

A SEMI-AUTOMATIC APPROACH TO OBJECT EXTRACTION FROM A COMBINATION OF IMAGE AND LASER DATA

S. A. Mumtaz^{a,*}, K. Mooney^a

^a Dept. of Spatial Information Sciences, The Dublin Institute of Technology, Bolton Street, Dublin 1, Ireland
(salman.mumtaz, kevin.mooney)@dit.ie

Commission III, WG III/4

KEY WORDS: LiDAR, Object Extraction, Data fusion, Buildings, Trees, Roads

ABSTRACT:

The aim of the authors' research is to develop an automated or semi-automated workflow for the extraction of objects such as buildings, trees and roads for noise mapping and road safety purposes. The workflow must utilise national airborne spatial data available throughout the country and be capable of robust incorporation in the noise modelling systems of a national roads authority. This paper focuses on the extraction of multiple objects by fusing data captured by two independent sensors, namely the Leica ADS40 aerial camera and the Leica ALS50 airborne laser scanner (LiDAR). A workflow has been developed for the extraction of objects utilizing height values from a normalised DSM generated using LiDAR or aerial images, multiple LiDAR echo data and NDVI (Normalized Difference Vegetation Index) data computed from multispectral ADS40 data.

Major tasks include LiDAR data classification, segmentation and its integration with the information extracted from aerial images. Buildings are extracted first and this facilitates the extraction of other objects. Preliminary results of this semi-automated process indicate high completeness rates for buildings trees and roads but 60% quality rates (e.g. buildings). Quality may be improved by manual extraction of small objects but continuing research is focussed on reducing reliance on such manual intervention.

1. INTRODUCTION

The National Roads Authority (NRA) in Ireland is responsible for generating noise maps in the environment of roads used by more than 8220 vehicles per day. According to the EU noise directive this exercise must be repeated every five years. Inputs for generating the noise maps include terrain model, location and dimension of buildings, trees, noise barriers and the geometric properties of roads. Capturing this data using field surveys or digital images is time consuming and expensive, especially if the same exercise must be repeated every five years. It is the intention of this work that all required objects be extracted using automatic or semiautomatic techniques from LiDAR and aerial image data of the type available from the National Mapping Agency of Ireland, OSi (Ordnance Survey of Ireland). Later the extracted information can be easily combined and analyzed along with noise data in a GIS system. For noise mapping, building detail or tree models are not required. Buildings or trees boundaries with height information are sufficient.

High resolution image and LiDAR sensors (ADS40 & ALS50) were used to capture the data for a part of County Sligo in the northwest of Ireland. Digital images were captured in April 2007 with a ground resolution of 15 cm. LiDAR data were captured separately in May 2007 at a flying height of 1241 m with a swath width of 800 m, resulting in an average point density of approximately 2 points/m². The ALS50 sensor recorded position, multiple echoes and intensity of the returning pulse.

The area selected for processing is about 3 km² and is covered by a single image strip and four LiDAR strips. This eliminates the necessity for bundle block adjustment and ground control

point acquisition. The reason for relying completely on direct geo-referencing in this research is the fact that in many situations ground control points may not be available. Strip adjustment of the LiDAR data was performed using the Terra Match application from TerraSolid.

1.1 Motivation

In recent years, research on automated object extraction has increased because of the increased use of GIS (Geographical Information Systems) with the consequential need for data acquisition and update.

Digital Photogrammetry is considered to be one of the most precise methods for capturing large scale data for GIS analysis from high resolution aerial images. However, it requires significant resources to digitize all objects of interest. As detailed high resolution digital images are regularly acquired as part of the national programme of OSi, it is considered important to develop automatic or semi-automatic techniques to exploit their potential for applications such as noise modelling involving the extraction of objects such as buildings, trees and roads.

LiDAR can provide high density 3D point clouds in a very short time with acceptable horizontal and high vertical accuracy. OSi also acquires national LiDAR data using the ALS50 sensor from Leica Geosystems. The availability to the national roads authority of Ireland of both of these high resolution data sources provides the impetus for this research.

However, the development in sensor technology is far more rapid than the advancements in automatic or semi automatic object extraction. Moreover there is still a large gap between

* Salman Ali Mumtaz

the theoretical work on fully automated object extraction and practical applications of the same (Mayer 2008). Success in automatic object extraction will also help in determining changes that occur between noise surveys (5 years) by comparing the extracted objects at the different epochs and should speed up the updating process of the GIS database.

1.2 Related Work

LiDAR has been extensively used for the generation of both DSMs (Digital Surface Model) and DTMs (Digital Terrain Model). Different classification methods have been used for the classification of terrain and off terrain points (Sithole & Vosselman, 2003). Different approaches have been used for the detection and reconstruction of buildings from LiDAR data (Brunn & Weidner, 1997 and Clode et al., 2004). Haitao et al. (2007) used aerial images and LiDAR data for land cover classification based on SVM (Scalable Vector Machine). Haala & Brenner (1999) also used the combination of multispectral imagery and LiDAR data for the extraction of buildings, trees and grass covered areas. Trees and grass covered areas were classified easily from the multispectral imagery but were found difficult to separate. Similarly, trees and buildings were separated using height differences between DSM and DTM. Both data sources were combined in order to identify the three classification types. Rottensteiner et al. (2004) classified land cover into four different classes namely, buildings, trees, grass lands, and bare soil. This was achieved by combining LiDAR data and multispectral images. Prior to performing building detection by data fusion based on the theory of Dempster-Shafer, the LiDAR data was pre-processed to generate a DTM. For the extraction of roads different information sources such as multispectral images from airborne and space borne sensors were used. Clode et al. (2007) used only LiDAR for road extraction. Despite encouraging results, there are still many fundamental questions to be answered for road extraction in urban areas (Mayer et al., 2008).

2. METHOD

The method under investigation is based on a workflow that identifies and classifies buildings, trees and other objects by fusing the information from LiDAR and aerial image data. This information includes the normalised digital surface model (NDSM) and multiple echoes from the LiDAR data together with Normalized Difference Vegetation Index (NDVI) data generated from the airborne imagery. The method is depicted in Figure 1. Three major task groups may be identified, namely the image group, LiDAR group and object extraction group.

3. WORKFLOW

Within the image group of tasks, the first step is to produce orthophotos for each spectral channel of the ADS40 sensor, i.e. Red (R), Green (G), Blue (B) & Near Infrared (NIR). For these orthophotos, the required DSM can be created relatively automatically using the panchromatic forward and backward image data captured by the ADS40 sensor. The effect of DSM quality on orthophoto generation is shown in Figure 2. The upper part of the figure shows a rectified building using a DSM generated by aerial images and the lower part shows the same building rectified using a DSM from LiDAR data.

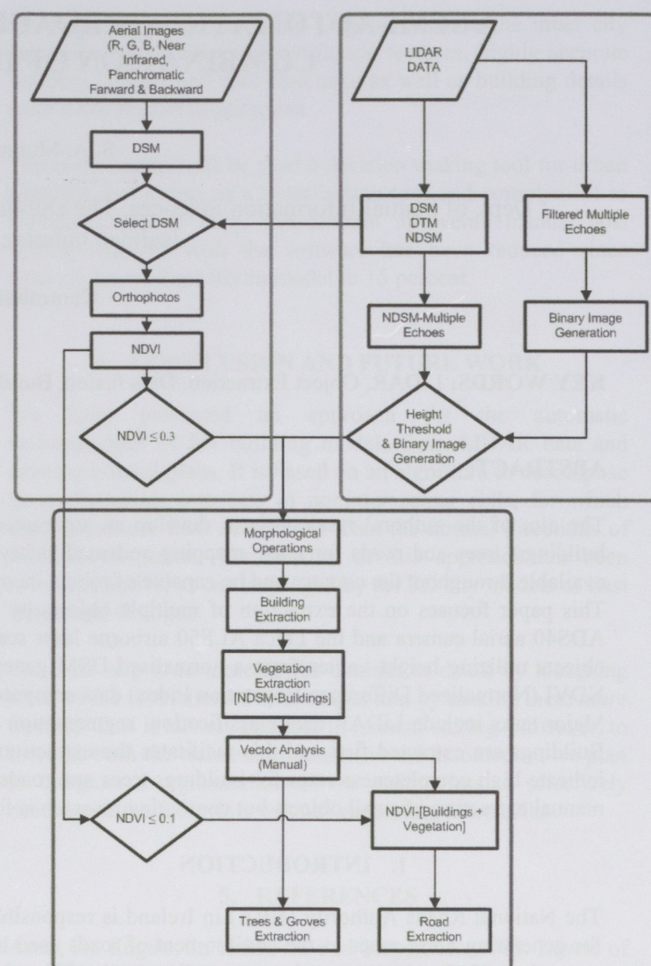


Figure 1: Method Workflow



Figure 2: Effect of DSM on Building Rectification

In case of the building illustrated in Figure 2, a DSM created from LiDAR data with a resolution of 0.5 m (the lower example) provided sharper edges compared to that generated from the image DSM and was used in the generation of the orthophotos. As a prerequisite to this step the quality of the registration between the airborne imagery and the LiDAR data must be verified. The Nearest Neighbourhood method was used as a sampling method for orthophoto generation. Separate

orthophotos were generated from the R, G, B and NIR channels (a true colour orthophoto was also generated) and the NDVI was calculated using the following formula:

$$\text{NDVI} = (\text{NIR} - \text{R}) / (\text{NIR} + \text{R})$$

NDVI values range from -1 to +1 which suggests that if the pixel value is close to -1 it does not belong to healthy vegetation or vice versa. As a result, NDVI data could assist in separating vegetation from buildings in a DSM.

Figure 3 shows the effect of sun position on selecting an NDVI threshold to separate buildings from vegetation. In the shadow area NDVI values are larger than the portion of the building directly facing the sun. A larger threshold value of 0.3 was selected to differentiate between buildings and vegetation. Because of this large threshold value some vegetation also appears with the buildings.

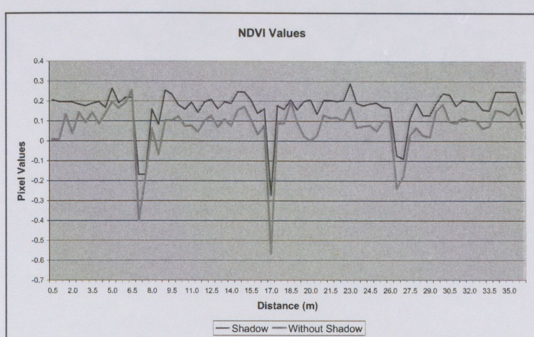


Figure 3: NDVI Threshold for Buildings

Within the LiDAR group of tasks, the first step is the generation of a DSM and DTM from the LiDAR data (TerraSolid software was used). In order to get the absolute height of the objects the DTM was subtracted from the DSM to give the NDSM. A further refinement of the NDSM can then be achieved by making use of multiple LiDAR echo data. These occur from building edges and trees. Figure 4 (Clode et al., 2005) shows how the laser beam interacts with building edges and trees.

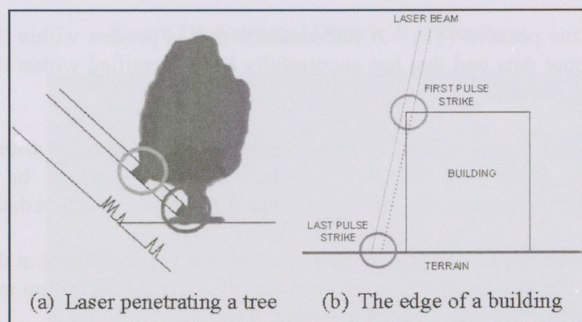


Figure 4: Laser Interactions

Firstly, the filtered multiple echoes (figure 5) were converted into an image. Gaps between pixels of less than 3 metres were filled and a binary image was generated. Selecting a value for gap filling depends on the density of the original point data. If the density is high a small value can serve the purpose but it should not be too high that it causes individual trees close to each other to merge.

The separation of multiple echo data (Figure 5) from the NDSM, by multiplication by the binary data, results in data only from those objects that record a single reflection. These include buildings and other solid objects but also vegetation that returned single echoes.



Figure 5: Filtered Multiple Echoes

The next step was to apply a height threshold of 2.5m to the NDSM to eliminate objects such as hedges, cars etc and the resultant NDSM containing buildings, vegetation and other tall objects was converted to a binary image. All pixels having a value lower than or equal to 2.5 m were assigned a zero value and the remainder a value of one (Figure 6).

A morphological operation such as closing and opening was used for filling small gaps in the binary image. Care should be taken as too many repetitions can result in rounding of the sharp building edges and loss of important detail.

This binary image contains pixels that belong to buildings and remaining trees and needs further classification. This was achieved by introducing the NDVI image described as part of the image group of tasks.



Figure 6: Binary Image with Height Threshold

The NDSM and the NDVI images were combined and the maximum likelihood classification method was used for the

extraction of buildings (Figure 7). The area was calculated for each building after conversion to vector format and used as a threshold for separating main buildings from smaller structures in front or behind the main buildings. These areas also include trucks and vans on the roads or parked near to buildings and are detected as buildings. These areas were retained and classified later into buildings, vegetation or other objects.



Figure 7: Extracted Buildings

To assign each building an individual height, building centroids were determined. For each centroid, heights were determined from the DSM by bilinear interpolation and the same process was used for determining vegetation height.

For the purpose of vegetation extraction, the final building data was subtracted from the NDSM. This resulted in vegetation present in the NDSM layer that is higher than 2.5m. However, the filtered multiple echo data (Figure 5) was also processed further. First, intermediate and last echoes represent reflections from the edges of buildings and trees. Once buildings were classified, building boundaries were used as an input to remove all multiple echo points that belong to building edges. Multiple reflections from large trees, together with compactness (area/perimeter^2) were used to classify large single trees and groves (Figure 8).



Figure 8: Extracted Trees and Groves

The remaining objects to be classified were roads which are part of the generated DTM. NDVI data below a threshold of 0.1 and the previously classified objects were used. Using the threshold eliminates most of the area having vegetation but does not help much in the areas with barren land. Their spectral signature value is also very close to the roads. Even for the roads the reflection value is not constant. It varies with age and type of material used in the road surface. Previously extracted buildings and vegetation were subtracted from the NDVI to extract road candidates.

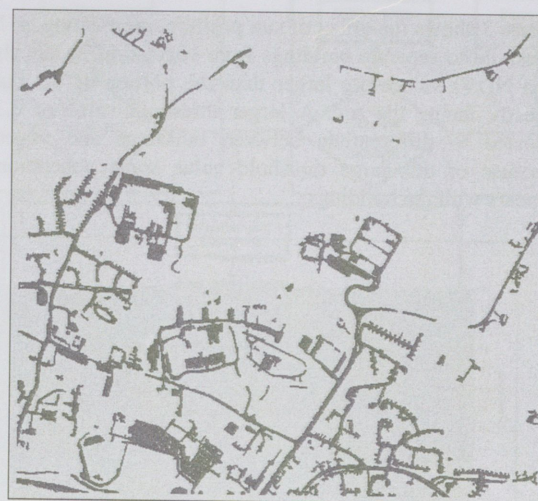


Figure 9: Extracted Roads

Gaps are present in the extracted roads and this is because of the trees and building shadows. This section of the road is not visible in aerial photos and LiDAR and needs further research for its successful extraction. Finally all the extracted objects were combined together and integrated into a specialized GIS system (Figures 10 and 11).

4. CLASSIFICATION ASSESSMENT

Results of the workflow were evaluated after the method of Heipke et al. (1997). Using this method, three different states for any feature can be identified (Hatger, 2006).

True positive (TP) - A phenomenon that is present within the input data and that has successfully been identified within the output data.

False positive (FP) - A phenomenon that is not present within the input data but that has falsely been found to be a phenomenon by the algorithm. Thus it is written to output data.

False negative (FN) - A phenomenon that is present within the input data but that has not been identified by the algorithm and therefore has been omitted from the output data.

Then we define Completeness, Correctness and Quality by

$$\text{Completeness} = TP / (TP + FN)$$

$$\text{Correctness} = TP / (TP + FP)$$

$$\text{Quality} = TP / (TP + FP + FN)$$

Object	Completeness (%)	Correctness (%)	Quality (%)
Buildings	98	72	60
Trees	99	83	80
Roads	78	72	62

Table 1: Accuracy Assessment



Figure 10: Extracted Objects & Orthophoto



Figure 11: 3D Model (County Sligo)

5. DISCUSSION

The classification results, indicated in Table 1, are the result of automated processes, depending on the choice of appropriate parameters and thresholds.

Building extraction is the first step in the classification process and is therefore important for the extraction of further objects. A completeness value of 98% implies that the adopted strategy has been successful in the identification of these objects. However correctness and quality values are significantly less than completeness due to the influence of FP values. FP values in buildings arise from large trucks or industrial installations incorrectly identified as buildings. However, FN values (i.e. missed buildings) mostly occurred for small buildings less than 50 sq. m.

In a subsequent manual step, these small buildings were individually identified and included in the final building layer. The identification of all buildings in the area in this way allowed the extracted vegetation data to be improved, which later helped in the extraction of roads. Commercial or residential buildings, having glass roofs or green colour were missed in the NDVI layer but they existed in the NDSM and were added to the building layer manually. Many small sheds were identified in the backyards of houses which are not part of the buildings, which significantly reduced the correctness value.

Vegetation was extracted by subtracting the building layer from the NDSM. Very small buildings which appeared in the vegetation layer were identified and manually added to the building layer. Continuing research is targeted at reducing the dependence on such manual steps.

Multiple reflections, size and compactness were used to separate single trees from groves. However, the LiDAR sensor can efficiently differentiate between multiple reflections only where their height differences are significant.

Roads appear to be the most difficult objects to extract. They are part of the DTM and have spectral reflectance, which varies a lot in a single image. Setting a NDVI threshold helps identify the areas where there is vegetation or not. Reflections from barren land or walking trails in the fields also have very low NDVI values. Roads which are not covered by building shadows or trees are detected successfully. Road markings of different colours also affect the extraction process. Roads connecting houses to the road are of different materials and need to be classified separately.

6. CONCLUSION

The accuracy of the generated orthophoto is critical for any classification technique using LiDAR and aerial images. Due to the nature of the push broom sensor and the configuration of the test flight (no overlap along strip and 15% overlap between strips) there is no possibility to combat limitations in the Red, and NIR orthoimages. Occluded areas and ghosting of building roofs (in the across flight direction) cannot be corrected adequately and the building roof structure is completely damaged in the areas close to strip edges. This is a major disadvantage in the identification and modelling of building roof structures. Ground control points, where available, should be used for the verification of the registration of the LiDAR point cloud and aerial images. In this approach we relied completely on orientation from GPS/INS data but for future research ground control points will be acquired and the accuracy of the image registration will be measured.

For the purpose of the NDVI image, the Red and NIR channels exhibit excessive tree pixels in the extremities due to tree lean and the structure of the resulting NDVI image will therefore not match the structure of the DSM. This requires further investigation. The DSM quality can also be improved by incorporating building foot prints if available.

LiDAR and aerial images should, ideally, not be captured separately. Objects which exist in the images might not exist in the LiDAR data and it is time consuming to identify and separate those points, especially vehicles on roads or in parking areas. In addition, if the time delay is significant, the vegetation may change considerably.

Results from the automatic and semi-automatic stages of this workflow are encouraging. Limitations identified above are the subject of continuing research.

7. REFERENCES

- Clode, S. and Rottensteiner, F. 2005. Classification of Trees and Powerlines from Medium Resolution Airborne Laser Scanner Data in Urban Environments. Proceedings of Workshop on Digital Image Computing, Brisbane, Australia, pp. 97-102.
- Clode, S., Rottensteiner, F., Kootsookos, P. and Zelniker, E., 2007. Detection and Vectorization of Roads from LiDAR Data. *Photogrammetric Engineering and Remote Sensing*, 73(5), pp. 517-535.
- Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., Lovell, B. 2004. Using the DempsterShafer Method for the Fusion of LiDAR Data and Multi-spectral Images for Building Detection. Proceedings of the 17th International Conference on Pattern Recognition, Vol. 2, pp. 339 – 342.
- Fugro International. FLI-MAP 400 Specifications. [Online]. Available at: <http://www.flimap.com/site47.php> (accessed: 26th June 2009).
- Haitao, L., Haiyan, G., Yanshun, H. and Jinghui, H., 2007. Fusion of High Resolution Aerial Imagery and LiDAR Data for Object-Oriented Urban Land-Cover Classification Based on SVM. ISPRS Workshop on Updating Geo-spatial Databases with Imagery & The 5th ISPRS Workshop on DMGISs, Urumchi, Xingjiang, China. [Online]. Available at: www.commission4.isprs.org/urumchi/papers/179-184%20Haitao%20Li.pdf (accessed: 26th June 2009).
- Hatger, C., 2005. On the Use of Airborne Laser Scanning Data to Verify and Enrich Road Network Features, Proceedings of ISPRS Technical Commission III Symposium, Enschede, Netherlands.
- Heipke, C., Mayr, H., Wiedemann, C. and Jame, O., 1997. Evaluation of Automatic Road Extraction. *International Archives of Photogrammetry and Remote Sensing*, XXXII (3/2W3), 56.
- Leica Geosystems Incorporation. [Online]. Available at: http://www.leica-geosystems.com/corporate/en/lgs_57627.htm (accessed: 12th October 2008).
- Mayer, H., 2008. Object Extraction in Photogrammetric Computer Vision. *ISPRS Journal of Photogrammetry and Remote Sensing*, Volume 63, Issue 2, pp. 213-222.
- Mayer, H., Hinz, S. and Stilla, U. 2008. Automated Extraction of Roads, Buildings and Vegetation from Multi-source Data, *Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences: ISPRS Congress Book*.
- N. Haala and C. Brenner, 1999. Extraction of Buildings and Trees in Urban Environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 54, pp. 130–137.
- Sithole, G and Vosselman, G., 2003. Comparison of Filtering Algorithms. Proceedings of the ISPRS working group III/3 workshop on 3-D reconstruction from airborne laserscanner and InSAR data, Dresden, Germany, 8-10 October.