

# Maximum Likelihood Classification of LIDAR Data incorporating multiple 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 supervised classification algorithm based on Maximum Likelihood is presented using high resolution first, last echo and intensity LIDAR data and co-registered line scanner bands such as aerial photos and near infra-red photos. The issues regarding feature and class selection as well as accuracy assessment are addressed in this paper. The presented results show the suitability of the classification approach for 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 more and more by the public sector and commercial world since the early 1990s [1]. 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 [2].

In one of the early investigations on height in LIDAR data, Weidner and Förstner [3] separated ground and object points in estimating a normalised Digital Surface Model (nDSM) by subtracting a morphologically filtered Digital Terrain Model (DTM) from the original Digital Surface Model (DSM). An approach to model buildings from LIDAR data in a less sloped area was developed by Maas

*et al.* [4]. First, a texture based segmentation employing Maximum Likelihood (ML) applied to LIDAR data and two further height derived bands [5, 6] and second, the actual building modelling from the point cloud [7].

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 by Brenner and Haala *et al.* [8, 9] by reconstructing buildings, facades and vegetation using multiple data sources, such as near infra-red and terrestrially captured digital images [10]. Building ground planes were estimated with a 2D GIS and a cadastral map. Building reconstruction by fusing LIDAR data and aerial photos was presented by Rottensteiner *et al.* [11] who detected building regions in ungridded data and roofs using a curvature-based segmentation technique. Additional planar faces were estimated with aerial photos. A noise robust texture-based segmentation approach for LIDAR data using wavelet packets, co-occurrence matrices and normalised modified histogram thresholding has been proposed in [12]. A supervised parametric classification algorithm based on the Gaussian mixture model and Expectation Maximisation was used by Charaniya *et al.* [13] to classify roads, roofs, trees and grass in LIDAR by fusing an nDSM, intensity data, height variation, difference of first and last echo, a grey scale aerial photo and an available Digital Elevation Model (DEM).

This paper presents 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 photos to categorise buildings, vegetation, cars and ground. The theoretical background is derived in Section 2. Section 3 presents the multivariate features and introduces the classes. In Section 4, achieved results and accuracy are discussed. The paper concludes and outlines future work in Section 5.

## 2. Theoretical Background

The ML classifier is suitable for multivariate problems involving multiple simultaneously recorded bands [14]. Fusing LIDAR data and co-registered channels yields a  $D$  dimensional feature space  $\mathbf{F}$  of the size  $M \times N$ . Given class  $\Omega_i$ ,  $i \in \{1, 2, \dots, R\}$ , where  $R$  is the number of classes, the probability of a pixel represented by its feature vector  $\vec{f}_{x,y} \in \mathbf{F}$  at position  $(x, y) \in \{M \times N\}$  actually belongs to class  $\Omega_i$ , is defined by the Bayes' rule [15]:

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

where  $P(\Omega_i)$  and  $P(\vec{f}_{x,y})$  are the prior probabilities of class  $\Omega_i$  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 channels [14]. 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  $\mathbf{F}$ , are Gaussian distributed as stated by the *central limit theorem* [16]. Thus,  $P(\vec{f}_{x,y} | \Omega_i)$  on the right hand side of Equation (2) can be defined as [16]

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

with  $M$  as the Mahalanobis distance [16]

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

where  $\vec{\mu}_i$  denotes the mean vector and  $\mathbf{C}_i$  the  $D \times D$  covariance matrix of class  $\Omega_i$ , with  $|\mathbf{C}_i|$  and  $\mathbf{C}_i^{-1}$  as its determinant and inverse, respectively [17]. For notation purposes,  $T$  refers to transposing vector  $(\vec{f}_{x,y} - \vec{\mu}_i)$ . Equation (3) can be further simplified as

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

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

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

$\tilde{P}$  as denoted in Equation (6) is calculated for each pixel and each class. Finally, after normalising,  $\tilde{P}$  has to be *minimised* to obtain a labelled classified image.

## 3. Feature Space and Classes

### 3.1. Feature Space

As depicted in Figure 1, five co-registered channels are employed in this study with a spatial resolution of  $0.5m$  per pixel: first, last echo and intensity LIDAR data, an aerial photo and a near infra-red (NIR) photo. All objects to be classified are resolved, even cars 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(d) and 1(e). When selecting the features, the determinant  $|\mathbf{C}_i|$  of the covariance matrix of  $\Omega_i$  has to be non-singular, *i.e.*  $|\mathbf{C}_i| \neq 0$  has to be satisfied in order to calculate  $\mathbf{C}_i^{-1}$ , *i.e.* the feature vectors have to be linear independent from each other [17]. Otherwise, redundant features have to be removed from or compensated in  $\mathbf{F}$  [14]. In this study, the red and the green channel of the aerial photo are even equal to their equivalents of the NIR photo. Thus, both photos are converted each into grey scale of 256 quantisation levels to achieve  $|\mathbf{C}_i| \neq 0$ .

### 3.2. Classes

Four distinctive classes are defined: building, vegetation, car and ground. At this stage, trees, low vegetation and grass are summarised as vegetation. High buildings and sheds are generalised as buildings. In the LIDAR community, the first two classes (building, vegetation) are referred as *detached objects* [18] and the fourth class (ground), including the top layer soil, thin man-made layering such as asphalt or tarmac are defined as *bare earth* [18]. The choice of defining the class car is justified because cars can be spotted due to the high resolution data set. Though those cars are not permanently resided in a scene, these objects still have to be removed when it comes to generating DTMs or nDSMs as discussed in Section 1. Table 1 lists the total number of sample pixels  $N_i$  for each class  $\Omega_i$  and its ratio towards the total number of pixels  $N_k = D \times M \times N = 5 \times 211 \times 356 = 375580$  in  $\mathbf{F}$ . For each class, this ratio is less than 2.45% for *detached objects* and *bare earth*. Note the small portion of samples for the class car of about 0.39% due to the limited number of cars in the scene. In general, the sample sizes are rather small considering the results in Section 4. They comprise in total a fraction of  $\mathbf{F}$  of approximately 6.07%.

Class	Building	Vegetation	Car	Ground	Total
$N_i$	3680	8475	1480	9150	22785
$N_i/N_k$	0.98%	2.26%	0.39%	2.44%	6.07%

**Table 1. Ratio of classes and feature space**

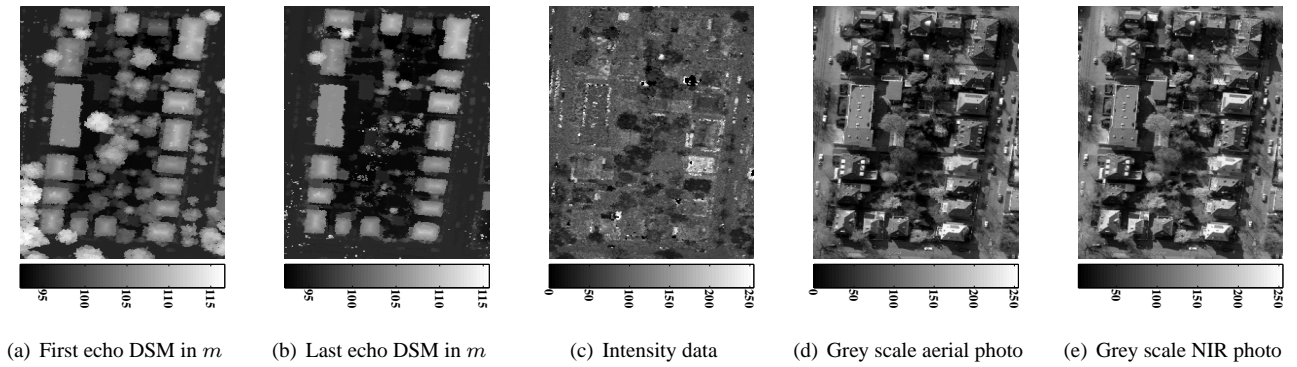


Figure 1. Fused features used in the classification

#### 4. Results and Accuracy Assessment

Figure 2(a) depicts the labelled classified scene. *Detached objects*, *i.e.* building (black) and vegetation (green), are clearly categorised with a limited number of false positives. Even grass is correctly classified despite their insignificance towards the height information in LIDAR data. Hence, as expected, the bands of the visible spectrum 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. However, as expected, though the class car (red) is identified, there are more false positives than the other three classes. The reason is the ambiguity of these pixels among the features in  $F$  as discussed in Section 3.1.

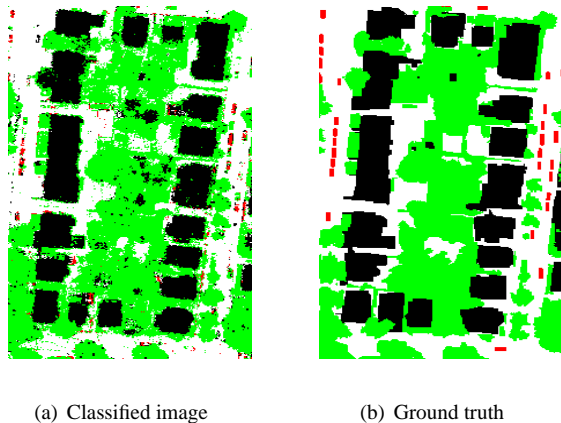


Figure 2. Classification results: building (black), vegetation (green), car (red) and ground (white)

From the results and the ground truth as depicted in Figure 2, a confusion matrix, as listed in Table 2, is produced. The rows of the confusion matrix represent the labels assigned by the classifier whereas the columns indicate the ground truth. Noteworthy is the significant main diagonal representing all classified pixels for each class which agree with the ground truth. These pixels have been correctly classified. Thus, from the confusion matrix, the overall accuracy can be estimated by the ratio of the sum of the main diagonal and the total number of classified pixels [14].

		Ground Truth				
		B	V	C	G	Total
Classified Data	B	<b>16326</b>	2178	224	1486	20214
	V	1709	<b>24635</b>	62	1535	27941
	C	111	252	<b>263</b>	327	953
	G	667	4827	257	<b>20257</b>	26008
	Total	18813	31892	806	23605	<b>75116</b>

Table 2. Confusion matrix (B = building, V = vegetation, C = car and G = ground)

The overall accuracy alone does not distinguish among the individual classes. 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 column [14]. 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 [14]. Table 3 lists producer's and user's accuracies for each individual class.

The multivariate measure  $\kappa$  taking the whole confusion matrix into account, is defined as [14]

$$\kappa = \frac{\theta_1 - \theta_2}{1 - \theta_2} \quad (7)$$

where  $\theta_1$  is the quotient of the sum of all diagonal elements of the confusion matrix and the total number of pixels

Class	Producer	User
Building	86.78%	80.77%
Vegetation	77.25%	88.17%
Car	32.63%	27.60%
Ground	85.82%	77.89%

**Table 3. Producer's and user's accuracy**

in the confusion matrix.  $\theta_2$  denotes the sum of the products of the marginal totals of the corresponding row and column in the confusion matrix divided by the square of the total number of pixels in the confusion matrix [14].

To summarise the results, an overall accuracy of approximately 81.85% and a  $\kappa$  of about 72.73% is achieved. The class with the best producer's accuracy is building with 86.78%, followed by ground and vegetation with 85.82% and 77.25%, respectively. However, class car has produced an accuracy of 32.63%. Regarding the user's accuracy, classes vegetation, building and ground are correctly classified with 88.17%, 80.77% and 77.89%, respectively. Class car, however, breaks ranks with 27.60% which is due to its ambiguity among the several bands as discussed in Section 3.1 and that it is insufficiently represented by its samples. Thus, errors have a stronger impact on the accuracy of this class than on other categories. To solve this problem, more representative samples have to be collected to increase the performance of class car. For this, a larger area beyond the scope of this study will have to be taken into consideration.

## 5. Conclusions and Future Work

In this paper, a parametric supervised classification algorithm based on Maximum Likelihood applied to LIDAR data using data fusion is presented. Several bands such as first, last echo and intensity LIDAR data and co-registered line scanner bands such as aerial and near infra-red photos are employed to build up a feature space. Four classes are classified and their individual accuracy is assessed. The results show that *detached objects* (buildings, vegetation) and *bare earth* are correctly classified up to 88.17%. The performance of class car, however, shows potential for further improvement due to its ambiguous appearance within the different features and the limited number of samples which can be collected compared to other classes.

For future work, the classification approach will be improved by introducing further uncorrelated features. Currently, the use of Gabor filters and wavelets to obtain further height derived features is explored. Further investigation will focus on the influence of each individual band on the results to improve the algorithm's performance.

## Acknowledgements

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

## References

- [1] H.-G. Maas. Akquisition von 3D-GIS Daten durch Flugzeuglaser-scanning. *Kartographische Nachrichten*, 55(1):3–11, 2005.
- [2] E. J. Huising and L. M. Gomes Pereira. Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications. *ISPRS Journal of Photogrammetry & Remote Sensing*, 53:245–261, 1998.
- [3] U. Weidner and W. Förstner. Towards automatic building extraction from high resolution digital elevation models. *ISPRS Journal of Photogrammetry & Remote Sensing*, 50(4):38 – 49, 1995.
- [4] H.-G. Maas and G. Vosselman. Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of Photogrammetry & Remote Sensing*, 54:153–163, 1999.
- [5] H.-G. Maas. Fast determination of parametric house models from dense airborne laserscanner data. *International Workshop on Mobile Mapping Technology*, 1999.
- [6] H.-G. Maas. 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*, 1999.
- [7] H.-G. Maas. 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*, 1999.
- [8] N. Haala and C. Brenner. Generation of 3D city models from airborne laser scanning data. *EARSEL Workshop on LIDAR remote sensing of land and sea*, pages 105–112, 1997.
- [9] N. Haala and C. Brenner. Fast production of virtual reality city models. *International Archives of Photogrammetry and Remote Sensing*, 32(4):77–84, 1998.
- [10] N. Haala, C. Brenner, and C. Staetter. An integrated system for urban model generation. *Proceedings ISPRS Congress Commission II, Working Group 6*, pages 96–103, 1998.
- [11] F. Rottensteiner and C. Briese. 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):174 – 180, 2003.
- [12] M. Bartels, H. Wei, and D. C. Mason. Wavelet packets and co-occurrence matrices for texture-based image segmentation. *IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 1:428–433, 2005.
- [13] A. P. Charaniya, R. Manduchi, and S. K. Lodha. Supervised parametric classification of aerial lidar data. *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, pages 30 – 38, 2004.
- [14] B. Tso and P. M. Mather. *Classification Methods for Remotely Sensed Data*. London: Taylor & Francis, 2001.
- [15] F. V. Jensen. *Bayesian Networks and Decision Graphs*. New York: Springer-Verlag, 2001.
- [16] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. New York: Wiley, 2001.
- [17] I. L. Thomas, V. M. Benning, and N. P. Ching. *Classification of Remotely Sensed Images*. Bristol: Adam Hilger, 1987.
- [18] G. Sithole and G. Vosselman. 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)*, pages 67–71, 2003.