

# Multi-scale Conditional Random Fields for Over-segmented Irregular 3D Point Clouds Classification

Ee Hui Lim

Institute for Vision System Engineering  
Monash University, Clayton VIC Australia  
Eehui.lim@eng.monash.edu.au

David Suter

Institute for Vision System Engineering  
Monash University, Clayton VIC Australia  
d.suter@eng.monash.edu.au

## Abstract

*In this paper, we propose using multi-scale Conditional Random Fields to classify 3D outdoor terrestrial laser scanned data. We improved Lim and Suter's methods [1] by introducing regional edge potentials in addition to the local edge and node potentials in the multi-scale Conditional Random Fields, and only a relatively small amount of increment in the computation time is required to achieve the improved recognition rate. In the model, the raw data points are over-segmented into an improved mid-level representation, "super-voxels". Local and regional features are then extracted from the super-voxel and parameters learnt by the multi-scale Conditional Random Fields. The classification accuracy is improved by 5% to 10% with our proposed model compared to labeling with Conditional Random Fields in [1]. The overall computation time by labeling the super-voxels instead of individual points is lower than the previous 3D data labeling approaches.*

## 1. Introduction

3D outdoor data classifiers are not new. In order to model the urban from scanned data, data classification is necessary to pre-process the data instead of directly triangulating the raw data. This is due to the properties of outdoor data; irregularly sampled, varies in density and containing multiple structures (where simple triangulations may triangulate across nearby data from different classes).

Research in Airborne LIDAR modeling showed that it is possible to divide the raw data into: terrain, man-made object and vegetation (via linear classifiers or clustering methods with features extracted from the height difference [2-5], variations of surface normal vectors [3] and colors [6]). LIDAR data collected at terrestrial level differs from data collected at airborne level - the airborne data is more similar to a 2.5D range image, where the height difference of vegetation data and other flat structure is relatively easy to be observed. In our experiment, which utilized a Riegl

LMS-Z420i terrestrial laser scanner to obtain 3D data, we need a classification technique that deals with the full 3D data as opposed to 2.5D.

We propose using a multi-scale Conditional Random Fields to divide the 3D outdoor LIDAR data acquired from a terrestrial laser scanner into different classes. The solution provides a higher level of constraint than classification with a graphical model, which is resolution and density invariant. Yet our approach only requires relatively small computation time. We validated our algorithm with real world outdoor data acquired from a long range terrestrial laser scanner.

## 2. Background

### 2.1. Graphical model

In data classification via supervised learning, previous work has shown the advantages of global classification (that takes neighboring points into account) over local classification. For irregularly sampled outdoor data, the ambiguity of a single point (sampled from different classes) can be solved using a graphical model. Previous work has shown possible classification solutions with generative [7] and discriminative graphical models [8, 9]. A comparison (mainly the advantages of discriminative graphical models over generative graphical models) of the different graphical models can be found in [10].

#### 2.1.1 Multi-scale Conditional Random Fields (CRFs)

However, the edge potentials in a graphical model are unable to provide a long range correlation, especially for high resolution data. Besides that, classifying every point using only the features of the points and of its neighboring points, in graphical models like Conditional Random Fields (CRFs) [10], can be affected by the difference in resolution among different scans and scanner technology. The density of the point clouds also varies with respect to the different distance from the laser scanner. In order to solve the insufficient constraints in the local smoothing interaction in the image labeling problem, tree structured [11, 12] and multi-scaled [13] approaches are introduced

to the CRFs. Our work is based on the multi-scale CRF approach and we extended the approach for 3D data labeling.

### 2.1.2 Regional and Local Features

3D data labeling in previous work was mostly point based. For example, Anguelov's [8] model. In that model, for every point to be labeled, a fixed number of neighboring points were randomly picked such that: three points were taken randomly in a fixed radius sphere and another three points were taken randomly in a fixed radius cylinder. These provide vertical and horizontal restrictions for pair-wise features. The features were then extracted from every point and its neighbors, for training and probabilistic inference in a graphical model.

However, classifying the neighboring points that belong to the same data class individually is unnecessary and a large amount of computation is required for high resolution data. In order to reduce the data points to be processed, Triebel et al [9] performed kd-tree pruning in the 3D data labeling process which prunes according to position of the point and its label. Therefore, the method only reduces 3D data used for model training and cannot be applied to inference of the model.

For 2D images, another method to avoid point-based classification is to over-segment the raw image into higher level individual pixels. The Normalized Cut segmentation algorithm has been proposed to groups similar regions [14] in 2D images. The resulting 'super-pixel' image is irregularly sampled and similar in that respect to 3D data. Regional and global features are then extracted from the super-pixels. This work showed the redundancy of classification at individual pixels level. With a similar concept, our work eliminates the redundant computation required to label 3D data at individual point level.

## 3. Model Architecture

Our model consists of multi-scale Conditional Random Fields with local and regional features extracted from super-voxels which is an adaptive over-segmentation technique. Super-voxels were first introduced in [1]. The similar neighboring points are grouped together in 3D space. We improve the labeling accuracy shown in [1] by extracting features not only from within super-voxels but as well from neighboring super voxels, therefore providing a higher level constraint. We have also modified the algorithm to include color similarity in the over-segmentation to avoid grouping data from different classes on the same flat surface.

## 3.1. Over-segmentation

Similar to He's model for 2D image segmentation, we over-segment the 3D data before the data classification takes place, using algorithms modified from 3D scale theory [15, 16] instead. The individual 3D points are

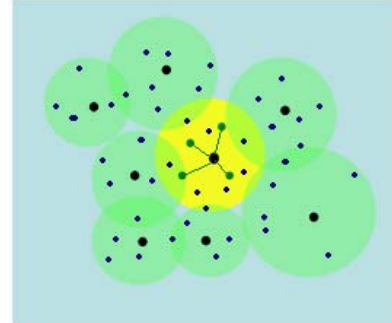


Fig. 1. Over-segmentation of 3D point clouds into super-voxels  $m=4$  neighboring points are randomly selected within every super-voxel.

clustered together to form a higher level representation as shown in Fig 1. For  $p$  data points (Fig. 1), the number of  $n$  super-voxels, where  $n \ll p$ , are computed based on the normalized color similarity and the geometry of the data structure. The radius of the super-voxel,  $r$  is iteratively estimated with the following equations [1] that depend on the estimated curvature, density, noise and the colours of the points:

---

Randomly pick a point  $p$

- Check if the  $p$  belongs to any previous super-voxel
- Start with  $k=15$
- Iterate and refine (maximum of 10 steps) to estimate the optimal size of the super-voxel

- Compute radius of super-voxel  $r$ , estimated density  $\rho$ , estimated curvature  $\kappa$  [17] locally
  - $r$ : distance from  $p$  to its  $k$ -th nearest neighbours

$$\rho \leftarrow k / \pi r^2 \quad (1)$$

$$\kappa \leftarrow 2d / \mu^2 \quad (2)$$

- Use them to compute  $r_{new}$ 

$$r_{new} = (1 - damp) \times r_{old} \quad (3)$$

$$+ damp \times \left\{ \left[ \frac{1}{\kappa} \left( d_1 \frac{\sigma_n}{\sqrt{\epsilon \rho}} + d_2 \sigma_n^2 \right) \right]^3 \times [1 - \max(\text{var}(\text{colour}))] \right\}$$

- Compute  $k_{new} \leftarrow \left\lfloor \pi \rho r_{new}^2 \right\rfloor \quad (4)$

- Stop if  $k_{new} > \text{threshold}$  or  $k_{new}$  saturates

---

Until all points belong to a super-voxel

---

$d$  in eq.2 is the shortest distance between the point and the least square fitted plane to the  $k^{th}$  nearest neighboring points. In eq.3,  $\epsilon$  is a very small number. The noise constant  $\sigma$  can be estimated experimentally by computing the average distance of every point (acquired from a single plane) to the least square fitted plane for different

resolutions.  $d_1$  and  $d_2$  are obtained by solving a minimization problem (details can be found in Lalonde et al. [16]). The damping factor  $damp$  set to be 0.2 is to avoid the iterations going into a marginally stable state.

We have included color properties of the data in eq.4 to provide color constraints within the super-voxel. The variances of the normalized RGB values are computed for the super-voxels. The maximum of  $[var_R \ var_G \ var_B]$  is then used as a factor to reduce the radius if the colour within the super-voxel is inconsistent.

For every super-voxel,  $m$  neighboring points from surrounding super-voxels were randomly selected to be included as local edges in the graphical model. The optimal super-voxel radius is then estimated from every point  $m$  using the same algorithm in Eq. 1-4. This is done in order to extract features that will be fed into the graphical model, to ensure continuity between the super-voxels. The effect of different resolution (between scans and varying density across different distance from the scanner) is minimized with the mid-level super-voxels representation.

The over-segmented super-voxels can *overlap* each

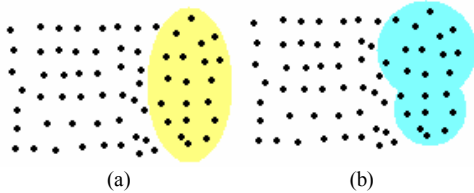


Fig. 2. Multi-scale Conditional Random Fields with local edges (green) and regional edges (black).

other. As the dividing boundary is a 3D sphere in our segmentation, it is possible for some points to belong to more than one region. The advantage of overlap is to provide maximum coverage of data with similar properties. The reason for choosing a sphere over an irregular shape (such as ellipsoid) is to avoid the effect of different shapes in features extraction. For example, if the shape of the segmented super-voxel region is long in one direction (as shown in Fig.2a due to the noise and surface scanned not being exactly flat), the resulting saliency features that depend on the eigenvalues of the covariance matrix of the super-voxel will be linear-like even if the data is planar. The effect of this problem can be reduced using sphere with equal radius as shown in Fig.2b.

### 3.2. Multi-scale Conditional Random Field

Given the over-segmentation in Section 3.1, we construct a probabilistic multi-scale Conditional Random

Fields (mCRFs), as shown in Fig.3, for super-voxels labeling with the following context:

Let  $s = s_1, \dots, s_N$  be the observed feature vectors of some  $N$  super-voxels. Each feature vectors consists of a combination of feature descriptors such as heights, colors,

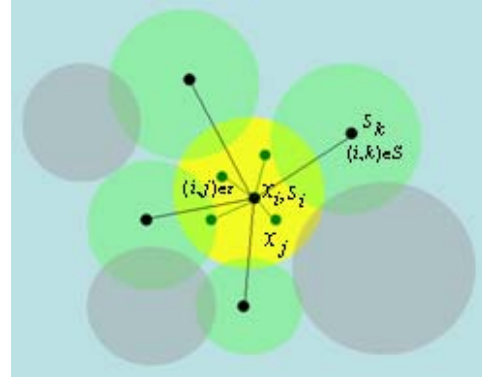


Fig. 3. Multi-scale Conditional Random Fields with local edges (green) and regional edges (black).

SPIN image and estimated normals (see Section 4.1).

Let  $c = c_1, \dots, c_N$  be the of labels in  $C$  given the observable super-voxel. In urban modeling, we select the labels from low level such as ‘planar’ and ‘cluttered’; or higher level such as ‘building’, ‘vegetation’, ‘tree trunk’, ‘grass’ and ‘man-made pathway’.

Let  $x = x_1, \dots, x_M$  be the observed feature vectors of some  $M$  points of point clouds data randomly selected within every super-voxel. Similar to  $s$ , each feature vectors can consist of a combination of feature descriptors such as heights, colors, SPIN image and estimated normals.

The mCRF with parameters  $\theta = \{l, r\}$ ;  $l = \{\lambda_i, \lambda_{ij}\}$ ;  $r = \{\lambda_i, \lambda_{ik}\}$  where  $\lambda_i = \{\lambda_i^1, \dots, \lambda_i^C\}$ ,  $\lambda_{ij} = \{\lambda_{ij}^1, \dots, \lambda_{ij}^C\}$ ,  $\lambda_{ik} = \{\lambda_{ik}^1, \dots, \lambda_{ik}^C\}$  define the conditional probability

for a state sequence to give an observable sequence of:

$$P_l(c | x) = \frac{1}{Z_l} \prod_{(i,j) \in \mathcal{E}} \Psi_{ij}(c_i, c_j, x_i, x_j) \prod_{i=1}^N \Psi_i(c_i, x_i)$$

$$P_r(c | s) = \frac{1}{Z_r} \prod_{(i,k) \in \mathcal{S}} \Psi_{ik}(c_i, c_k, s_i, s_j) \prod_{i=1}^N \Psi_i(c_i, s_i)$$

$$P_\theta(c | s, x) = P_r(c | s) \times P_l(c | x)$$

(5)

$P_l$  is the probability of the super-voxel being labeled as class  $c$  given the features of the mid-point of the super-voxels and its neighbors within the super-voxel.  $P_r$  is the probability of the super-voxel being labeled as class  $c$  given the features of the mid-point of the super-voxels and

the mid-point of its super-voxel neighbors. The final conditional probability of the super-voxel being of class  $c$  is the product of the mentioned probabilities as shown in Eq.5, (with the assumption of independency between the regional and local features). In eq.5,  $Z_l$  and  $Z_r$  are the normalization constants that make the conditional probabilities sum to one. The local edge potential, region edge potential and node potential in eq.5 are defined as follows:

Local edge potential:

$$\begin{aligned} \Psi_{ij}(c_i, c_j, x_i, x_j) &= \exp\left\{\sum_C \lambda_{ij}^C f(c_i, c_j, x_i, x_j)\right\} \\ &= \exp\left\{\sum_C (\lambda_{ij}^C x_i x_j) c_i^C c_j^C\right\} \end{aligned} \quad (6)$$

As shown in Fig. 3, the local edge potential represents the label interactions between the point  $x_i$  with  $m$  randomly selected neighbors within the super-voxel.

Region edge potential:

$$\begin{aligned} \Psi_{ik}(c_i, c_k, s_i, s_k) &= \exp\left\{\sum_C \lambda_{ik}^C f(c_i, c_k, s_i, s_k)\right\} \\ &= \exp\left\{\sum_C (\lambda_{ik}^C s_i s_k) c_i^C c_k^C\right\} \end{aligned} \quad (7)$$

The regional edge potential provides a coarser constraint than the local edge potential.  $l$  number of the closest neighboring super-voxels that are selected as the regional edges.

Node potential:

$$\begin{aligned} \Psi_i(c_i, x) &= \exp\left\{\sum_C \lambda_i^C f(c_i, x)\right\} \\ &= \exp\left\{\sum_C (\lambda_i^C x) c_i^C\right\} \end{aligned} \quad (8)$$

The node potential is a discriminative logistic regression (maximum entropy classifier) that model each label  $c$  as a linear function of  $x$  or  $s$ .

$f(c_i, c_j, x_i, x_j)$ ,  $f(c_i, c_k, s_i, s_k)$  and  $f(c_i, x)$  are feature functions. These are often binary valued for categorical classes (such as in text applications), but in our application with ordinal observations, the feature functions were real-valued; the feature functions were defined over all the local data points feature (for example, the logarithm of saliency features) observation sequence  $x$  and  $s$ , the current state  $c_i$  and the neighboring state  $c_j$  and  $c_k$ .

mCRFs learn by finding the node, local and regional edge weights vector to maximize the log-likelihood. With a Gaussian prior with variance  $\sigma_c^2$ , the log-likelihood is penalized as follows:

$$C_\theta = \sum_{i=1}^N \log P_\theta(c | s, x) - \sum_C \frac{\lambda_c^2}{2\sigma_c^2} \quad (9)$$

where the second summation provides smoothing to avoid over-fitting [18]. The scaled conjugate gradient optimization algorithm is used for the maximization in our experiment.

Given the observation sequence  $x$ , inference in mCRFs is to find a state sequence  $c_{max}$  which is the most likely:

$$c_{max} = \arg \max_c p_\theta(c | s, x) \quad (10)$$

Since exact inference can be intractable in such models, approximate inference using belief propagation is performed for finding  $c_{max}$  in the experiment.

As mentioned in Section 3.1, it is possible for the super-voxels to overlap. In inference, the points belonging to more than one super-voxel will be labeled as the maximum of the product of the conditional probabilities from the overlapped super-voxels. Let  $V$  be the super-voxels that include point  $p$ .

$$c_{max} = \arg \max_c \prod_{v \in V} p_{\theta,v}(c | p) \quad (11)$$

## 4. Experiment setup

We tested our algorithm with real-world data acquired from a terrestrial laser scanner shown in Fig. 4. The mCRFs is trained with hand labeled outdoor laser scanned data. The original segmented 57,734 training data was reduced to 5,850 super-voxels automatically, using the mentioned algorithm. The training data is hand-labeled and the data is chosen from three scans with different density, scene configuration and lighting condition. The total training time is around 5 hours with an Intel Core 2 Duo 2.13GHz CPU and 2GB of RAM.



Fig. 4. Riegl LMS-Z420i Terrestrial Laser Scanner equipped with a calibrated Nikon D100 6 Mega Pixel digital camera

Four neighbors are randomly selected for local edge features and four nearest super-voxels are selected for the regional edge features. Therefore, for every support region, we need to compute feature descriptors for only five points instead of every point. The main computation time is in feature extraction: we can save up to 50% in feature extraction computation. For example, we need to compute  $5,850 \times 5 = 29,250$  feature vectors, which is around half of the total data points.

The 3D data are classified into 5 classes: [Vegetation, Trunk, Man-made objects (building, signboard), Pathway, Terrain (Grass)]. It is important to select good feature descriptors that are capable of differentiating the different data classes for the probabilistic mCRFs.

### 4.1. Feature Descriptors

The features extracted include geometry features such as:

a. Saliency features:

Let  $\lambda_3 > \lambda_2 > \lambda_1$  be the eigenvalues of the covariance matrix of the  $k$  nearest neighbors. In case of clutter,  $\lambda_1 \approx \lambda_2 \approx \lambda_3$  and there is no dominant direction. For points on surfaces where  $\lambda_2, \lambda_3 \gg \lambda_1$  and for linear structures where  $\lambda_3 \gg \lambda_1, \lambda_2$  [16], the saliency features were evaluated using Eq. 12.:

$$\begin{bmatrix} \text{clutter-ness} \\ \text{surface-ness} \\ \text{curve-ness} \end{bmatrix} = \begin{bmatrix} \log(\lambda_1) \\ \log(\lambda_2 / r - \lambda_1) \\ \log(\lambda_3 / r - \lambda_2 / r) \end{bmatrix} \quad (12)$$

As opposed to [16], we normalize the 1<sup>st</sup> and 2<sup>nd</sup> largest eigenvalues with the radius of the super-voxel. This is so that the eigenvalues are invariant to the different size of super-voxel especially for planar-like data (as shown in Fig 5) and linear-like data. For point-like data, the super-voxels are usually relatively small and similar in size; therefore normalization of the 3<sup>rd</sup> largest eigenvalue is unnecessary.

b. Estimated normal vector

The estimated normal vector was calculated as the eigenvector with the largest eigenvalue (from the previous computation of saliency features). The angle between the normal vector and the vertical vector was computed using Eq. 10. This feature is useful in differentiating terrain and building data.

$$\theta = \arccos(n \bullet [0 \ 1 \ 0]) \quad (10)$$

c. Color and intensity

In addition, we also included color and intensity (direct intensity from laser scanner that is illumination invariant)

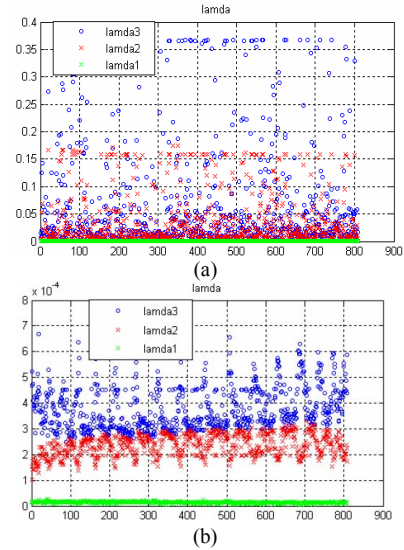


Fig. 5. (a) Three largest eigenvalues (lamda 1, 2 and 3) for planar data

(b) Normalized three largest eigenvalues for planar data

features. The colors of the data points were obtained from the calibrated camera as shown in Fig. 4. Due to the reason that laser scanning and camera images were not total synchronous, inconsistency may have occurred between the images and the point clouds data if moving objects appear in data acquiring environment. We removed inconsistencies caused by in the image and laser scans [19] and the color features were computed from the normalized RGB of these clean images.

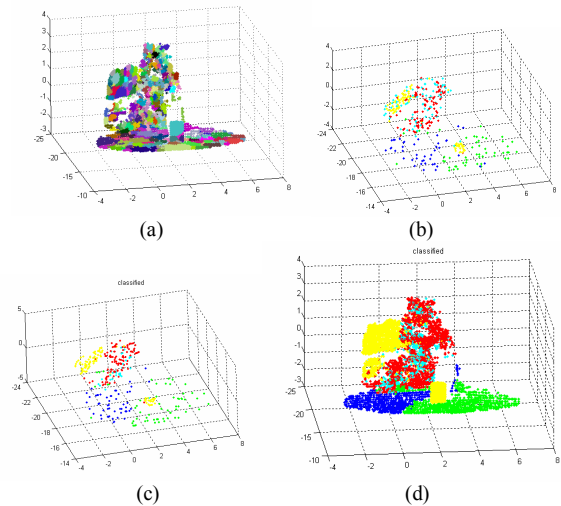


Fig. 6. Data set 1

- (a) 3D points over-segmented into super-voxels
- (b) Labeled super-voxel's mid point with CRFs  
Yellow – Man-made objects, Red – Vegetations, Light blue – Trunks,  
Green – Terrain, Dark blue – Pathways
- (c) Labeled super-voxel's mid point with mCRFs
- (d) Labeled original data with mCRFs

## 5. Results

We validated our algorithm with the following three sets of outdoor scanned data:

a. Data set 1:

With super-voxel over-segmentation, 10,660 points as shown in Fig.6 are reduced into 538 super-voxels. Fig 6a shows the over-segmentation result. Note the bigger super-voxel in the geometrically flat data, such as building and terrain. The computation time taken for feature extraction is around 53.7s; and 0.2s for inference. The authors in [9] reported 2.5 minutes computation time to label 6558 points using kd-tree pruning with 7 features (normals, point distribution histogram and normalized height). The labeled data is shown in Fig. 6d and our classification accuracy is around 86% for mCRFs (Fig. 6c) compared to 78% for CRFs (Fig. 6b).

b. Data set 2:

158,922 points as shown in Fig. 7a is over-segmented into 8,330 super-voxels as shown in Fig. 7b. The computation time taken for feature extraction is around 15 minutes; 87s for inference in mCRFs; and 39s for inference in CRFs. The time difference is 48s which is almost negligible compare to the time taken for feature extractions.

The labeled super-voxels with CRFs are shown in Fig. 7c with classification accuracy around 72%. With the

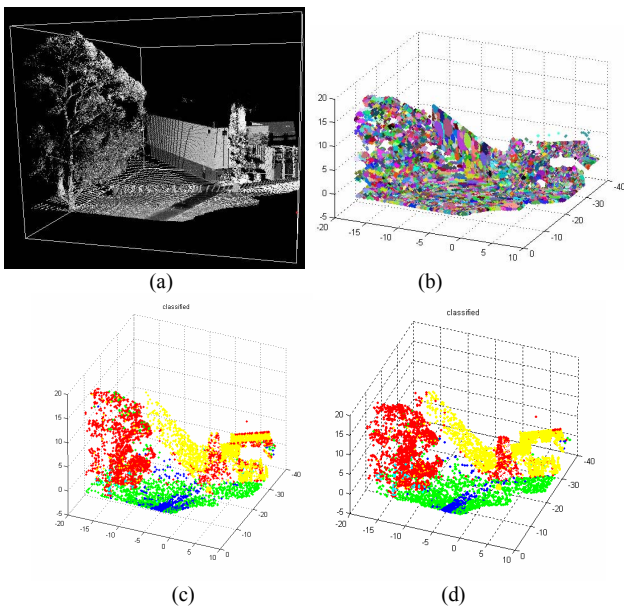


Fig. 7. Data set 2

- (a) 3D points shown in laser direct intensity
  - (b) 3D points over-segmented into super-voxels
  - (c) Labeled super-voxel's mid point with CRFs
  - (d) Labeled super-voxel's mid point with mCRFs
- Yellow – Man-made objects, Red – Vegetations, Blue – Trunks, Green – Terrain, Blue - Pathways

longer range of interaction provided by the regional edge

feature in mCRFs, the label accuracy improved to 79% as shown in Fig. 7d.

c. Data set 3:

We also tested our mCRFs model for large data labeling that contains more than 10 million points. The raw data are divided into small voxels before being processed to reduce the processing time (eliminating data too far away to be included as a candidate in the least square plane fitting).

The total time needed to label a single scan (1,009,942 points) reduced from around 17 hours (without data division) to 5.8 hours for  $100 \times 100 = 10,000$  divisions; to

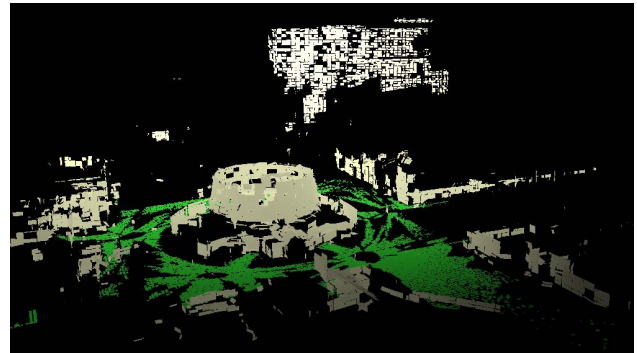


Fig. 8. Plane fitting on labeled building and terrain data

5.1 hours for  $200 \times 200 = 40,000$  divisions; and to 1.6 hours with  $400 \times 400 = 160,000$  divisions. The accuracy dropped from 0.853 for no division to 0.795 for 400x400 divisions. However, this is not a problem due to the dense scan of the objects and the post-processing steps that refine the object modeling.

For a complete scan of the area, a total of seven scans are stitched together (7,086,588 points). The classification accuracy with  $400 \times 400 = 160,000$  divisions is still acceptable as shown in Figure 8. A total of 12.8 hrs are required for the computation of the scans.

Plane patches are fitted onto the labeled building, terrain and floor data using the RANSAC algorithm as a post-processing step to geometrically model the scene. During the process, it is found that the misclassified data or outliers are filtered during the RANSAC plane fitting; while the 'holes' caused by occlusion or misclassification (for eg. Building points labeled as vegetation) are mostly recovered during the plane fitting or surface reconstruction. This step refined the data labeling, and therefore the increment in the classification errors due to data division is not that much of an issue.

## 6. Conclusion

We have shown an efficient and accurate method for 3D data labeling that provides connectivity at local, edge and regional levels. The regional feature between the super-voxels in the mCRFs improves the classification accuracy of CRFs by 5% to 10% while only requiring a negligible increment in the computation time. The computation time and memory requirement can be further reduced by dividing the raw data into smaller grid cells before feature extraction and data labeling take place.

## References

- [1] E. H. Lim, and D. Suter, "Conditional Random Field for 3D Point Clouds with Adaptive Data Reduction" in *International Conference on Cyberworlds, 2007. CW '07.*, 2007, pp. 404-408.
- [2] P. Krishnamoorthy, K. L. Boyer, and P. J. Flynn, "Robust detection of buildings in digital surface models." *Proceedings 16th International Conference on Pattern Recognition*, Quebec City, Canada, 2002, pp. 159-63.
- [3] F. Rottensteiner, "Automatic generation of high-quality building models from lidar data," *IEEE Computer Graphics and Applications*, vol. 23, pp. 42-50, 2003.
- [4] L. Matikainen, J. Hyypä, and H. Hyypä, "Automatic detection of buildings from laser scanner data for map updating," *ISPRS Commission III. Workshop 3-d reconstruction from airborne laserscanner and InSAR data*, 2003.
- [5] G. Vosselman, "Fusion of laser scanning data, maps, and aerial photographs for building reconstruction." *International Geoscience and Remote Sensing Symposium (IGARSS)*, Toronto, Ont., Canada, 2002, pp. 85-88.
- [6] N. Haala and C. Brenner, "Extraction of buildings and trees in urban environments," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 54, pp. 130-137, 1999.
- [7] D. F. Wolf, G. S. Sukhatme, D. Fox, and W. Burgard, "Autonomous Terrain Mapping and Classification Using Hidden Markov Models," in *(ICRA)*. *Proc. of the IEEE International Conference on Robotics and Automation*, 2005, pp. 2026-2031.
- [8] D. Anguelov, B. Taskar, V. Chatalbashev, D. Koller, D. Gupta, G. Heitz, and A. Ng, "Discriminative learning of Markov random fields for segmentation of 3D scan data." *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, San Diego, CA, United States, 2005, pp. 169-176.
- [9] R. Triebel, K. Kersting, and W. Burgard, "Robust 3D scan point classification using associative Markov networks." *Proceedings - IEEE International Conference on Robotics and Automation*, Orlando, FL, United States, 2006, pp. 2603-2608.
- [10] C. Sutton and A. McCallum, "An Introduction to Conditional Random Fields for Relational Learning," *Introduction to Statistical Relational Learning*. Edited by Lise Getoor and Ben Taskar. MIT Press, 2006.
- [11] JieTang, H. Mingcai, L. Juanzi, and L. Bangyong, "Tree-structured conditional random fields for semantic annotation." *The Semantic Web - ISWC 2006. OTM 2006 Workshops. 5th International Semantic Web Conference, ISWC 2006. Proceedings (Lecture Notes in Computer Science Vol. 4273)*, Athens, GA, USA, 2006, pp. 640-53.
- [12] P. Awasthi, A. Gagrani, and B. Ravindran, "Image Modeling using Tree Structured Conditional Random Fields," *IJCAI*, pp. 2054-2059, 2007.
- [13] X. He, R. S. Zemel, and M. A. Carreira-Perpinan, "Multiscale conditional random fields for image labeling." *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, United States, 2004, pp. 695-702.
- [14] H. Xuming, R. S. Zemel, and D. Ray, "Learning and incorporating top-down cues in image segmentation." *Computer Vision - ECCV 2006. 9th European Conference on Computer Vision. Proceedings, Part I (Lecture Notes in Computer Science Vol. 3951)*, Graz, Austria, 2006, pp. 338-51.
- [15] N. J. Mitra, A. Nguyen, and L. J. Guibas, "Estimating surface normals in noisy point cloud data," *Int. J. Comput. Geometry Appl.* 14, vol. (4-5), pp. 261-276, 2004.
- [16] J. F. Lalonde, R. Unnikrishnan, N. Vandapel, and M. Hebert, "Scale selection for classification of point-sampled 3D surfaces." *Proceedings. Fifth International Conference on 3-D Digital Imaging and Modeling*, Ottawa, Ont., Canada, 2005, pp. 285-92.
- [17] S. Gumhold, X. Wang, and R. MacLeod, "Feature extraction from point clouds," *International Meshing Roundtable, Sandia National Laboratories*, October 2001.
- [18] S. F. Chen and R. Rosenfeld, "A Gaussian prior for smoothing maximum entropy models," *Technical Report CMUCS-99-108, Carnegie Mellon University*, 1999.
- [19] E. H. Lim and D. Suter., "Occlusion removal in image for 3d urban modelling," *Image and Vision Computing, New Zealand*, pp. 191-196, Nov. 2006 2006.