**AERIAL IMAGES AND LIDAR DATA FUSION FOR AUTOMATIC FEATURE**

**EXTRACTION USING THE SELF-ORGANIZING MAP (SOM) CLASSIFIER**

M. Salah a, \*, J.Trinder a, A.Shaker b, M.Hamedb, A.Elsagheer b

a School of Surveying and Spatial Information Systems, The University of New South Wales,

UNSW SYDNEY NSW 2052, Australia - (m.gomah, j.trinder)@unsw.edu.au

b Dept. of Surveying, Faculty of Engineering Shoubra, Benha University, 108 Shoubra Street, Cairo, Egypt -

ahmshaker@link.net, prof.mahmoudhamed@ yahoo.com, alielsagheer@yahoo.com

**ISPRS** WG III/2 "Point Cloud Processing"

**KEY WORDS:** Aerial Images, Lidar, GLCM, Attributes, Building Detection, Classification, Self-Organizing Map.

**ABSTRACT:**

This paper presents work on the development of automatic feature extraction from multispectral aerial images and lidar data based

on test data from two different study areas with different characteristics. First, we filtered the lidar point clouds to generate a Digital

Terrain Model (DTM) using a novel filtering technique based on a linear first-order equation which describes a tilted plane surface,

and then the Digital Surface Model (DSM) and the Normalised Digital Surface Model (nDSM) were generated. After that a total of

22 uncorrelated feature attributes have been generated from the aerial images, the lidar intensity image, DSM and nDSM. The

attributes include those derived from the Grey Level Co-occurrence Matrix (GLCM), Normalized Difference Vegetation Indices

(NDVI) and slope. Finally, a SOM was used to detect buildings, trees, roads and grass from the aerial image, lidar data and the

generated attributes. The results show that using lidar data in the SOM improves the accuracy of feature detection by 38% compared

with using aerial photography alone, while using the generated attributes as well improve the detection results by a further 10%. The

results also show that the following attributes contributed most significantly to detection of buildings, trees, roads and grass

respectively: entropy (from GLCM) derived from nDSM; slope derived from nDSM; homogeneity (from the GLCM) derived from

nDSM; and homogeneity derived from nDSM.

\* Corresponding author.

**1. INTRODUCTION**

Research on automated feature extraction from aerial images

and lidar data has been fuelled in recent years by the need for

data acquisition and updating for GIS. The high dimensionality

of aerial and satellite imagery presents a challenge for

traditional classification methods based on statistical

assumptions. Artificial Neural Networks (ANNs) on the other

hand may represent a valuable alternative approach for land

cover mapping for such highly dimensional imagery. ANNs

require no assumption regarding the statistical distribution of

the input pattern classes (Hugo *et al.*, 2007) and they have two

important properties: the ability to *learn* from input data; and to

generalize and predict unseen patterns based on the data source,

rather than on any particular *a priori* model. The Self-

Organizing Map is one of the most commonly used neural

network classifiers. It can be adjusted to adapt to the probability

distribution of the inputs (Seto and Liu, 2003).

In this paper we applied the SOM algorithm for combining

multispectral aerial imagery and lidar data so that the individual

strengths of each data source can compensate for the weakness

of the other. The low contrast, occlusions and shadow effects in

the image were compensated by the accurately detected planes

in the lidar data. However, edges of features are not located

accurately in lidar point clouds because of the lidar’s system

discrete sampling interval of 0.5m to 1m, (Li and Wu, 2008).

Therefore, we have derived 22 attributes from both aerial image

and lidar data by a number of algorithms to alleviate this

problem. To evaluate the contribution of the lidar data and the

generated attributes in the detection process, three separate

SOM classification tests were carried out using different input

data to determine the accuracy of feature detection against a

reference map:

1. The aerial image, the lidar data and the derived

attributes,

2. The aerial image and the lidar data,

3. The aerial image only as input data for the SOM.

Finally, the contributions of the individual attributes to the

quality of the classification results were evaluated.

**2. RELATED WORK**

There have been many research efforts on the application of

aerial images and lidar data for building extraction.

Rottensteiner *et al.,* (2005) evaluated a method for building

detection by the Dempster-Shafer fusion of lidar data and

multispectral images. The heuristic model for the probability

mass assignments for the method was validated, and rules for

tuning the parameters of this model were discussed. Further,

they evaluated the contributions of the individual cues used in

the classification process to the quality of the classification

results, which showed that the overall correctness of the results

can be improved by fusing lidar data with multispectral images.

Matikainen *et al.,* (2007) used a classification tree approach for

building detection. A digital surface model (DSM) derived from

last pulse laser scanner data was first segmented into classes

‘ground’ and ‘building or tree’. Different combinations of 44

input attributes were used. The attributes were derived from the

last pulse DSM, first pulse DSM and a colour aerial ortho

image. In addition, shape attributes calculated for the segments

were used. Compared with a building reference map, a mean

accuracy of almost 90% was achieved for extracting buildings.

The numbers of studies that have utilized ANNs for highly

spectrally dimensional image analysis are limited. Jen-Hon and

Din-Chang (2000) applied the self-organized map classification

(SOM) method for SPOT scene land cover classification. Hugo

*et al.* (2007) assessed the potential of the SOM neural network

to extract complex land cover information from medium

resolution satellite imagery using MERIS Full Resolution data.

**3. STUDY AREA AND DATA SOURCES**

Two test data sets of different characteristics were used in this

study. The first area is a part of the University of New South

Wales campus; Sydney Australia, covering approximately

500m x 500m. It is a largely urban area that contains residential

buildings, large Campus buildings, a network of main roads as

well as minor roads, trees and green areas. Lidar data were

acquired over the study area in April 2005, using an Optech

ALTM 1225 with a pulse repetition frequency (PRF) of 25kHz

at a wavelength of 1.047μm. The multispectral imagery was

captured by film camera by AAMHatch on June 2005 at 1:6000

scale. The film was scanned in three colour bands (red, green

and blue) in TIFF format, with 15μm pixel size (GSD of 0.09m)

and radiometric resolution of 16-bit as shown in Figure 1(left).

The second study area is a part of Bathurst city; NSW Australia,

covering approximately 1000m x 1000m. It is a largely rural

area that contains small sized residential buildings, road

networks, trees and green areas. Lidar data was acquired over

the area by a Leica ALS50 sensor in August 2008, operating

with a PRF of 150kHz at a wavelength of 1.064μm. The multispectral

imagery was captured by a Leica ADS40 sensor on

October 2007. Three colour band (red, green and blue) images

were collected at 50cm GSD as shown in Figure 1(right).

Figure 1. Orthophotos for UNSW (left), Bathurst (right).

**4. METHODOLOGY**

Feature extraction of the study area was implemented in several

stages as follow:

**4.1 Filtering of lidar point clouds**

Filtering is the process of separating on-terrain points (DTM)

from points falling onto natural and human made objects.

Axelsson (2000) developed an adaptive Triangulated Irregular

Network (TIN) method to find ground points based on selected

seed ground measurements. Whitman *et al.,* (2003) used an

elevation threshold and an expanding search window to remove

non-ground points. Abo Akel *et al.,* (2004) used a robust

method with orthogonal polynomials and road network for

filtering of lidar data in urban areas.

The basic assumption of the approach adopted in this paper is

that the height of a ground point is lower than the heights of

neighbouring non-ground points and the terrain can be

described using a simple tilted plane within small areas. The

method started by dividing the data into small 50m x 50m

square patches. In principle, the patch should be larger than the

largest building within the test area in such a way that no object

within the study area can totally cover the patch. Otherwise,

points falling over buildings will be classified as on-terrain

points. Then, the algorithm constructed a matrix, A (m, n),

where m and n are the number of patches in both X and Y

directions respectively, see figure 2(left). Then, the lower left

and the upper right coordinates for each patch were determined

and stored. Data from both the first and the last pulse echoes

were used in order to obtain denser terrain data and hence a

more accurate filtering process. For each patch we fitted tilted

plane surfaces to the terrain points using equation (1):

Z = a + b \* x + c (1)

where X, Y and Z = coordinates of lidar point clouds.

The process of plane surface construction started with the

detection of two points, one on each patch border, in the Y

direction, which represent the minimum elevations on these

borders. The two points were then shifted in X directions by a

reasonable value, for example 1000m, while Z values remained

constant, see figure 2(middle). The reason behind the shifting

process is to create a new set of two points to construct a

comparison plane, see figure 2(right), which includes the four

detected points (two old and two new) and represents the

general slope of the patch. The main assumption here was that

the surface varies slowly from region to region over the patch

of interest. The four points were then used to determine the best

estimates of the coefficients of the plane by a least squares

solution. Based on the computed coefficient values of a, b and

c, equation (1) was applied for each individual point i with

coordinates Xi, Yi in the lidar point clouds to find the Z value of

its corresponding point on the plane. From a comparison of the

elevation of each data point with its corresponding elevation on

the generated plane surface, all points below, on or above this

plane within the threshold t (=15cm), were classified as onterrain

points. Threshold t was equal to the lidar system

accuracy. Figure 2 demonstrates the steps of the filtering

process, while figure 3 shows a part of the results for UNSW

data.

Figure 2. Dividing the area into small square patches (left),

detecting and shifting the lowest two points of the

patch (middle) and constructing the tilted plane and

removing the non-ground features (right).

Figure 3. Points filtered as on-terrain points in green colour

(left) compared to the aerial image (right).

Finally, the filtered lidar points were converted into an image

DTM, the DSM was generated from the original Lidar point

clouds (first and last pulses) and the nDSM was generated by

subtracting the DTM from the DSM, see figures 4. These are

grey scale images where tones range from dark for low

elevations to bright for high elevations.

Figure 4. DSM (left), DTM (middle) and the nDSM (right).

In order to analyze the produced filtering errors, a sample of

100 well distributed filtered points has been selected, overlaid

on the orthophoto and classified visually as ground and nonground.

Compared to those results, our algorithm has achieved

commission errors, classifying non-ground points as ground

points, and omission errors, classifying ground points as nonground

points, of about 3.1% and 5.2% for UNSW case study

and 5.9% and 9.4% respectively for Bathurst case study.

Compared with other methods, this technique is simple and

requires no work tuning parameters except for the patch size.

Also, fitting a simple tilted plane into a small square area

effectively removes most of the non-ground points especially

those on low vegetation.

**4.2 Generation of attributes**

Features or attributes commonly used for feature extraction

from aerial images and lidar data include height texture (Maas

and Vosselman, 1999) or surface roughness (Brunn and

Weidner, 1998) of the lidar data, reflectance information from

aerial images (Vögtle and Steinle, 2000) or lidar data (Hug,

1997), the difference between first and last pulses of the lidar

data (Alharthy and Bethel, 2002). The attributes calculated for

predefined segments or single pixels are presented as input data

for a classification method. Before generating the attributes, the

aerial photographs (already orthorectified by AAMHatch) were

registered to the lidar intensity image using a projective

transformation. The Root Mean Square (RMS) errors from the

modelling process were 0.01m and 0.01m in X and Y

respectively and the total RMS error was 0.02m, indicating an

accurate registration between image and lidar data and

demonstrating that most of the geometric distortions had

already been removed by the orthorectification process.

Following the transformation, the image was resampled to

30cm x 30cm and 50cm x 50cm cell size in case of UNSW and

Bathurst respectively to match the resolution of the lidar data. A

bilinear interpolation was used for resampling, which results in

a better quality image than nearest neighbour resampling and

requires less processing than cubic convolution.

In our test, a set of 78 possible attributes were selected as

shown in Table 1. Because of the way the texture equations

derived from the GLCM (Haralick, 1979) are constructed, many

of them are strongly correlated with one another. Clausi (2002)

analysed the correlations among the texture measures to

determine the best subset of measures and showed that

Contrast, Correlation and Entropy used together outperformed

any one of them alone. If only one can be used, he

recommended choosing from amongst Contrast, Dissimilarity

or Homogeneity. Based on these experiments, only 22 of the 78

possible attributes were uncorrelated and hence available for the

classification process as shown in the shaded cells of Table 1.

The attributes include those derived from the GLCM,

Normalized Difference Vegetation Indices (NDVI), standard

deviation of elevations, slope and the polymorphic texture

strength based on the Förstner operator (Förstner and Gülch,

1987).

Attributes Attribute R G B I DSM NDSM

Mean ● ● ● ● ● ●

St. Deviation ● ● ● ● ● ●

Spectral

Strength ● ● ● ● ● ●

Contrast ● ● ● ● ● ●

Dissimilarity ● ● ● ● ● ●

Homogeneity ● ● ● ● ● ●

A.S.M ● ● ● ● ● ●

Entropy ● ● ● ● ● ●

Mean ● ● ● ● ● ●

Variance ● ● ● ● ● ●

GLCM

Correlation ● ● ● ● ● ●

Height SD ● ● ● ● ● ●

Slope ● ● ● ● ● ●

Table 1. The full set of the attributes; attributes available for the

classification are shown by shading.

**4.3 Land cover classification**

The SOM (Kohonen, 1999) was used for classifying the images.

Figure 5 illustrates the basic architecture of an SOM. The input

layer represents the input feature vector and thus has neurons

for each measurement dimension. In our study, we applied a

separate neuron for each band. Therefore, the SOM has 29

input neurons which are: 22 generated attributes, 3 image bands

(R, G and B), intensity image, DTM, DSM and nDSM. For the

output layer of an SOM, we used a 15 x 15 array of neurons as

an output for the SOM. This number was selected because, as

recommended by Hugo *et al.* (2007), small networks result in

some unrepresented classes in the final labelled network, while

large networks lead to an improvement in the overall

classification accuracy. Each output layer neuron is connected

to all neurons in the input layer by synaptic weights.

Figure 5. Example of SOM with a 4 neurons input layer and

an equally spaced 5x5 neurons output layer.