of spatial arrangement between labels of neighbouring nodes. Since the short range and long range CRF model share the same association term, this part will not be introduced again.  $Z_L$  is the normalization term.

#### 4.2.2 Long-range Interaction Term

The long-range dependency can be formulated using Bayesian rule, which is described in Equation 7.

$$L_{ij}(u_{ij}, y_i, y_j) = \log(P(y_{above} = l | y_{below} = k, u_{ij})) \\ = \frac{P(u_{ij} | y_{above} = l, y_{below} = k) P(y_{above} = l, y_{below} = k)}{\sum_{y_{above}} P(u_{ij} | y_{above} = l, y_{below} = k) P(y_{above} = l, y_{below} = k)} \quad (7)$$

where  $(y_i, y_j)$  is a pair of lines forming an edge in  $G_L$ ;  $y_{above}$  is defined if one of  $(y_i, y_j)$  is placed higher than the other, otherwise as  $y_{below}$ .

Equation 7 estimates the probability of  $y_{above}$  labelled as  $l$ , given edge feature  $u_{ij}$  and  $y_{below}$  labelled as  $k$ . In Equation 7,  $P(y_{above} = l, y_{below} = k)$  models a co-occurrence rate of class  $l$  that is placed above class  $k$ . This prior was calculated over all label pairs, which represents *a priori* knowledge of spatial arrangements between object pairs. This statistically-derived knowledge was formed in a look-up table shown in Figure 4(a). The likelihood in Equation 7 is the probability distribution of edge feature  $u_{ij}$  given a configuration of that class  $l$  is above class  $k$ . The edge feature  $u_{ij} = \{|h_i - h_j|, |o_i - o_j|, |l_i - l_j|\}$ ,  $h$ , height;  $o$ , orientation;  $l$ , length. We assumed that the likelihood follows a multivariate Gaussian distribution (mean vector:  $\mu_{l,k}$ ; covariance matrix  $\Sigma_{l,k}$ ) described in Equation 8. The normalization term is a marginal probability over  $y_{above}$ . Figure 4(b) give an example of probability distribution of height difference when one low man-made object (LMO) is bellows other objects. The x axis represents value of height difference.

$$P(u_{ij} | y_{above} = l, y_{below} = k) \\ = \frac{\exp \left( -\frac{1}{2} (u_{ij} - \mu_{l,k})^T \Sigma_{l,k}^{-1} (u_{ij} - \mu_{l,k}) \right)}{2\pi \sqrt{|\Sigma_{l,k}|}} \quad (8)$$

We use asymmetric pairwise interactive potential to reflect the spatial arrangement between adjacent (long range) objects. From Equation 7 we can see that  $\log P(y_{above} = l, y_{below} = k | u_{ij}) \neq \log P(y_{above} = k, y_{below} = l | u_{ij})$ , which made asymmetric long range potential  $L_{ij}(x, y_i, y_j) \neq L_{ji}(x, y_j, y_i)$ . For example, given the condition that node  $i$  is above node  $j$ , the asymmetric potential encourages  $y_i = \text{building}$ ,  $y_j = \text{LMO}$  but penalizes  $y_j = \text{building}$ ,

$y_i = \text{LMO}$ .

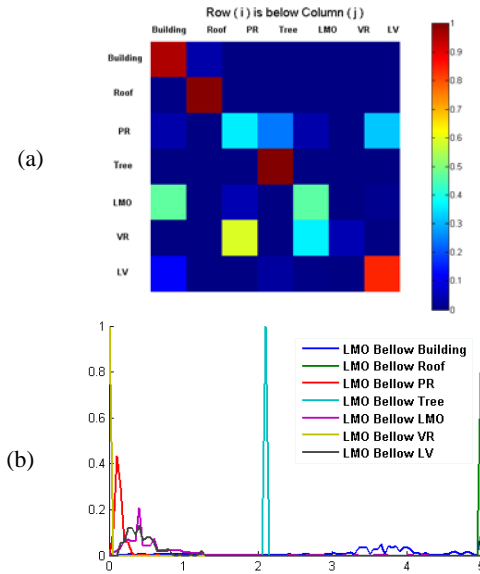


Figure 4. (a) Look-up table and (b) probability distribution of height difference when LMO is placed below the other objects.

#### 4.3 Parameters Learning

Since srCRF and lrCRF are independent of each other, their parameters are separately learnt. In srCRF, the parameter learning is divided into two stages. At the first stage, GMM's parameters were learned using EM algorithm. Once it was done, the posterior probabilities of GMM-EM classifier were used as the association term in Equation 3. Since lrCRF shares the same association term with srCRF, GMM's parameters can be shared.

Maximum likelihood is often used for estimating parameters by differentiating the likelihood function with respect to parameter. However, it is not possible to estimate parameter directly in CRFs by maximizing the log-likelihood, because it does not always yield a closed form solution. Gradient based method is an optimal option. An advanced Gradient based method for CRFs is limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method, which uses a limited memory to approximate the inverse Hessian matrix, the detail of which can be check in (Liu and Nocedal, 1989). However, the batch-only algorithm is wasteful when the training data consist of large and redundant data sets. Instead of updating parameters until they have scanned the entire training set, stochastic gradient descent (SGD) takes a small step in the direction given by the gradient for that instance only and so it converges faster. To make a balance between convergence quality and speed, Vishwanathan et al. (2006) proposed a stochastic gradient method to CRFs, which is also used to estimate parameters in our srCRF model. The last step is to estimate the parameters in long range interaction term (Equation 7). The look-up table was created from frequency statistics of co-occurrence over labels. Parameters in Equation 8 were estimated using Maximum Likelihood (ML).

#### 4.4 Inference

Both of short-range and long-range graphs have cyclic structure and so exact inference over cyclic graph is an intractable problem. Loopy belief propagation (LBP) is an exact approximation solution for graphs with cycles (Murphy et al., 1999). Computing an approximate gradient using LBP, and



learning CRF model parameters using stochastic gradient-based optimization method, has been approved to work well in Vishwanathan et al. (2006). Given estimated CRF parameters, labelling inference was conducted separately in short range graph and long range graph using LBP. The conditional probability of maCRF model is the product of prediction of srCRF and lrCRF models. The final decision was made using Maximum a Posterior Probability (MAP) decision rule.

## 5 EXPERIMENT

### 5.1 Dataset

The performance of our proposed classification method (maCRF) was evaluated with TLS data. The data set was collected by RIEGL LMS-Z390i at two different sites (Site A and Site B) located in Kidd Terrace, Toronto. The datasets show typical residential street scenes where both semi- and detached houses are densely built. Two sites contain our seven targeted object classes, which include building, roof, tree, low vegetation, low man-made objects (LMO), pedestrian road and vehicle road. Figure 5 shows manually labelled references of the two dataset.

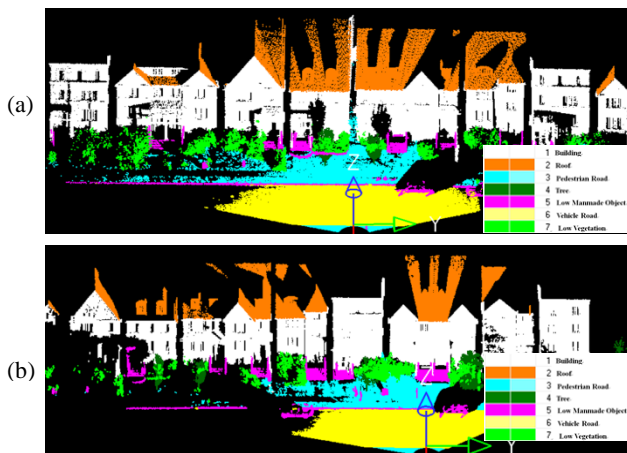


Figure 5. Ground truth of the Site A (a) and Site B (b).

Table 1 summarizes the total number of spatial entities used for constructing CRFs: laser points, scan profiles, line segments and edges used in two graphical models of srCRF and lrCRF for classifying TLS data.

	Site A	Site B
laser points	3,294,337	3,087,301
scan profiles	2,810	2,580
line segments	105,620	100,648
edges in srCRF	260,579	277,584
edges in lrCRF	158,276	156,023

Tabel 1. Total number of the entities: laser point, scan profiles, line segments and edges used in srCRF and lrCRF.

To evaluate the importance of multi-range interactions, we conducted a comparative analysis of classification results obtained from four different classifiers: 1) GMM-EM (local classifier without contextual label interactions); and CRF models 2) with short-range interaction (srCRF), 3) with long-range interaction (lrCRF) and 4) with both short-range and long-range interactions (maCRF). The two-fold cross validation was used for measuring respective classification performance.

For each classifier, the model parameters were learnt using one of the datasets, while the other site was used for testing the trained model. The classification performance was measured on each site and then averaged over both sites. The classification accuracy was estimated on line segment level, not point level. First, the ground truth of target object per laser point was manually labelled for evaluating the performance of respective classifiers. Then, the ground truth of line segments extracted was labelled according to the majority vote of labelling results of its member points.

### 5.2 Qualitative Evaluation

Figure 6 presents classification results of Site B obtained from four classifiers. We observed that context-based classifiers (srCRF, lrCRF and maCRF) shows better classification performance compared to local classifier (GMM-EM). It is clear to see that GMM-EM produced very noisy prediction results. Compared to GMM-EM, we confirmed that srCRF produced much “smoothed” classification results within individual objects, especially over building façade and trees (Figure 6(b)). As shown in Figure 6(c), lrCRF rectified local classification errors caused by srCRF (e.g., the roof below building and tree inside façade). This fact implies the scene-layout compatibility augmented by long-range contest works well. Finally, we can observe that maCRF shows slightly better classification results compared to lrCRF.

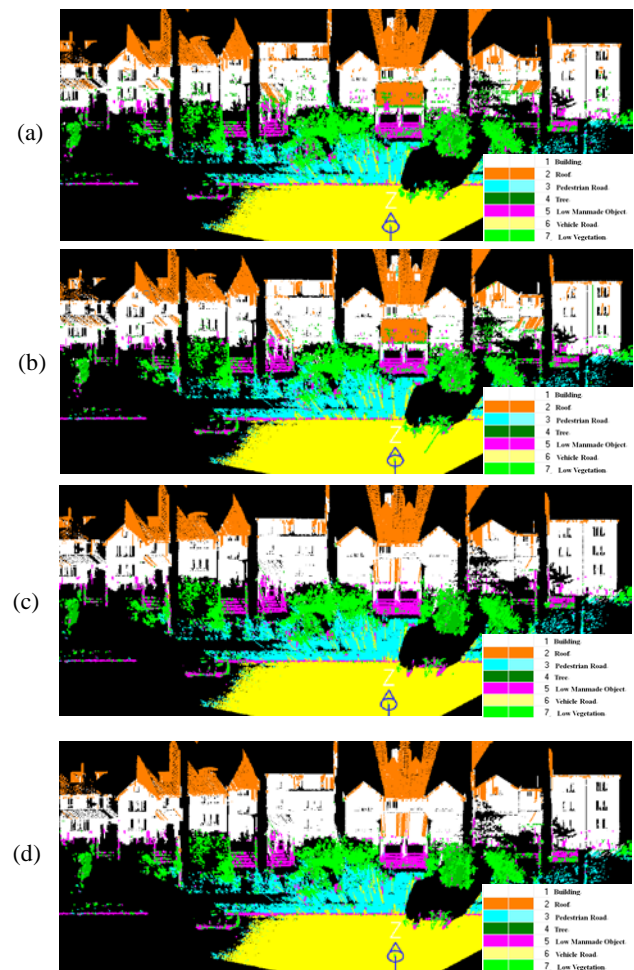


Figure 6. Classification results of Site B: (a) GMM-EM, (b) srCRF, (c) lrCRF and (d) maCRF.

For further investigation of respective classification performance of four classifiers, we arbitrarily selected one scan profile and visualized its labelling results as shown in Figure 7. It is clear to conclude that maCRF model yielded the most accurate classification results and produced significant improvement compared to the other classifiers. In Figure 7(a), GMM-EM produced the largest commission errors between pairwise classes, which include: building-tree, building-LMO and tree-LMO. However, Figure 7(b) shows some portion of those commission errors were rectified by srCRF through enforcing local label consistency among adjacent entities. However, this short-range interaction enhances a local smoothness only, but did not work effectively to guarantee global spatial arrangement. For instance, srCRF still produced classification errors, which do not follow common scene-layout compatibility such as: “trees are not placed within building façade” or “buildings are not placed at the treetops” (Figure 7(b)). As explained in Section 3, lrCRF model made constraints on global spatial arrangement by considering long-range interactions of line segments. Figure 7(c) suggests that lrCRF enables rectifying spatial arrangement errors in vertical direction, but failed to correct local inconsistency in short-range interaction. Two CRF models (srCRF and lrCRF) showed their respective benefits and limitations. By considering local smoothness and global spatial arrangement together, the combined maCRF model produced the most accurate classification results (Figure 5 (d)).

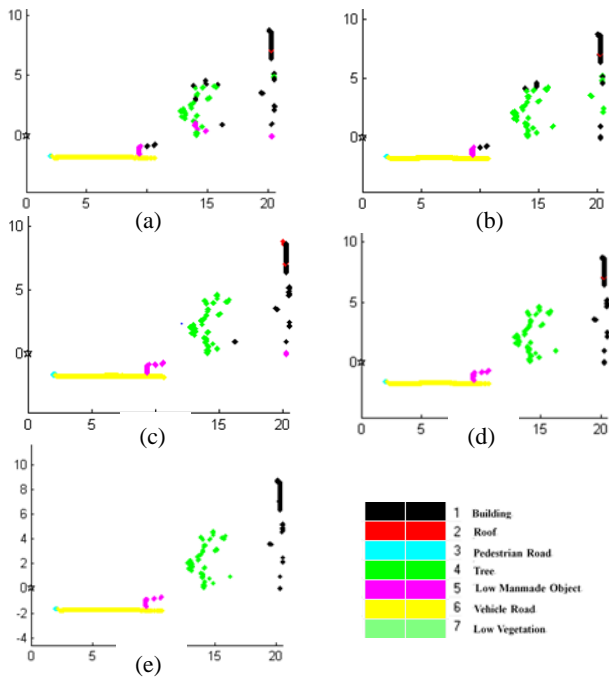


Figure 7. Per-line classification results of Site B: (a) GMM-EM classifier, (b) srCRF, (c) lrCRF, (d) maCRF and (e) ground truth.

### 5.3 Quantitative Evaluation

To quantitatively evaluate classification results, confusion matrices were computed for four classifiers. Based on the confusion matrices, the overall classification accuracy (i.e., precision and recall) were computed. The overall accuracy of four methods on each folder is presented in Table 2. We observed an obvious advantage of the contextual information; all three contextual classifiers showed higher accuracy than the GMM-EM classifier. lrCRF demonstrated better performance than srCRF. By combining the short range and long range interaction, maCRF improved its classification accuracy by

6.77% compared with GMM-EM. While the overall accuracy indicates the classification performance averaged over all the classes, the precision and recall evaluate the performance of a classifier per single class. In case of building class, the precision measures the percentage of lines that are correctly classified as building as a ratio of all the objects that truly be building (Table 3). The recall describes what proportion of lines that should be classified as building is in fact classified as building (Table 4).

	Site A	Site B	Averaged
GMM-EM	78.79	79.53	79.16
srCRF	81.77	81.34	81.56
lrCRF	85.97	84.72	85.35
maCRF	86.71	85.14	85.93

Table 2. The overall classification accuracy of four methods.

In Table 3, we can observe that maCRF has positive effects on the precision of all classes. Building and tree classes benefit most from multi-range context since enforcing the spatial arrangement largely can rectify relational errors (e.g., building on the top of tree, or tree in the building). From Table 4, we can also observe that maCRF positively contributed on the recall of all classes. The recall measure over the roof class was improved by 10% since maCRF enforced the scene-layout compatibility so that roof is likely to be labelled if they are located above the building classes.

	GMM-EM	srCRF	lrCRF	maCRF
Building	84.12	89.35	91.01	92.28
Roof	81.39	72.56	85.20	81.48
PR	74.10	80.19	76.06	79.17
Tree	78.73	67.33	93.29	91.48
LMO	68.75	68.10	77.75	76.30
VR	91.27	91.35	91.94	91.86
LV	70.36	77.85	76.10	76.23

Table 3. Precision of each class in four methods.

	GMM-EM	srCRF	lrCRF	maCRF
Building	92.90	90.24	94.49	94.08
Roof	59.61	65.43	65.11	69.01
PR	89.42	85.43	91.21	89.86
Tree	61.62	77.35	80.37	83.45
LMO	64.95	68.90	74.01	75.13
VR	87.44	87.55	87.84	88.56
LV	62.38	66.88	72.80	72.82

Table 4. Recall of each class in four methods.

## 5. CONCLUSION

In this study, we address the problem of incorporating context constraint in TLS point clouds classification. A new multi-range and asymmetric CRF model (maCRF) was proposed to explicitly model the scene-layout compatibility with respect to a long-range dependency. maCRF constructs two independent graphical models, short-range (srCRF) for enhancing a local labelling smoothness, while long-range (lrCRF) for imposing a global and asymmetric regularity on spatial arrangement between different object classes. Therefore, the maCRF is expected to combine local smoothness and global scene-layout compatibility together. To evaluate the importance of multi-range interactions, classification performance of GMM-EM and srCRF, lrCRF and maCRF were comparatively evaluated. Experimental results show that context-based classifiers

(srCRF, lrCRF and maCRF) have better classification performance than local classifier (GMM-EM). Moreover, maCRF is slightly better than individual srCRF and lrCRF, which validates the advantages of multi range context constraints. In our future research, we will extend the concept of the long-range dependency to horizontal direction. Furthermore, multi-range contexts will be integrated in the same graphical model, rather than independently. In addition to the multi-range context, we also will exploit the benefit of multi-scale context.

## 6. ACKNOWLEDGEMENT

This research has been supported by the Chinese State Scholarship Fund and Korean Ministry of Land, Infrastructure and Transportation for R&D project on the information retrieval of railway corridor infrastructure.

## REFERENCE

- Anguelov, D., Taskarf, B., Chatalbashev, V., Koller, D., Gupta, D., Heitz, G., Ng, A., 2005. Discriminative learning of markov random fields for segmentation of 3d scan data. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 169-176.
- Bar M., Ullman S., 1996. Spatial context in recognition. *Perception* 25, pp. 343-352.
- Belton, D., Lichti, D. D., 2006. Classification and segmentation of terrestrial laserscanner point clouds using local variance information. *IAPRS, XXXVI*, 5, pp. 44-49.
- Cross, G. R., Jain, A. K., 1983. Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI (1), pp. 25-39.
- 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.
- He, X., Zemel, R. S., Carreira-Perpindn, M. A., 2004. Multiscale conditional random fields for image labeling. In: *CVPR 2004, Proceedings of the 2004 IEEE on computer vision and pattern recogniton*, Vol. 2, pp. 695-702.
- Hu, X., Ye, L., 2013. A fast and simple method of building detection from LiDAR data based on scan profile analysis. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-3/W1*, pp. 308-312.
- Jiang, X., H. Bunke., 1994. Fast segmentation of range images into planar regions by scan line grouping. *Machine Vision and Applications*, 7(2), pp. 115-122.
- Kumar, S., Hebert, M., 2006. Discriminative random fields. *International Journal of Computer Vision*, 68(2), pp. 179-201.
- Li, Y., Huttenlocher, D. P., 2008. Sparse long-range random field and its application to image denoising. In: *Computer Vision-ECCV*, pp. 344-357.
- Lim, E. H., Suter, D., 2009. 3D terrestrial LIDAR classifications with super-voxels and multi-scale Conditional Random Fields. *Computer-Aided Design*, 41(10), pp. 701-710.
- Liu, D. C., Nocedal, J., 1989. On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 45(1-3), pp. 503-528.
- Luo C., Sohn G., 2013. Line-based Classification of Terrestrial Laser Scanning Data using Conditional Random Field. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1(2), pp. 155-160.
- Manandhar, D., Shibasaki, R., 2001. Feature extraction from range data. In *Paper presented at the 22nd Asian Conference on Remote Sensing*, Vol. 5, 6p (on CDROM).
- Munoz, D., Vandapel, N., Hebert, M., 2008. Directional associative markov network for 3-d point cloud classification. In: *Proc. 4th International Symposium on 3D Data Process, Visualization and Transmission*, Atlanta, 18–20 Jun, pp. 1–8.
- Murphy, K. P., Weiss, Y., Jordan, M. I., 1999. Loopy belief propagation for approximate inference: An empirical study. In: *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pp. 467-475.
- Niemeyer, J., Rottensteiner, F., Soergel, U. 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 87, pp. 152-165.
- Krzanowski, W. J., 2000. *Principles of multivariate analysis*. Oxford University Press, Oxford, pp187-189.
- Pu, S., Vosselman, G., 2009. Knowledge based reconstruction of building models from terrestrial laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(6), pp. 575-584.
- Shapovalov, R., Velizhev, A., Barinova, O., 2010. Non-associative markov networks for 3d point cloud classification. *Photogrammetric Computer Vision and Image Analysis*, 38(3A), pp. 103-108.
- Sithole, G., Vosselman, G., 2003. Automatic structure detection in a point-cloud of an urban landscape. In: *Remote Sensing and Data Fusion over Urban Areas, 2nd GRSS/ISPRS Joint Workshop*, pp. 67-71.
- Vishwanathan, S. V. N., Schraudolph, N. N., Schmidt, M. W., Murphy, K. P., 2006. Accelerated training of conditional random fields with stochastic gradient methods. In: *Proceedings of the 23rd international conference on Machine learning*, pp. 969-976.
- Triebel, R., et al., 2006. Robust 3D scan point classification using associative markov networks. In: *Proceedings of the 2006 IEEE International Conference on Robotics and Automation, Orlando, Florida*, pp. 2603-2608.
- Yang, M.Y., Förstner, W., 2011. A hierarchical conditional random field model for labeling and classifying images of man-made scenes. In: *2011 IEEE International Conference on Computer Vision*, pp. 196-203.
- Zhao, H., et al., 2010. Scene understanding in a large dynamic environment through a laser-based sensing. In: *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 127-133.