

Rule-based Improvement of Maximum Likelihood Classified LIDAR Data fused with co-registered Bands

Marc Bartels and Hong Wei

*Computational Vision Group, School of Systems Engineering
The University of Reading
Whiteknights, Reading RG6 6AY
United Kingdom*

*m.bartels@reading.ac.uk, h.wei@reading.ac.uk
<http://www.cvg.reading.ac.uk/projects/LIDAR>*

Abstract

In the past decade, Light Detection And Ranging (LIDAR) has been recognised by both the commercial and public sector as a reliable and accurate source for land surveying. Object classification in LIDAR data tends towards data fusion by employing additional simultaneously recorded bands. In this paper, a rule-based approach is presented for improving classification accuracy obtained in a supervised classification based on Maximum Likelihood using high resolution LIDAR data, co-registered bands, and further height derived texture features. The issues regarding feature and class selection as well as differentiated accuracy assessment are addressed. The results show that incorporating additional knowledge and considering relationship among classes is beneficial for improving classification accuracy in fused LIDAR data sets.

1 Introduction

Light Detection And Ranging (LIDAR) for terrain and land surveying has made significant contributions to many environmental, engineering and civil applications. It is therefore not surprising that LIDAR data is being used increasingly by the public sector and commercial world since the early 1990s (Maas, 2005). Applications such as forestry, building reconstruction, flood modelling and corridor mapping are based on post processing of LIDAR data point clouds, as they are accurate for less hilly terrain (Huising and Pereira, 1998). Mounted on an airborne platform, a LIDAR data acquisition system estimates the distance between the instrument and a point on the surface by measuring the time the laser pulse needs to hit its receiver (Huising and Pereira, 1998). A Global Positioning System (GPS) receiver and an Inertial Navigation System (INS) complement the data with position and orientation, respectively (Huising and Pereira, 1998).

Early investigations on height in LIDAR data (Weidner and Förstner, 1995) aimed at ground and object point separation in estimating a normalised Digital Surface Model (nDSM) by subtracting a morphologically filtered Digital Terrain Model (DTM) from the original Digital Surface Model (DSM). Building modelling from LIDAR data in a less hilly area was presented in two parts (Maas and Vosselman, 1999): first, a texture based segmentation employing Maximum Likelihood (ML) applied to LIDAR data and two further height derived bands (Maas, 1999b; Maas, 1999c); and second, the actual building modelling from the point cloud (Maas, 1999a). Separating ground and object points on gridded LIDAR data using wavelets and K-means on height to assign pixels to buildings, motorway, boundary and two tree types was presented in (Vu and Tokunaga, 2001). A noise robust texture-based segmentation approach using wavelet packets, co-occurrence matrices and normalised modified histogram thresholding has been proposed in (Bartels *et al.*, 2005) where the authors partitioned ground and objects into rivers, fields and residential areas.

In the course of increasing the LIDAR data resolution and the possibility of recording complementary bands, researchers developed further classification techniques using data fusion. Realistic 3D city models from LIDAR data were presented in (Haala and Brenner, 1997; Haala and Brenner, 1998) where buildings, facades and vegetation were reconstructed using multiple data sources, such as near infra-red (NIR) and terrestrially captured digital images (Haala *et al.*, 1998). The building ground planes were estimated with a 2D Geographical Information System (GIS) map complemented with a cadastral map. Building reconstruction by fusing LIDAR data and aerial photos was presented in (Rottensteiner and Briese, 2003). First, the authors detected building regions in ungridded data. Then, roofs were detected using a curvature-based segmentation technique and additional planar faces were estimated with aerial photos. A supervised parametric classification algorithm based on the Gaussian mixture model and Expectation Maximisation (EM) was used in (Charaniya *et al.*, 2004) to classify roads, roofs, trees and grass in LIDAR data using data fusion with an nDSM (*i.e.* the difference between LIDAR DSM and a given DTM), LIDAR intensity data, height variation, difference of first and last echo and a grey scale aerial photo.

This paper presents a rule-based refinement for accuracy improvement of supervised classification using ML of high resolution first, last echo and intensity LIDAR data and co-registered line scanner bands such as aerial and near infra-red photo to categorise buildings, vegetation, cars and ground. Further height derived texture features are introduced such as the difference between first and last echo LIDAR, wavelets, mean, standard deviation and variance of height. It is commonly agreed that the accuracy of ML is limited. For this reason, contextual knowledge is used to improve the results by exploiting the inter-class relationship. The theoretical background is derived in Section 2. Section 3 presents the multivariate features and introduces the classes. Section 4 discusses the accuracy assessment approach, while results are discussed in Section 5. The paper concludes and outlines future work in Section 6.

2 Theoretical Background

2.1 Maximum Likelihood

The ML classifier is suitable for multivariate problems involving multiple simultaneously recorded bands (Tso and Mather, 2001). Fusing LIDAR data and co-registered bands yields a D dimensional feature space F of size $M \times N$. Given class Ω_i , $i \in \{1, 2, \dots, R\}$, where R is the number of classes, the probability of a pixel represented by its feature vector $\vec{f}_{x,y} \in F$ at position $(x, y) \in \{M \times N\}$ actually belonging to class Ω_i , is defined by the Bayes' rule (Jensen, 2001):

$$P(\Omega_i | \vec{f}_{x,y}) = \frac{P(\vec{f}_{x,y} | \Omega_i) P(\Omega_i)}{P(\vec{f}_{x,y})} \quad (1)$$

where $P(\Omega_i)$ and $P(\vec{f}_{x,y})$ are the prior probabilities of class Ω_i and the pixel represented by the feature vector $\vec{f}_{x,y}$, respectively. By estimating $P(\Omega_i | \vec{f}_{x,y})$, the lack of knowledge of these two prior probabilities can be bypassed by assuming both to be uniformly distributed for all bands (Tso and Mather, 2001). Therefore, Equation (1) can be reduced to

$$P(\Omega_i | \vec{f}_{x,y}) \propto P(\vec{f}_{x,y} | \Omega_i) \quad (2)$$

Another important assumption is that measured natural samples, such as F , are Gaussian distributed as stated by the *central limit theorem* (Duda *et al.*, 2001). Thus, $P(\vec{f}_{x,y} | \Omega_i)$ on the right

hand side of Equation (2) can be defined as (Duda *et al.*, 2001)

$$P(\vec{f}_{x,y} | \Omega_i) = \frac{1}{\sqrt[2]{2\pi} \sqrt{|C_i|}} e^{-\frac{1}{2}M} \quad (3)$$

with M as the squared Mahalanobis distance (Duda *et al.*, 2001)

$$M = (\vec{f}_{x,y} - \vec{\mu}_i)^T C_i^{-1} (\vec{f}_{x,y} - \vec{\mu}_i) \quad (4)$$

where $\vec{\mu}_i$ denotes the mean vector and C_i the $D \times D$ covariance matrix of class i , with $|C_i|$ and C_i^{-1} as its determinant and inverse, respectively (Thomas *et al.*, 1987). For notation purposes, T refers to transposing vector $(\vec{f}_{x,y} - \vec{\mu}_i)$. Equation (3) can be further simplified as

$$\ln(P(\vec{f}_{x,y} | \Omega_i)) = -\frac{D}{2} \ln(2\pi) - \frac{1}{2} \ln|C_i| - \frac{1}{2}M \quad (5)$$

Since the term $-\frac{D}{2} \ln(2\pi)$ is constant for all classes it can be discarded from Equation (5). Multiplying the remaining terms in Equation (5) with the factor -2 results in

$$\tilde{P}(\vec{f}_{x,y} | \Omega_i) = -2 \ln(P(\vec{f}_{x,y} | \Omega_i)) = \ln|C_i| + M \quad (6)$$

\tilde{P} as denoted in Equation (6) is calculated for each pixel and each class, normalised and finally minimised to obtain a labelled classified image.

2.2 Rule-based Classification Refinement

In order to reduce false positives in an ML classification, the spatial context of the classes and their relation among each other is incorporated. Considering the spatial resolution of LIDAR data of $0.5m$ per pixel as used in this study, the following rules are derived.

- The edges of the smallest objects, cars, are expected to take at least an area $0.5m^2$, equivalent to 2×1 pixels in either direction. Therefore, none of the detected objects can be smaller than two pixels. For this reason, single pixels are flipped into the majority of their adjacent pixels in a 3×3 neighbourhood.
- No object can be within a hard opaque object such as cars and buildings. *I.e.*, Cars cannot be within buildings and vice versa. Likewise, vegetation cannot be within cars or buildings. Misclassified pixels within those hard objects of an area $1.0m^2$ or 2×2 pixels within a 4×4 window can therefore be assumed to have actually the circumjacent pixel value.
- Conventional cars cannot exceed an area of $5.5m^2$ or 22 pixels. Thus, if the number of clustered pixels representing cars in a 5×5 environment is greater than 22, they are flipped to buildings as the probability is high that they are actually building pixels.
- Buildings do not occupy an area less than $5.5m^2$. Therefore, if the area of a building is less than $1.0m^2$ and the area of a car is greater than $1.0m^2$ in a 5×5 environment, building pixels are corrected to car pixels.

3 Feature Space and Classes

3.1 Feature Space

Seven co-registered bands and five height derived features are employed in this study as shown in Figure 1. The first echo LIDAR DSM in Figure 1(a) is a height profile of the scene and resolves *detached objects* (both hard and soft) such as buildings, sheds, trees, bushes and streets. The last echo in Figure 1(b) is a representation of the scene after the laser has penetrated soft objects such as deciduous trees and bushes. Both DSMs are reported to have a high vertical accuracy, especially in less hilly terrain (Huising and Pereira, 1998). Additionally, LIDAR data does not contain shadows. The difference between first and last echo in Figure 1(c) is often used to estimate vegetation height (Kraus and Pfeifer, 1998; Maas, 2005; St-Onge and Achaichia, 2001). LIDAR intensity data in Figure 1(d) represents the amount of power returning to the recording device and can distinguish among objects of similar height but different absorbing characteristics, such as water and tarmac. Bands of the aerial and NIR photos in Figure 1(e), 1(f) and 1(g), 1(h), respectively, contribute the characteristics of the spectrum which is beneficial in describing vegetation and other heat emitting objects. From the first echo DSM in Figure 1(a), the following height derived features are produced: The normalised magnitude of the details of wavelet response in Figure 1(i) is an ideal tool to separate global and local slopes of height (Bartels *et al.*, 2005). The mean in Figure 1(j), the standard deviation 1(k) as used in (Charaniya *et al.*, 2004) and the variance of height in Figure 1(l) are further height textures and are calculated in a window of 3×3 pixels.

The determinant $|C_i|$ of the covariance matrix of Ω_i has to be non-singular, *i.e.* $|C_i| \neq 0$ has to be satisfied to calculate C_i^{-1} . Consequently, the feature vectors have to be linear independent from each other (Thomas *et al.*, 1987) or redundant features have to be removed from or compensated in F (Tso and Mather, 2001). In this study, two channel pairs are identical: first, the red channel of the aerial photo and the green channel of the NIR photo and second, the green channel of the aerial photo and the blue channel of the NIR photo. Thus, the red channel of the aerial and the blue channel of the NIR photos are removed from F to achieve $|C_i| \neq 0$.

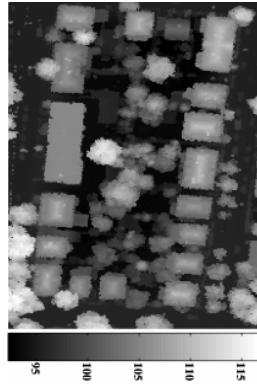
The most accurate classification result is obtained by estimating the best feature combination. Given $n = 12$ features, there are $2^n - 1 = 4095$ combinations to train the classifier to be validated in Section 4 using confusion matrices. The number of combinations without repetition $C_{(k)}^n$, while removing $k \in \{0, 1, \dots, 11\}$ features, is defined in Equation (7)

$$C_{(k)}^n = \binom{n}{k} = \frac{(n!)}{k!(n-k)!} \quad (7)$$

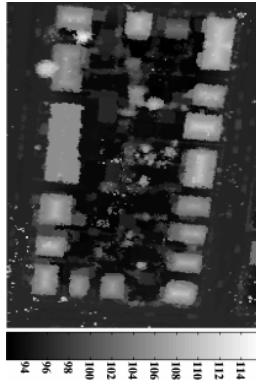
3.2 Classes

Four distinctive classes are defined: building, vegetation, car and ground. At this stage, trees, low vegetation and grass are defined as vegetation. High buildings and sheds are generalised to the class building. In the LIDAR community, the first two classes (building, vegetation) are referred as *detached objects* (Sithole and Vosselman, 2003) and the fourth class (ground), including the top layer soil, thin man-made layering such as asphalt or tarmac are defined as *bare earth* (Sithole and Vosselman, 2003). Defining the class car is justified because cars can be spotted in the high resolution data set. They are clearly visible in bands penetrating canopies of deciduous vegetation, *i.e.* the last echo LIDAR in Figure 1(b) and the line scanner data in Figures 1(e), 1(f), 1(g) and 1(h). Though not permanently situated in a scene, those cars should be removed when generating DTMs or nDSMs as discussed in Section 1.

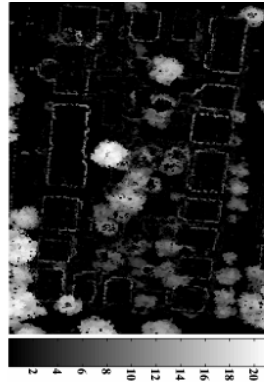
Compared to the total number of pixels $D \times M \times N = 6 \times 211 \times 356 = 450,696$ in F , each training sample has the following ratio: building 0.98%, vegetation 1.62%, car 0.28% and ground 2.44%. Hence, the sample sizes for *detached objects* and *bare earth* are less than 2% and 3%, respectively. Note the small portion of samples for class car as they are limited in the scene. All in all, the training sample sizes comprise a fraction of 5.31% in F .



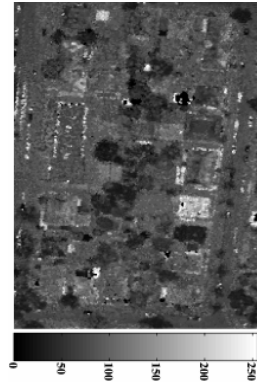
(a) First echo LIDAR in metres



(b) Last echo LIDAR in metres



(c) Difference first/ last echo LIDAR in metres



(d) LIDAR intensity data



(e) Green channel of aerial photo



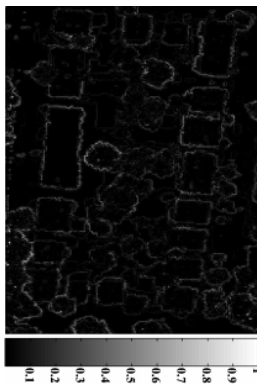
(f) Blue channel of aerial photo



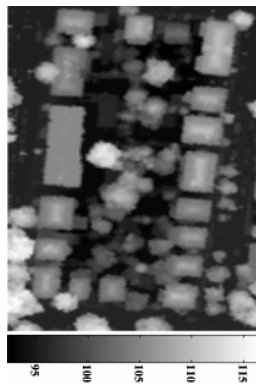
(g) Red channel of NIR photo



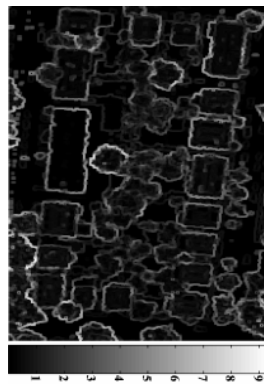
(h) Green channel of NIR photo



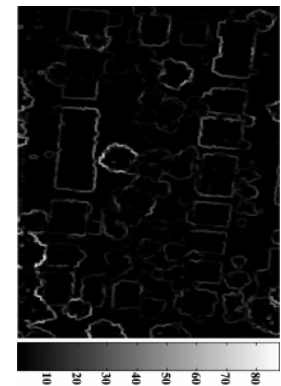
(i) Normed wavelet response to height in first echo LIDAR



(j) Mean of height in first echo LIDAR in metres



(k) Standard derivation of height in first echo LIDAR in metres



(l) Variance of height in first echo LIDAR in square metres

Figure 1. LIDAR data, co-registered bands and height derived features

4 Accuracy Assessment

Based on ground truth, a confusion matrix can be produced to assess the classification accuracy. The overall accuracy θ_1 can be estimated by the ratio of the sum of its main diagonal and the total number of classified pixels (Tso and Mather, 2001) as defined in Equation (9). Focusing individual classes, producer's and user's accuracy can be estimated. The producer's accuracy evaluates the proportion of correctly classified pixels from the collected class samples, and is estimated for each class by the ratio of its diagonal element and the sum of its columns (Tso and Mather, 2001). The user's accuracy stands for the proportion of pixels which were correctly assigned to one particular class, estimated for each class by the ratio of the diagonal element and the sum of its row (Tso and Mather, 2001). The multivariate measure κ (kappa) examining the whole confusion matrix, is defined as (Tso and Mather, 2001)

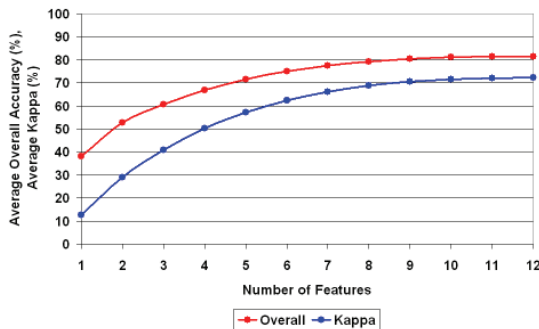
$$\kappa = \frac{\theta_1 - \theta_2}{1 - \theta_2} \quad (8)$$

$$\theta_1 = \frac{1}{(M \times N)} \sum_{r=1}^R x_{rr} \quad (9)$$

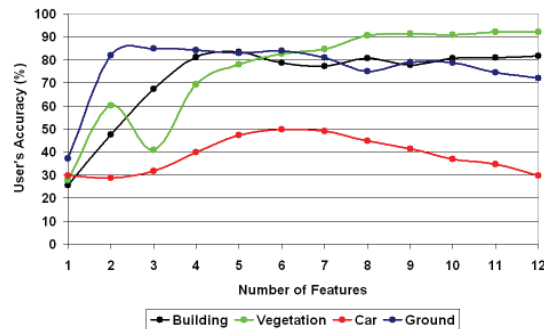
$$\theta_2 = \frac{1}{(M \times N)^2} \sum_{r=1}^R x_{r+} \cdot x_{+r} \quad (10)$$

where x_{rr} represents the diagonal elements of the confusion matrix at index r and x_{r+} and x_{+r} are the sum of row r and column r , respectively.

Figure 2 depicts an accuracy analysis for all 4095 combinations. As expected, the average overall accuracy and κ increase with the number of features as shown in Figure 2(a). With feature space F with $D \geq 9$, they are greater than 80% and 70%, respectively. The conclusion would be to take the richest feature space F with $D = 12$. However, as class car is the most difficult one to be classified, its user's accuracy has to be further analysed. Figure 2(b) shows the user's accuracy of all classes for each feature space size where in each case the maximum of the user's accuracy of class car is selected. The performance of class car is at its maximum of about 50% at F with $D = 6$. Note, that even an increase of features does not yield a better classification result: 12 feature layers produce a user's accuracy of approximately 30%. Thus, this feature space with $D = 6$ is chosen as it produces the best user's accuracy of class car: green and blue channel of the aerial photo; red and green channel of the NIR photo, and two height derived features from the first echo LIDAR DSM, *i.e.* its mean and standard deviation.



(a) Average overall accuracy and κ



(b) User's accuracy from the perspective of class car

Figure 2. Accuracy of feature combinations

5 Results and Discussion

Figure 3 depicts the Maximum Likelihood classification result, the successive rule-based improved result and the ground truth. It can be seen in Figure 3(a) that *detached objects*, i.e. buildings (black) and vegetation (green), as well as cars (red) are categorised. Even grass is correctly classified despite its insignificance towards the height information in LIDAR data. Hence, as expected, the bands of the line scanner have more impact on detecting grass and low vegetation. The class ground as *bare earth* (white) also shows a good response to the classifier despite its various appearance as top layer soil, thin man-made layering such as asphalt or tarmac. For the results in Figure 3(a) from the confusion matrix in Table 1 (left values), the overall accuracy is about 81.61% and $\kappa \approx 72.64\%$. However, it can also be seen that there is a number of false positives for each class due to the ambiguity of the classes among the features in F as discussed in Section 3.1. Another reason is that class car is insufficiently represented by their samples as there are only a limited number of cars present in the scene. Consequently, errors have a stronger impact on the accuracy of this class than on other classes. These misclassified pixels are corrected by employing spatial inter-class information as introduced in Section 2.2. The improved result is shown in Figure 3(b) and from confusion matrix in Table 1 (right values), an overall accuracy of approximately 83.11% and κ of about 72.64% are achieved.

Table 1. Confusion matrices before (left values) and after spatial refinement (right values)

		Ground Truth				
		Building	Vegetation	Cars	Ground	Total
Classified Data	Building	15902/ 16743	1178/ 874	396/ 156	2738/ 2441	20214/ 20214
	Vegetation	2838/ 2995	23129/ 23456	171/ 38	1803/ 1452	27941/ 27941
	Car	79/ 61	37/ 0	475/ 483	362/ 409	953/ 953
	Ground	906/ 1358	2668/ 2732	637/ 173	21797/ 21745	26008/ 26008
	Total	19725/ 21157	27012/ 27062	1679/ 850	26700/ 26047	75116/ 75116

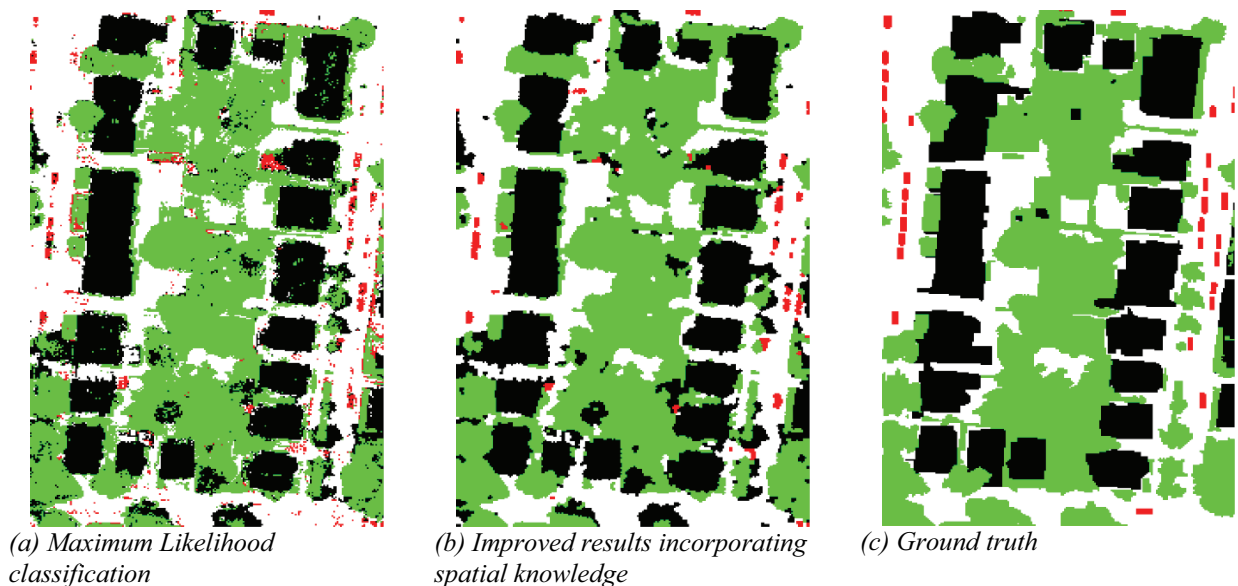


Figure 3. Results and ground truth: building (black), vegetation (green), car (red) and ground (white)

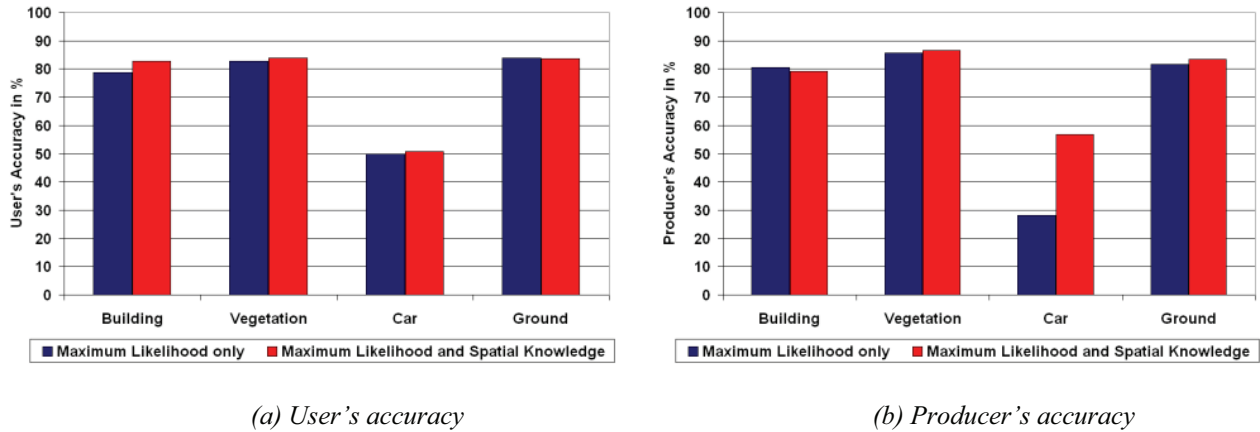


Figure 4. User's and producer's accuracies

Figure 4 depicts the user's and producer's accuracies after Maximum Likelihood classification before and after rule-based refinement in Figures 4(a) and 4(b), respectively. The producer's and user's accuracies of the classes building, vegetation and ground range between 78.67% minimum and 86.68% maximum and remain fairly stable within this range. However, the producer's accuracy of class car is improved by about 28.53% by incorporating spatial knowledge as described in Section 2.2.

6 Conclusions and Future Work

In this paper, a rule-based improvement of parametric supervised classification based on Maximum Likelihood applied to LIDAR data using data fusion is presented. To build up a feature space, first and last echo LIDAR, intensity data, several co-registered bands such as aerial and near infra-red photos and further height derived texture features are employed such as the difference between first and last echo LIDAR, wavelets, mean, standard deviation and variance of height. Four classes are classified and their individual accuracy is assessed. The results show that *detached objects* (buildings, vegetation), *bare earth* and cars are correctly classified. False positives are significantly reduced by using a rule-based approach which incorporates spatial information and inter-class relationships. Outstanding is the producer's accuracy improvement of class car which is outperformed by nearly 30% compared to a simple Maximum Likelihood classification.

For future work, the classification approach will be improved by introducing further uncorrelated features. Currently, the use of Gabor filters to obtain further height derived features is explored. Further investigation will focus on expanding the developed algorithm to different scenes in order to categorise more classes.

7 Acknowledgements

The project is RETF funded by the University of Reading. The authors would like to thank TopoSys GmbH, and the Stadt Mannheim, Germany, for LIDAR data supply.

References

- Bartels, M., Wei, H. and Mason, D. C., 2005. **Wavelet packets and co-occurrence matrices for texture-based image segmentation**. *IEEE International Conference on Advanced Video and Signal-Based Surveillance* 1, pp. 428–433.
- Charaniya, A. P., Manduchi, R. and Lodha, S. K., 2004. **Supervised parametric classification of aerial LIDAR data**. *IEEE Conference on Computer Vision and Pattern Recognition Workshop* pp. 30 – 38.
- Duda, R. O., Hart, P. E. and Stork, D. G., 2001. **Pattern classification**. *New York: Wiley*.
- Haala, N. and Brenner, C., 1997. **Generation of 3D city models from airborne laser scanning data**.

EARSEL Workshop on LIDAR remote sensing of land and sea pp. 105–112.

Haala, N. and Brenner, C., 1998. **Fast production of virtual reality city models.** *International Archives of Photogrammetry and Remote Sensing* 32(4), pp. 77–84.

Haala, N., Brenner, C. and Staetter, C., 1998. **An integrated system for urban model generation.** *Proceedings ISPRS Congress Commission II, Working Group 6* pp. 96–103.

Huising, E. J. and Pereira, L. M. G., 1998. **Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications.** *ISPRS Journal of Photogrammetry & Remote Sensing* 53, pp. 245–261.

Jensen, F. V., 2001. **Bayesian Networks and Decision Graphs.** *New York: Springer-Verlag.*

Kraus, K. and Pfeifer, N., 1998. **Determination of terrain models in wooded areas with airborne laser scanner data.** *ISPRS Journal of Photogrammetry & Remote Sensing* 53, pp. 193–203.

Maas, H.-G., 1999a. **Closed solutions for the determination of parametric building models from invariant moments of airborne laserscanner data.** *ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery.*

Maas, H.-G., 1999b. **Fast determination of parametric house models from dense airborne laserscanner data.** *International Workshop on Mobile Mapping Technology.*

Maas, H.-G., 1999c. **The potential of height texture measures for the segmentation of airborne laserscanner data.** *Fourth International Airborne Remote Sensing Conference and Exhibition / 21st Canadian Symposium on Remote Sensing.*

Maas, H.-G., 2005. **Akquisition von 3D-GIS Daten durch Flugzeuglaserscanning.** *Kartographische Nachrichten* 55(1), pp. 3–11.

Maas, H.-G. and Vosselman, G., 1999. **Two algorithms for extracting building models from raw laser altimetry data.** *ISPRS Journal of Photogrammetry & Remote Sensing* 54, pp. 153–163.

Rottensteiner, F. and Briese, C., 2003. **Automatic generation of building models from lidar data and the integration of aerial images.** *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences of the ISPRS* 34(3/W13), pp. 174 – 180.

Sithole, G. and Vosselman, G., 2003. **Automatic structure detection in a point cloud of an urban landscape.** *Proceedings of 2nd Joint Workshop on Remote Sensing and Data Fusion over Urban Areas (Urban 2003)* pp. 67– 71.

St-Onge, B. A. and Achaichia, N., 2001. **Measuring forest canopy height using a combination of lidar and aerial photography data.** *International Archives of Photogrammetry and Remote Sensing* 34(3/W4), pp. 131 – 137.

Thomas, I. L., Benning, V. M. and Ching, N. P., 1987. **Classification of Remotely Sensed Images.** *Bristol: Adam Hilger.*

Tso, B. and Mather, P. M., 2001. **Classification Methods for Remotely Sensed Data.** *London: Taylor & Francis.*

Vu, T. T. and Tokunaga, M., 2001. **Wavelet and scale-space theory in segmentation of airborne laser scanner data.** *22nd Asian Conference on Remote Sensing.*

Weidner, U. and Förstner, W., 1995. **Towards automatic building extraction from high resolution digital elevation models.** *ISPRS Journal of Photogrammetry & Remote Sensing* 50(4), pp. 38 – 49.