Influence of Aleatoric Uncertainty on Semantic Classification of Airborne LiDAR Point Clouds: A Case Study with Random Forest Classifier Using Multiscale Features

Jaya Sreevalsan-Nair, *Senior Member, IEEE,* and Pragyan Mohapatra

Abstract

For semantic classification of LiDAR point clouds, the features derived from the local geometric descriptors are routinely used as features in (supervised) learning algorithms. In this study, our goal is to determine if the aleatoric uncertainty in the input to a supervised semantic classifier influences the outcomes. We consider two sources of such uncertainty – one from the computation of multiscale local geometric descriptors and the other, from class ambiguities at object boundaries. We perform ensembles of experiments to measure the significance of these uncertainties in the semantic classification of airborne LiDAR point clouds, when using random forest classifier. Our case study shows that the presence of aleatoric uncertainty improves the classification outcomes.

Index Terms

Airborne LiDAR point clouds, semantic classification, probabilistic geometric classification, local geometric descriptors, Shannon entropy, semantic homogeneity, aleatoric uncertainty, random forest classifier, multiscale

I. Introduction

Airborne Light Detection and Ranging (LiDAR) point clouds capture topographical data for widespread regions. Unlike imagery, point clouds capture three-dimensional (3D) geometric information, and when integrated with the image data, they provide rich information of the region [1]. Semantic classification is a widely used data processing method, which is increasingly done using supervised learning methods [2]. The feature vector used for several algorithms is obtained from raw data, as well as the eigenvalue decomposition of local geometric descriptors computed at each point [2].

However, there are unanswered questions on the influence of the uncertainties inherent in the point cloud data or those introduced during computation of feature vectors, on the performance of the classifier. The uncertainty pertaining to the data is called *aleatoric uncertainty*, which is different from the uncertainty with respect to the model governing the data, namely, *epistemic uncertainty* [3]. We study the influence of the aleatoric uncertainty from two specific sources on the accuracy of the classifier. We also look at the uncertainty computation when using different methods for aggregating the information from multiple scales in the local geometric descriptor. Our work is an initial attempt to analyze uncertainty in the semantic classification of LiDAR point clouds. It can be further generalized to uncertainties in data acquisition methods, e.g. non-uniform point sampling density [2].

Here, we define terms, namely, *saliency map-based Shannon entropy*, E_{geom}, and *semantic homogeneity*, H_{sem}, which is computed from multiscale probabilistic geometric classification [4] and the semantic composition of the point cloud, respectively. We control E_{geom} and H_{sem} in our experiments to study their influence in the performance of semantic classification. Computation of E_{geom} depends on the local geometric descriptor [5], and its computational methodology [4].

Related Work: Data-dependent aleatoric uncertainty has been recently used for object detection from 3D LiDAR point clouds [6]. Shannon entropy has been used for measuring classification uncertainty, and total variance, for that of object detection. This work has been done on mobile LiDAR, different from the work done in airborne LiDAR for forest data analysis [7], and change detection in glaciology [8]. Thus, our work is novel in uncertainty assessment, specifically for aleatoric uncertainty, for LiDAR point clouds for semantic classification. Our case study focuses on urban regions. We use a random forest classifier, which is well-studied in the domain [2], and as of now, has performed better than the neural network models in the benchmark data [9].

This document is a preprint, as on May 01, 2020. The authors are with Graphics-Visualization-Computing Lab (GVCL), International Institute of Information Technology, Bangalore, 26/C Electronics City, Hosur Road, Bangalore 560100, India. *Corresponding author: Jaya Sreevalsan-Nair.* e-mail: jnair@iiitb.ac.in | {pragyan.mohapatra}@iiitb.org

This work was supported by the Early Career Research Award from Science and Engineering Research Board, awarded by the Department of Science and Technology, Government of India.

II. Our Proposed Method

Our goal is to quantify aleatoric uncertainty in the semantic classification of airborne LiDAR point clouds. We give a background of probabilistic geometric classification with the multiscale approaches and feature vector used for semantic classification. We then propose metrics to measure aleatoric uncertainty, and conduct specific experiments, where uncertainty used for training in a supervised classifier is controlled.

Probabilistic Geometric Classification: Local geometric descriptors are used for identifying the shape of the local neighborhood [10]. Here, we use covariance tensor T_{3DCM} as the local geometric descriptor. The descriptor is computed from either a k-nearest or a spherical (local) neighborhood. The parameter, k or the radius r of the neighborhood, is the *scale*. The neighborhood shape can be classified into "line-", "surface-" and (critical) "point-" type features, which are the geometric classes [4]. The likelihood of the point belonging to these classes, C_l, C_s, C_p, respectively, are computed using eigenvalues of the descriptors at multiple scales [4]. The tuple $\{C_l, C_s, C_p\}$, called as saliency map, gives the probabilistic geometric classification of the point [4].

Multiscale Approaches: It has been widely understood that the outdoor environmental scenes captured by airborne LiDAR point clouds have larger uncertainties involved in scene understanding, which makes the single-scale analysis of the local neighborhood of the points insufficient [2], [4], [9], [10], [11]. Therefore, information is gathered from multiple scales, and then either the descriptors or features, such as $\{C_l, C_s, C_p\}$, are aggregated across scales. We use two computational methods for the descriptors. Firstly, we use k-nearest neighborhood, for $10 \leq k \leq 100$, with step-size of 10, aggregated as a descriptor at optimal scale [10]. Secondly, we use three scales of the spherical neighborhood, where only the saliency maps are aggregated using averages, justified as they are treated as probabilities of multiple independent events [4]. The saliency map computed using these methods are T_{3DCM}(OptmSc) and T_{3DCM}(MultiSc), respectively.

Semantic Classification: We use a random forest classifier for semantic classification with N features per point per scale [2], [12]. The (N=12) features include : $R_{\lambda,2D}$, ΔZ (maximum height difference), $\sigma(z)$ standard deviation of height values of the local neighborhood, D (local point dimensionality), V (verticality), A_λ (anisotropy), E_λ (eigenentropy), $\sum_\lambda$ (the sum of eigenvalues), C_λ (change of curvature), L_λ (linearity), S_λ (Sphericity), P_λ (Planarity). We use the point coordinates, (x, y, z), as additional features. Thus, the feature vector obtained from the descriptor with optimal scale has (N+3) features, and that with multiple scales with averaged saliency map has ($n_r \cdot N$+3) features, where n_r is the number of scales used for spherical neighborhood.

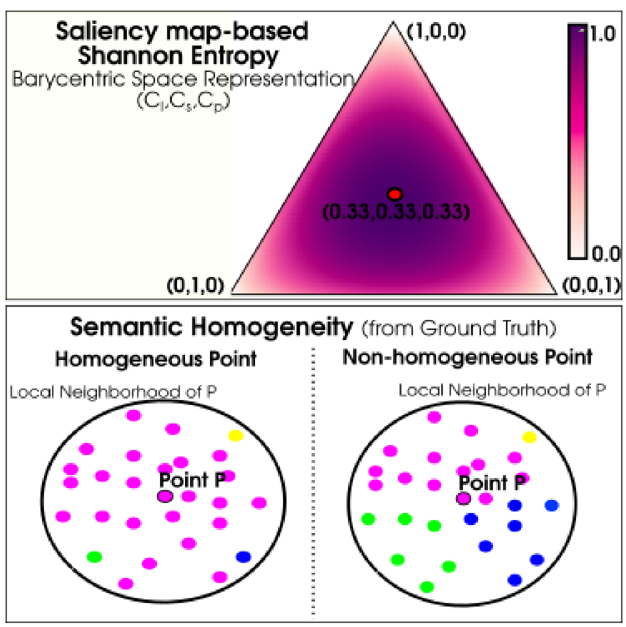

Fig. 1: (Top) Saliency map-based Shannon entropy E_{geom} increases from vertices to the centroid of the triangle in its barycentric representation. (Bottom) Semantic homogeneity H_{sem} is determined based on the percentage of the local neighborhood of a point belonging to its semantic class. As per definitions, $0.0 \leq E_{\text{geom}} \leq ln(3)$, and $0.0 \leq H_{\text{sem}} \leq 1.0$.

A. Measuring Uncertainty

We consider two metrics to quantify uncertainty arising from different processes in the data analytics workflow. These metrics, namely *saliency map-based Shannon entropy* and *semantic homogeneity*, are defined and illustrated (§Figure 1) here. We then design an experimental setting for using these measures of uncertainties in a controlled manner.

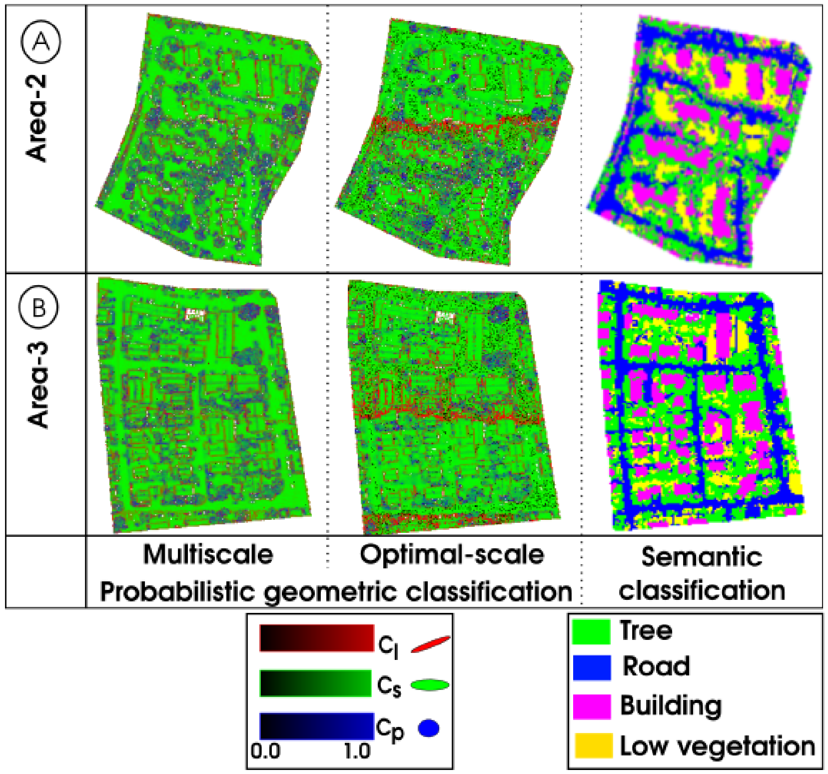

Fig. 2: Top view of Area-2 and Area-3 of Vaihingen site showing probabilistic geometric and semantic classifications – geometric classification when using covariance tensor, with (left) multiscale averaging of saliency map in spherical neighborhood, T_{3DCM}(MultiSc), and (middle) optimal scale with k-nearest neighborhood, T_{3DCM}(OptmSc); (right) semantic class composition of Area-2 with high-rise buildings, and Area-3 with small houses.

Definition II.1. *Saliency map-based Shannon entropy* E_{geom} of a point x is the Shannon entropy of the saliency map. Using $\{C_l, C_s, C_p\}(x)$:

$$E_{\text{geom}} = -C_l.ln(C_l) - C_s.ln(C_s) - C_p.ln(C_p).$$

E_{geom} is proportional to the uncertainty in the probabilistic geometric classification. Since E_{geom} depends on the local geometric descriptor, the aleatoric uncertainty measured using E_{geom} is a *task-dependent* or *homoscedastic* uncertainty.

Definition II.2. *Semantic homogeneity*, H_{sem} of a point x, is the proportion of points in the neighborhood of x that belong to the same semantic class as that of point x. It is expected that, in a segmented point cloud, H_{sem} of points in the interior of the segments is lesser than that at the inter-segment boundaries. Since H_{sem} depends on the ground truth (semantic) labels, the aleatoric uncertainty measured using H_{sem} is a *data-dependent* or *heteroscedastic* uncertainty.

Proposed Ensemble of Experiments: We compute the two metrics at each point. Using probabilistic geometric classification and E_{geom} at each point, using T_{3DCM}(MultiSc) as well as T_{3DCM}(OptmSc), we compute E_{geom}. We find H_{sem} using semantic class labels. For each metric, we train the supervised learning model for semantic classification, using subsets of the point cloud containing specific intervals of the metric, and then test the model on a point cloud. We repeat an ensemble of runs for different intervals of the metric. The use of intervals isolates the influence of the presence or absence of the specific uncertainty on the accuracy of the semantic classifier. We use bounds of the interval starting from 0.0 to the maximum value of the metric, in step-size of 0.1. A combination of lower bound (b_l) and upper bound (b_u) defines the interval chosen for each run, and ($b_l < b_u$). Overall, we run a total of 66 experiments for E_{geom} in an ensemble, and similarly, 55 for H_{sem}. We use two different point clouds for training and testing to compare the accuracy of the classifier for each run. We also compare the performances of the classifier between the descriptors, i.e., T_{3DCM}(MultiSc) and T_{3DCM}(OptmSc).

III. Experiments and Results

We have used the labeled point clouds of Area-2 and Area-3 of the Vaihingen site, given in the ISPRS benchmark data [13], for our ensembles of experiments. The probabilistic geometric classification of T_{3DCM}(MultiSc) and T_{3DCM}(OptmSc) have subtle differences (§Figure 2). The datasets are composed of four semantic classes, namely "Low vegetation," "Road," "Building," and "Tree" (§Figure 2). Training the random forest classifier with 70% of Area-3, and testing with the remaining

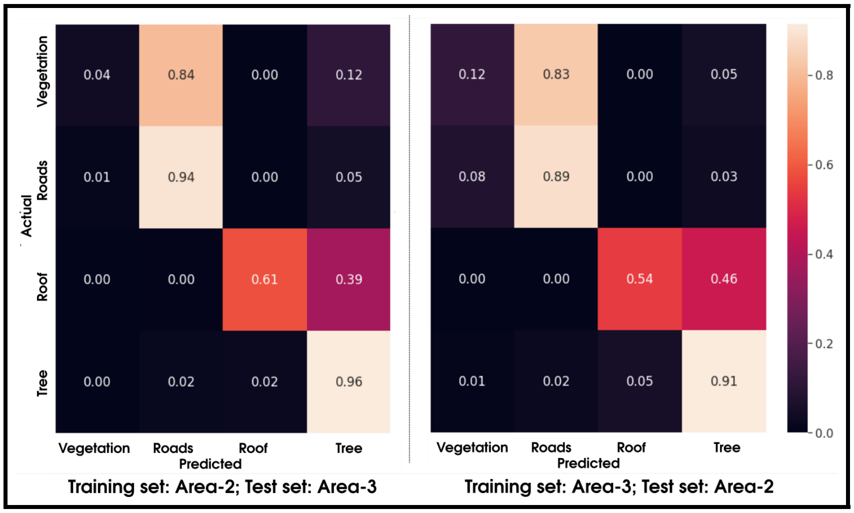

Fig. 3: Confusion matrix of a random forest classifier, when (left) training with Area-2 and testing with Area-3, and (right) training with Area-3 and testing with Area-2.

30% gave 95% accuracy. However, in order to get comparable results from running an ensemble, we train with and test on different point clouds. Training with Area-3, and testing on Area-2, gave 70% accuracy. On the other hand, training with Area-2, and testing on Area-3 gave 78% accuracy. The confusion matrices (§Figure 3) show that low vegetation has been mostly misclassified as road, and building as tree. The misclassification may be attributed to similarity in geometric properties between the classes. Owing to higher accuracy, we present results of the ensembles with Area-3 and Area-2, as training and testing data, respectively.

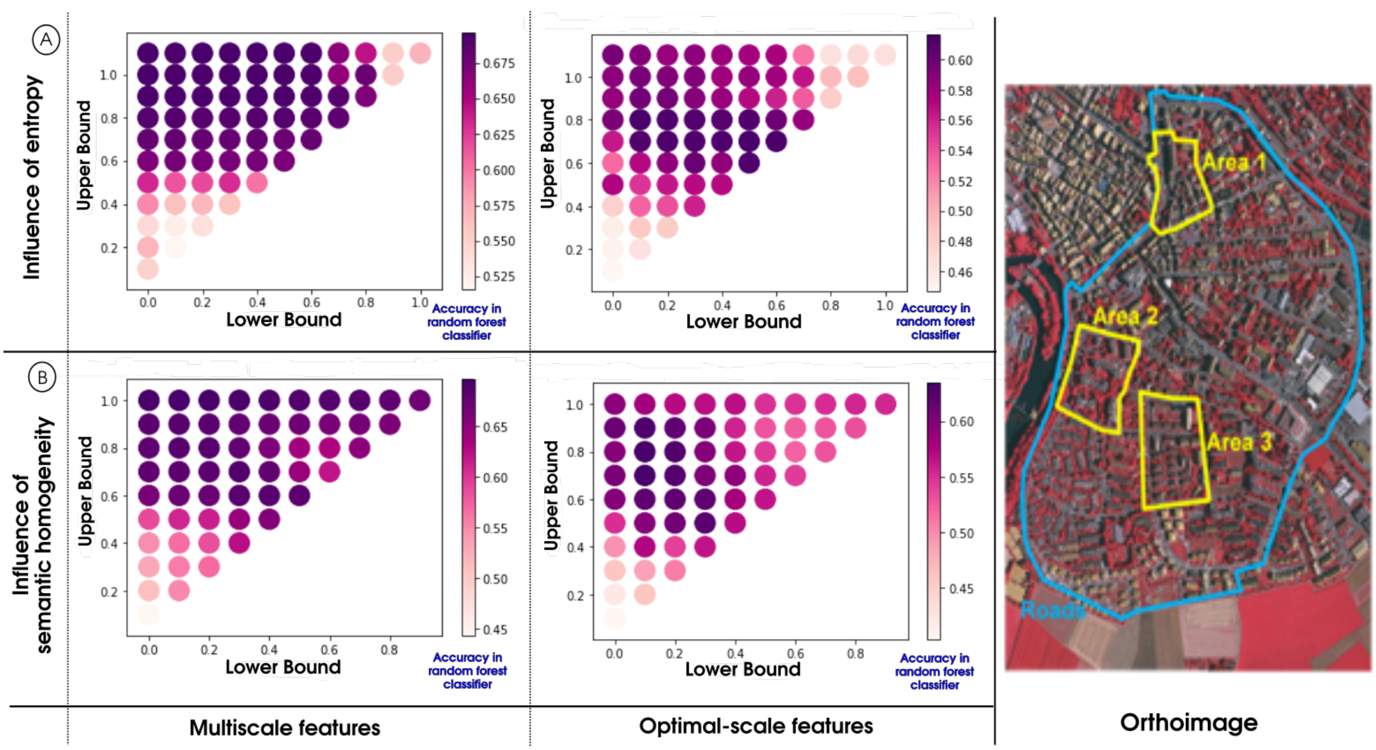

Fig. 4: (Left) Visualization of accuracy results from running ensembles of experiment by controlling the amount of E_{geom} and H_{sem} in the training set of the random forest classifier, using the values of lower and upper bounds. (Right) The orthoimage of Vaihingen site shows Area-2 and Area-3. (Image courtesy: http://www2.isprs.org/commissions/comm3/wg4/tests.html)

The training set of points with E_{geom} in the interval $[0.5, 0.8]$ performs at par with the models containing the entire point cloud (§Figure 4(A)). Thus, the training set containing points with $E_{\text{geom}} \in [0.5, 0.8]$ is sufficient for training using a random forest classifier, which is demonstrated by both T_{3DCM}(MultiSc) and T_{3DCM}(OptmSc). Training sets with points of lower E_{geom}, e.g. interval $[0, 0.5]$, or high E_{geom}, e.g. interval $[0.8, 1.1]$, give lower accuracy.

H_{sem} in intervals varying between $[0.1, 0.5]$ and $[0.2, 0.8]$ show good accuracy results (§Figure 4(B)), for both the descriptors, T_{3DCM}(MultiSc) and T_{3DCM}(OptmSc).

We see relatively better accuracy with T_{3DCM}(MultiSc) than with T_{3DCM}(OptmSc), possibly because the uncertainty owing to E_{geom} is reduced in optimal scale by design. Even though H_{sem} is semantic information-dependent, we find differences in results with T_{3DCM}(MultiSc) and those with T_{3DCM}(OptmSc). This is attributed to the improvement in accuracy results with the increase in the number of features in T_{3DCM}(MultiSc). The lower values of accuracy in the ensembles can be attributed to the variation in the location of Area-2 and Area-3 (§ Orthoimage, Figure 4), including proximity of Area-2 to River Enz, leading to the differences in the habitats and socio-economic trends between the regions.

IV. Conclusions

In this work, we have demonstrated how we can use saliency map-based Shannon entropy, E_{geom}, and semantic homogeneity H_{sem}, for determining aleatoric uncertainty in the semantic classification of airborne LiDAR point clouds. These metrics are computed using probabilistic geometric classification and ground truth semantic labels, respectively. We have shown how an ensemble of experiments is formulated to study the influence of the sources of uncertainty, using our proposed metrics. Our results show that the features aggregated across multiple scales show a better outcome in classification accuracy than the features at the optimal scale. Our work in the systematic analysis of aleatoric uncertainty in the semantic classification of airborne LiDAR point clouds is novel.

References

[1] Franz Rottensteiner, "Status and further prospects of object extraction from image and laser data," in *2009 Joint Urban Remote Sensing Event*. IEEE, 2009, pp. 1–10.

[2] Martin Weinmann, Boris Jutzi, Stefan Hinz, and Clément Mallet, "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 105, pp. 286–304, 2015.

[3] Alex Kendall and Yarin Gal, "What uncertainties do we need in bayesian deep learning for computer vision?," in *Advances in neural information processing systems*, 2017, pp. 5574–5584.

[4] Jaya Sreevalsan-Nair and Beena Kumari, *Local Geometric Descriptors for Multi-Scale Probabilistic Point Classification of Airborne LiDAR Point Clouds*, pp. 175–200, Springer Cham, Mathematics and Visualization, 2017.

[5] Jaya Sreevalsan-Nair and Akshay Jindal, "Using gradients and tensor voting in 3D local geometric descriptors for feature detection in airborne lidar point clouds in urban regions," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2017, pp. 5881–5884.

[6] Di Feng, Lars Rosenbaum, and Klaus Dietmayer, "Towards safe autonomous driving: Capture uncertainty in the deep neural network for lidar 3d vehicle detection," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 3266–3273.

[7] Joseph Mascaro, Matteo Detto, Gregory P Asner, and Helene C. Muller-Landau, "Evaluating uncertainty in mapping forest carbon with airborne LiDAR," *Remote Sensing of Environment*, vol. 115, no. 12, pp. 3770–3774, 2011.

[8] Philip Claudio Joerg, Felix Morsdorf, and Michael Zemp, "Uncertainty assessment of multi-temporal airborne laser scanning data: A case study on an Alpine glacier," *Remote Sensing of Environment*, vol. 127, pp. 118–129, 2012.

[9] Timo Hackel, N. Savinov, L. Ladicky, Jan D. Wegner, K. Schindler, and M. Pollefeys, "SEMANTIC3D.NET: A new large-scale point cloud classification benchmark," in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2017, vol. IV-1-W1, pp. 91–98.

[10] Jérôme Demantké, Clément Mallet, Nicolas David, and Bruno Vallet, "Dimensionality based Scale Selection in 3D LiDAR Point Clouds," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 38, no. Part 5, pp. W12, 2011.

[11] Hugues Thomas, François Goulette, Jean-Emmanuel Deschaud, and Beatriz Marcotegui, "Semantic classification of 3d point clouds with multiscale spherical neighborhoods," in *2018 International Conference on 3D Vision (3DV)*. IEEE, 2018, pp. 390–398.

[12] M. Weinmann, B. Jutzi, and C. Mallet, "Feature relevance assessment for the semantic interpretation of 3d point cloud data," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 5, pp. W2, 2013.

[13] Franz Rottensteiner, Gunho Sohn, Jaewook Jung, M Gerke, Caroline Baillard, Sebastien Benitez, and Uwe Breitkopf, "The ISPRS benchmark on urban object classification and 3D building reconstruction," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences I-3*, pp. 293–298, 2012.