

# Classification and Extraction of Trees and Buildings from Urban Scenes Using Discrete Return LiDAR and Aerial Color Imagery

Madhurima Bandyopadhyay<sup>\*a</sup>, Jan A.N. van Aardt<sup>a</sup>, and Kerry Cawse-Nicholson<sup>a</sup>

<sup>a</sup>Digital Imaging and Remote Sensing Laboratory, Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, NY 14623, USA

## ABSTRACT

Airborne Light Detection and Ranging (LiDAR) is used in many 3D applications, such as urban planning, city modeling, facility management, and environmental assessments. LiDAR systems generate dense 3D point clouds, which provide a distinct and comprehensive geometrical description of object surfaces. However, the challenge is that most of the applications require correct identification and extraction of objects from LiDAR point clouds to facilitate quantitative descriptions. This paper presents a feature-level fusion approach between LiDAR and aerial color (RGB) imagery to separate urban vegetation and buildings from other urban classes/cover types. The classification method used structural and spectral features derived from LiDAR and RGB imagery. Features such as flatness and distribution of normal vectors were estimated from LiDAR data, while the non-calibrated normalized difference vegetation index (NDVI) was calculated by combining LiDAR intensity at 1064 nm with the red channel from the RGB imagery. Building roof tops have regular surfaces with smaller variation in surface normal, whereas tree points generate irregular surfaces. Tree points, on the other hand, exhibit higher NDVI values when compared to returns from other classes. To identify vegetation points an NDVI map was used, while a vegetation mask was also derived from the RGB imagery. Accuracy was assessed by comparing the extraction result with manually digitized reference data generated from the high spatial resolution RGB image. Classification results indicated good separation between building and vegetation and exhibited overall accuracies greater than 85%.

**Keywords:** Classification, Fusion, LiDAR (Light Detection and Ranging), NDVI, Region-growing, Vegetation mask.

## 1. INTRODUCTION

Three-dimensional (3D) building models are used for variety of applications such as urban planning, city modeling etc. Accurate building models can be created manually by using aerial images, building blueprints, and other data sources<sup>1</sup>, but often rely on costly and time consuming processing. Apart from detecting and modeling buildings, there is a growing interest around urban forestry, i.e., the mapping of urban woody resource. Accurate information of urban vegetation is important to preserve the ecological environment within human habitats. Most of the tree statistics such as tree height, tree volume, species distribution etc. are measured manually in field surveys, which are expensive and tedious, while the advancement of remote sensing could help to reduce the cost of tree inventory<sup>2</sup>. However using only non-stereo optical imagery, it is impossible to acquire accurate 3D information of trees and buildings.

The invention of LiDAR technology provides a novel tool which can provide a detailed description of 3D structure of object surfaces<sup>1</sup>. Range scan output from a LiDAR system is commonly referred to as a “point cloud”, since it is comprised of a large number of points, each containing x, y, z coordinates and the associated per-point intensity information. With the advancement of technology, LiDAR provides increasingly denser point coverage over surfaces. Furthermore there is a growing demand for LiDAR data to model buildings<sup>1</sup> and trees<sup>3</sup>. Raw LiDAR data have been segmented in unsupervised classification using height texture measures to discriminate between buildings and vegetation<sup>4</sup>. In one study height data from laser scanner and existing ground plans were used to generate 3D building models automatically<sup>5</sup>. Recently fusion of LiDAR data with aerial imagery has become very popular for object extraction and quantification<sup>6,7</sup>. LiDAR data and a multispectral image were combined by using the Dempster – Shafer theory and the fused dataset was then used to classify land cover into building, tree, grassland, and bare soil classes<sup>6</sup>. Segmentation based weighted features such as height, texture map, height variation, and normal vector estimation from aerial images and LiDAR data were also used to detect trees<sup>7</sup>.

In our study we used the LiDAR data and high spatial resolution color (RGB) imagery to extract building and vegetation from other classes/cover types. Building rooftops have smoother surfaces compared to trees, thus a region-growing algorithm based on a smoothness constraint was used to identify different building regions. Once the building points were detected from LiDAR point clouds, an NDVI map and a vegetation mask were used to detect vegetation.

\*mxb2438@rit.edu; www.cis.rit.edu

Manually created vector files were used to measure the classification accuracy by calculating the number of detected building and tree points enclosed by their corresponding vector truth files.

This paper is organized in three major sections. We will start with a brief description of the study area and data used, next we present a short explanation on the preprocessing and the methodology used, followed by the results, discussion, and conclusion.

## 2. STUDY AREA AND DATA

LiDAR point clouds and the aerial color (RGB) imagery of Downtown Rochester were collected during fall 2011. A Leica ALS60 airborne laser scanner, provided by Kucera International, with a pulse rate up to 150,000 Hz was used to collect a LiDAR point clouds over the city. The density of the LiDAR point cloud is approximately 10 -15 points/m<sup>2</sup>. The LiDAR dataset contains the x, y, z location for each return, along with the return intensity, return number, number of returns, and vendor classification (either ground, non-ground, or noise). The aerial color (RGB) imagery was captured by Rochester Institute of Technology's (RIT) Wildfire Airborne Sensor Program (WASP) sensor (0.45µm to 9µm). The spatial resolution of color (RGB) imagery is 0.15m. The RGB imagery was orthorectified and processed into the same Universal Transverse Mercator (UTM) coordinate system as the LiDAR data.

Two different regions of Downtown Rochester were selected to test the classification algorithm. Figures 1 and 2 show these two regions. Region 1 is a residential area of size 150x143 m, with smaller size buildings and many trees, whereas Region 2 consists of large commercial buildings and it covers an area of 360x 260 m. These sites were selected to assess the initial algorithm efficiency and accuracy, with more variable sites planned for future efforts.

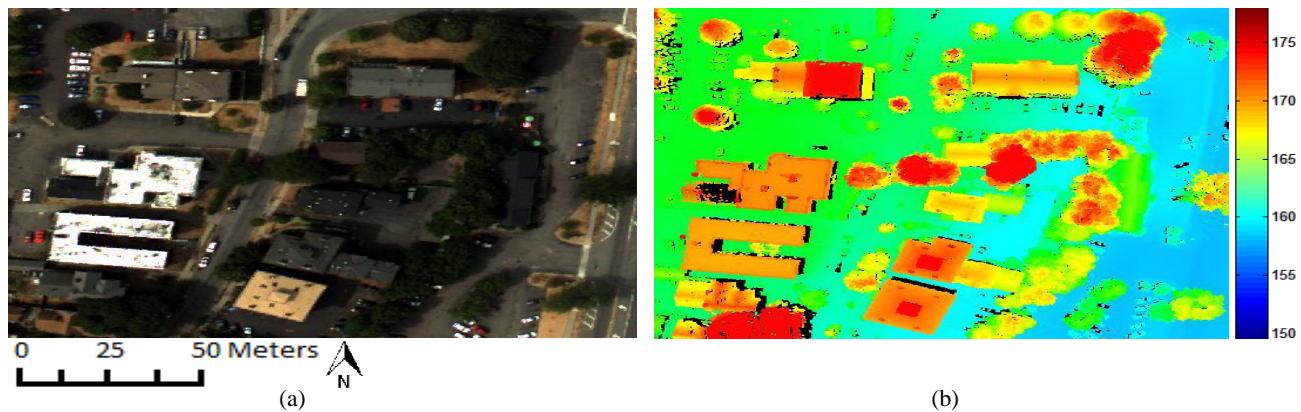


Figure 1. Region 1 is a residential area. This region is shown in (a) an RGB image, and (b) 3D LiDAR point clouds

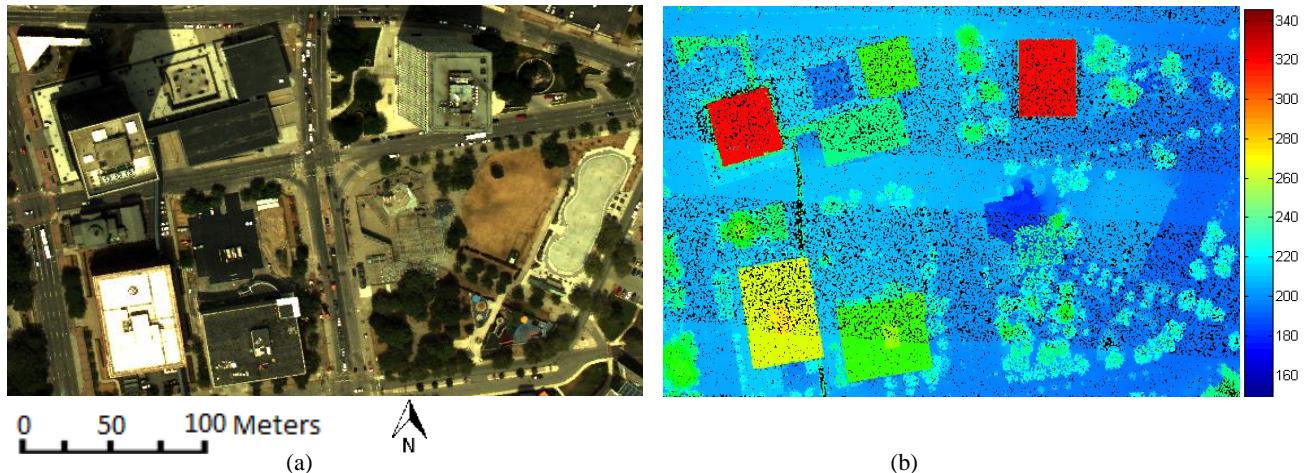


Figure 2. Region 2 consists of large commercial buildings. The region is shown as (a) an RGB image, and (b) 3D LiDAR point clouds

### 3. PREPROCESSING OF LiDAR POINT CLOUDS

Preprocessing the LiDAR data is an important step before performing any further operation on the point clouds. In the first preprocessing step, gross outlier points due to bird hits and noise were removed by gridding the point clouds into a 1x 1m tile size, followed by thresholding the point clouds for each tile by three standard deviations above and below the mean.



Figure 3. Preprocessing chain of LiDAR point clouds

Once the ‘outlier’ points were removed, the next step was to extract the ground points from the point cloud. The classification of LiDAR dataset into ground and non-ground returns is called filtering. There are a number of algorithms available for filtering, such as morphological, surface-based, and segment-based filters. The LiDAR dataset we were using here comes with a vendor-supplied classification field. Kucera International classified each point using Terrasolid software’s slope-based ground return classifier<sup>8</sup>. Using the supplied classification, the irregularly spaced ground points were interpolated into a 1 m resolution grid. Delaunay triangulation, using natural neighbor interpolation, was selected to construct the Digital Elevation Model (DEM). In the next step, a height model was created by subtracting the DEM from the point clouds. Thus the topography effect was eliminated from LiDAR data points and each point in the LiDAR data reflects the actual height of the object above the ground.

### 4. METHODOLOGY

The steps followed in the implemented methodology are shown in Figure 4. After preprocessing the LiDAR point clouds, the next step was to remove points below 1.5 m height. Thus the ground, roads, and small objects such as vehicles, shrubs, shades etc. were removed from the LiDAR data set and the remaining points consist of mainly buildings and/or vegetation points. Segmentation based on smoothness constraints was then applied to the remaining LiDAR points. Since building rooftops exhibit a regular pattern or flat surface compared to vegetation, this resulted in building regions that were separated from the rest of the points. However parts of tree canopies sometimes were included in the segmentation result; these were eliminated using the NDVI map that is introduced later. Once building points were detected, they were removed from the point clouds. In the next step a vegetation mask was used to detect all the vegetation points.

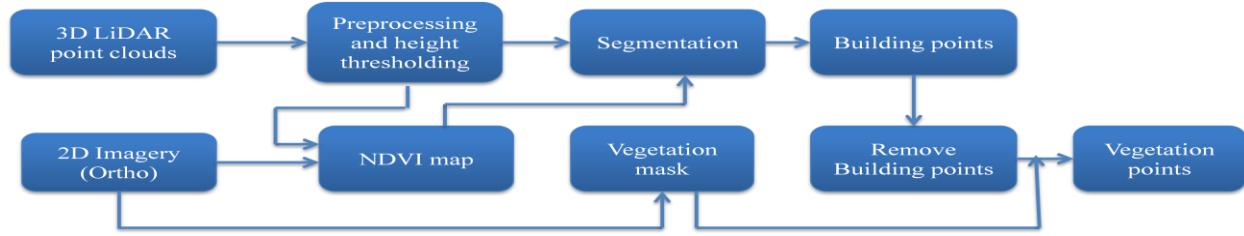


Figure 4. Workflow of classification method

The segmentation process, formations of NDVI map and vegetation mask are described in details below.

#### 4.1 Segmentation of the LiDAR point clouds

The segmentation method has two stages:

- 4.1.1 **Normal and Flatness estimation:** The normal for each point was estimated by fitting a plane to its neighboring points. The neighboring points can be selected based on K nearest neighbor (KNN) or fixed distance neighbor (FDN) methods<sup>9</sup>. Here we selected the 30 nearest neighbors for each individual point. The estimation of normal vectors and flatness measure can be performed by eigen-analysis of the covariance matrix of the point positions. Let  $p \in P$  be the sample point in the LiDAR point clouds, where  $p$  consists of x, y and z coordinates, and  $\bar{p}$  be the centroid of the neighborhood of  $p$ , i.e.,

$$\bar{p} = \frac{1}{|N_p|} \sum_{i \in N_p} p_i \quad (1)$$

The 3x3 covariance matrix  $C$  for the sample point  $p$  is given by

$$C = \frac{1}{|N_p|} \sum_{i \in N_p} (p_i - \bar{p})(p_i - \bar{p})^T \quad (2)$$

If  $\lambda_0 \leq \lambda_1 \leq \lambda_2$  are the eigenvalues sorted in the ascending order, then the eigenvector corresponding to the smallest eigenvalue, i.e.,  $\lambda_0$ , defines the normal vector at any point  $p$ <sup>9</sup>.

The flatness or surface variation<sup>1</sup> at any point  $p$  can be estimated as

$$F = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \quad (3)$$

4.1.2 **Region-growing:** This step uses the point normals and their flatness values to group points into smooth surfaces. Two constraints are followed in the region-growing algorithm<sup>9</sup>,

- i. Points belonging to a segment should be locally connected. The constraint is enforced by including the nearest neighbor in the region-growing process.
- ii. Points belonging to a segment should form a smooth surface, i.e., surface normals of the points should vary within a predefined threshold ( $\theta_{th}$ ). Region-growing starts with calculating the flatness ( $F$ ) of each point in the data set. Among all points, the one with minimum flatness is considered as a seed point and other points, based on their surface normal orientation, are iteratively added to the region. Here we wanted to prevent over-segmentation of surfaces, but we also did not want all points to form a single segment. Thus, depending on the rooftop structure, we set a threshold angle between surface normals of points ( $\theta_{th}$ ) at 10°. Future work will involve investigating the sensitivity of this threshold to make a robust segmentation method.

## 4.2 NDVI map from LiDAR point clouds and color (RGB) imagery

The Normalized Difference Vegetation Index (NDVI) has been used as an indicator to monitor plant growth, vegetation cover-type etc. The typical vegetation class shows a higher NDVI value than the other land cover-classes<sup>10</sup>.

$$NDVI = \frac{(NIR - VIS)}{(NIR + VIS)} \quad (4)$$

Where  $VIS$  and  $NIR$  represents the spectral reflectance measured at the red visible and near-infrared spectral regions respectively.

Since we are using an RGB image in our study, we utilized the intensity recorded by the LiDAR system for each point return to derive the  $NDVI$ . The wavelength of the LiDAR pulse is 1064 nm, which is in the near infrared region and the range of the red sensor of WASP is 600nm – 800 nm. The RGB image is geo-referenced and both LiDAR and RGB images were processed in the same UTM coordinate system, thus these were used to calculate  $NDVI$ .

For each LiDAR point we found the closest pixel from the RGB image and assigned the red value to that LiDAR point, thus for each point we calculated the  $NDVI$  value using equation (4). It should be noted that the LiDAR intensity is not calibrated, so the intensity values could exhibit more spectral variability across flight lines than those for a calibrated sensor. Figures 5(c), and 5(d) show the NDVI map of Region 1 and Region 2.

## 4.3 Vegetation mask from RGB image

As a validation of the NDVI approach we have created another vegetation mask from the RGB image, utilizing the three bands from the color image and calculating two different indices<sup>11</sup>.

4.3.1 *Ratio Index (RI):* This is the ratio of green radiance ( $\rho_{green}$ ) and the blue radiance ( $\rho_{blue}$ ).

$$RI = \frac{\rho_{green}}{\rho_{blue}} \quad (5)$$

4.3.2 *Visible Atmospherically Resistant Index (VARI)*: This index is expressed as,

$$VARI = \frac{\rho_{green} - \rho_{red}}{\rho_{green} + \rho_{red} - \rho_{blue}} \quad (6)$$

Both the index images have gray scale values between 0-255, which were then converted to the binary images based on a threshold value; this value was determined as a mid range value between highest and lowest gray level for each individual index image. The intersection of these two index images produced a vegetation mask image. Morphological operations such as connected component followed by a closing operation were performed to calculate the final vegetation mask. Figures 5(e) and 5(f) show the vegetation mask images of Region 1 and Region 2 respectively.

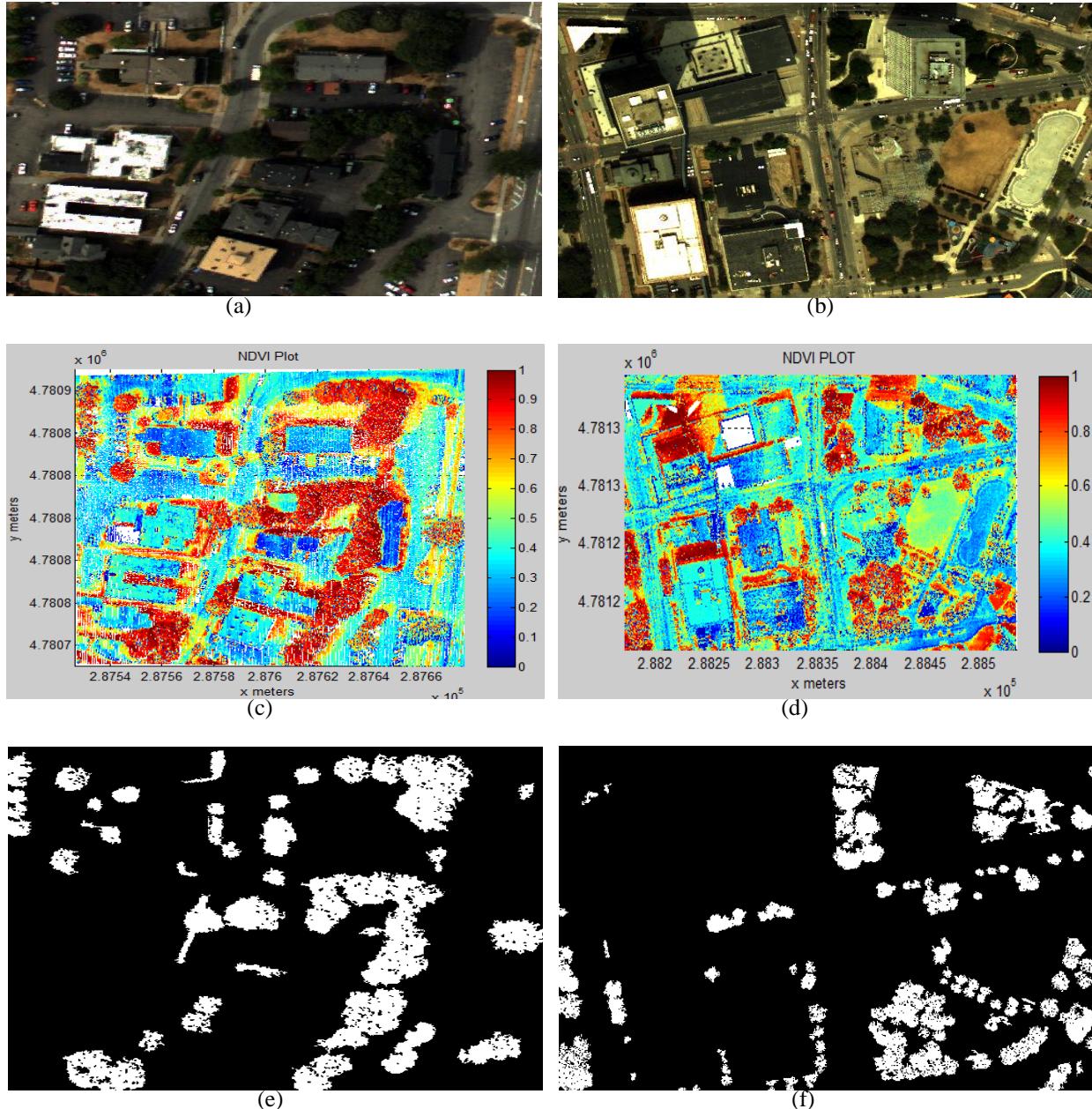


Figure 5. Demonstrating the (a) & (b) RGB images, (c) & (d) the NDVI map, (e) & (f) vegetation mask of Regions 1 and 2 respectively.

## 5. RESULTS AND DISCUSSION

The classification results for both study areas are presented here. As mentioned earlier, the classification accuracy was assessed by manually creating vector (GIS) files. Here, the whole scene was divided into three classes, namely, building/man-made objects, vegetation, and other classes. The “other” class arguably is very broad, but we want to focus on the building vs. vegetation separation in this study. For Region 2, we have considered building connectors and bridges within the building class. Thus three different vector files for each scene were created to determine the classification accuracy. These vectors files were overlaid onto the detected building and vegetation points to calculate the number of points that were contained within the boundary defined by their corresponding vector files. Figures (6), and (7) show the classification results of buildings and vegetation classes for Region 1 and Region 2 respectively. Tables 1 and 2 represent the confusion matrices for these two regions. Table 3 displays different accuracy measurements for these regions.

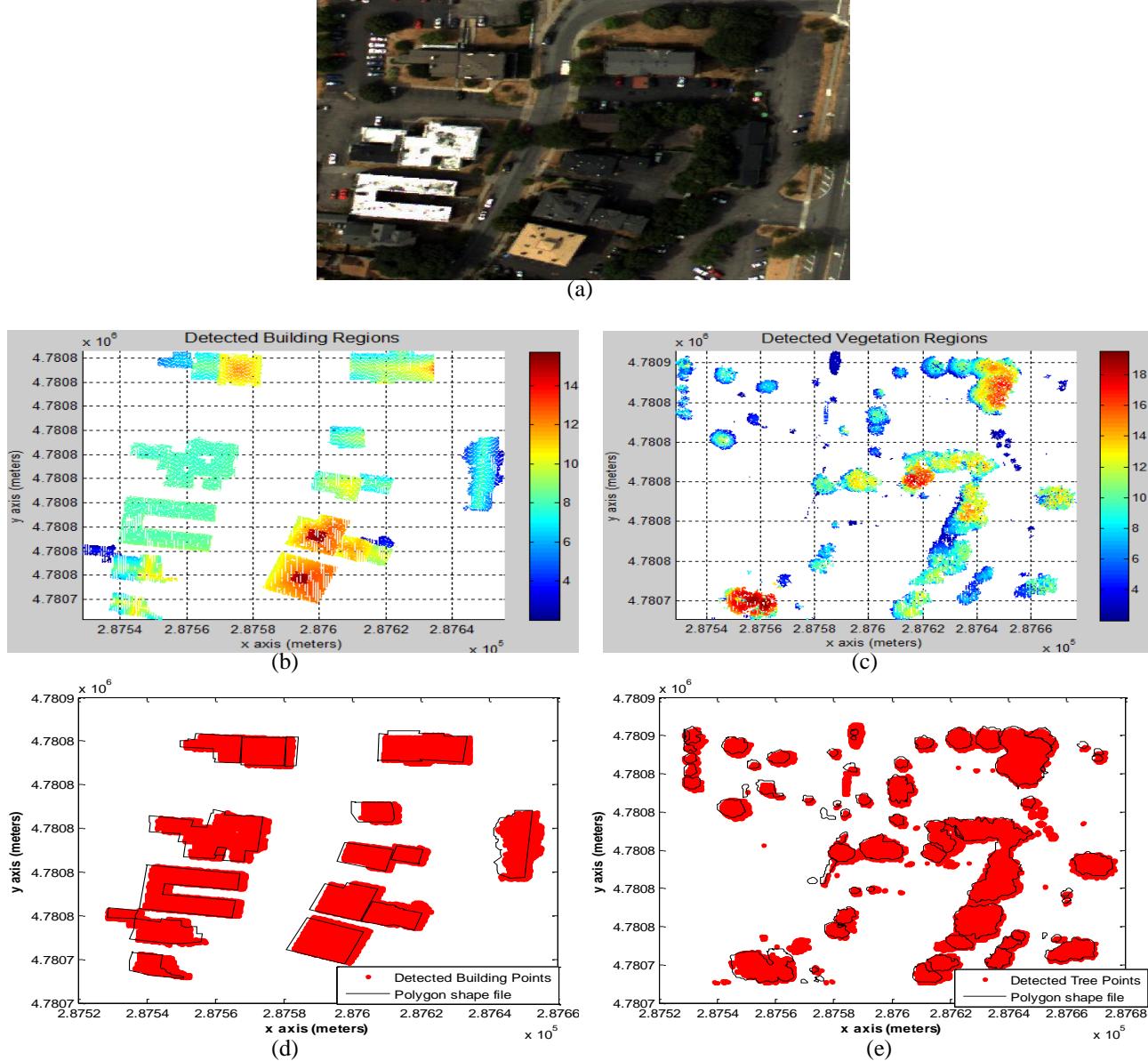
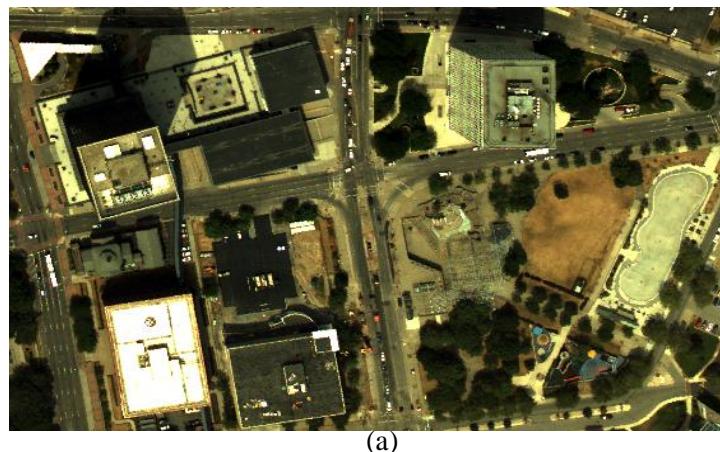


Figure 6. Building and vegetation classification results of Region 1, (a) RGB image, (b) Classified building regions, (c) classified vegetation regions, (d) GIS vector file overlaid on classified building points and (e) GIS vector file overlaid on classified vegetation points



(a)

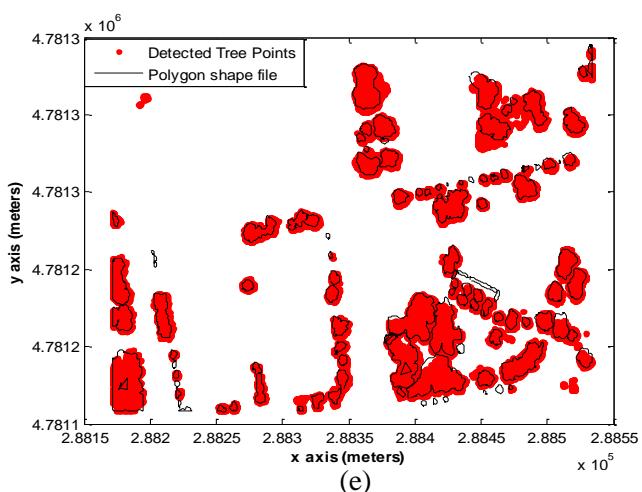
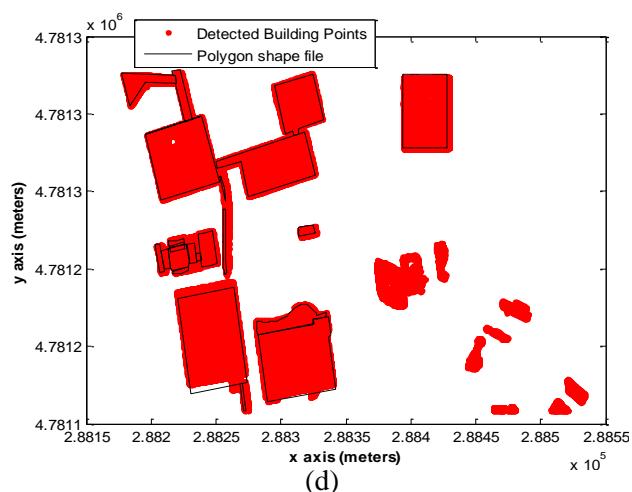
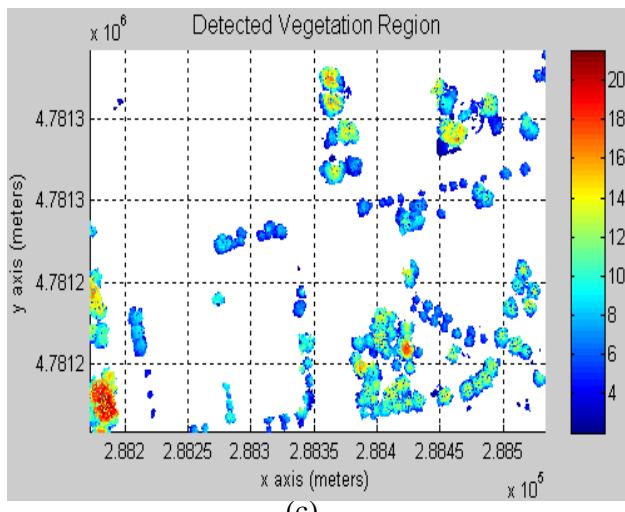
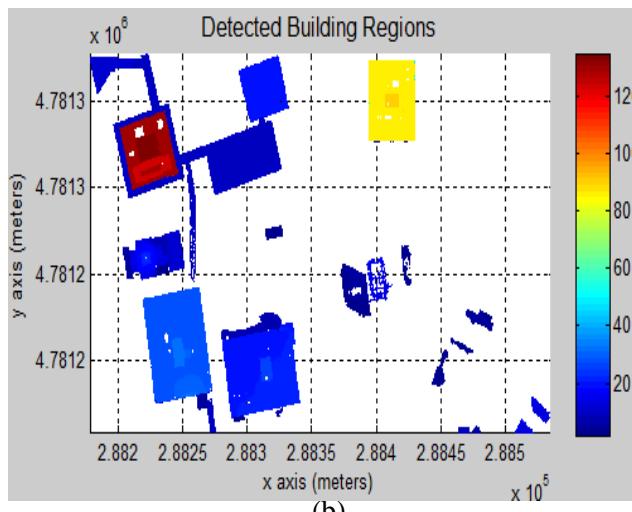


Figure 7. Building and vegetation classification results of Region 2, (a) RGB image, (b) Classified building regions, (c) classified vegetation regions, (d) GIS vector file overlaid on classified building points and (e) GIS vector file overlaid on classified vegetation points

Table 1. Classification results of Region 1

Classified Points	Reference Points			Row total
	Building	Vegetation	Others	
<b>Building</b>	21048	387	3128	24563
<b>Vegetation</b>	144	37385	6387	43916
<b>Others</b>	2291	8278	126716	137285
<b>Column total</b>	23483	46050	136231	205764

Table 2. Classification results of Region 2.

Classified Points	Reference Points			Row total
	Building	Tree	Others	
<b>Building</b>	263037	327	35353	298717
<b>Tree</b>	283	194261	24886	219430
<b>Others</b>	19667	44002	1479940	1543609
<b>Column total</b>	282987	238590	1540179	2061756

Producer's Accuracy: This indicates the probability of a certain pixel/point on land-cover map as being correctly classified. Producer's accuracy also called omission error<sup>12</sup>.

User's Accuracy: This indicates probability that a pixel/point will be labeled as a certain land-cover class in the classification map is really that class. User's accuracy also called commission error<sup>12</sup>.

Overall Accuracy: This is computed by dividing the total correct (i.e., the sum of the major diagonal) by the total number of pixels/points in the confusion matrix<sup>12</sup>.

Table 3. Producer's Accuracy, User's Accuracy, and Overall Accuracy for Region 1 and Region 2

	Region 1			Region 2		
	Building	Vegetation	Others	Building	Vegetation	Others
<b>Producer's Accuracy (%)</b>	89.63	81.18	93.01	92.95	81.42	96.08
<b>User's Accuracy (%)</b>	85.68	85.12	92.3	88.05	88.52	95.87
<b>Overall Accuracy (%)</b>	$185149/205764 = 89.98$			$1937238/2061756 = 93.96$		

For both the regions the overall classification accuracy and the accuracy for individual classes are shown in Table 1 and Table 2. From these tables, we can observe that the classification accuracy for the building class is greater than 85%. But for vegetation class the Producer's accuracy is comparatively low, near 81% for both regions. From the tables we can see that, there is very little confusion between the building and vegetation classes, but a comparatively large numbers of building and vegetation points are confused with the class containing other objects, which include ground, road, vehicles and shrubs etc. One of the reasons behind confusion among classes is that there is some misregistration (MSE nearly 2m) between LiDAR and RGB imagery<sup>13</sup>. These two data sets were collected from the same platform and processed under the same coordinate system, but misalignment between these two data-sets exists due to system and processing error and is unavoidable under standard processing protocol. We contend that this misregistration had an adverse effect on our classification algorithm. Also, the manually created vector files were created from the RGB image, thus misalignment affects the accuracy assessment procedure. In Figures 6(b) and 6(d) we can see that for Region 2, some portions were classified as buildings, which were not included in our reference vector file. After re-evaluating this region carefully, we found that there is an observatory tower, which has a steel mesh-like structure. In the RGB image this tower is not very prominent, but due to its mesh like structure it was detected in the LiDAR data set and our algorithm classified it in the building class.

## 6. CONCLUSIONS AND FUTURE WORK

Our classification algorithm, based on the fusion of LiDAR point clouds and color RGB imagery, resulted in a good classification result for both buildings and vegetation in an urban environment. On average the classification accuracy for both vegetation and buildings is greater than 85%, which can further be improved by correctly registering RGB imagery and LiDAR point clouds. An NDVI map, created by using the red channel of RGB and the intensity return of LiDAR proves to be very good indicator between vegetation and other classes (Figures 5(c) and (d)).

Future work will focus on (i) registering these two datasets to improve the classification accuracy by lowering the confusion among different classes, (ii) applying this algorithm to different sites that contain buildings and vegetation of different size and orientation, to evaluate its robustness, and (iii) expanding the “other” class to include a higher class resolution for assessment of between class confusion.

## REFERENCES

- [1] Zhou, Q-Y., Neumann, U., “Fast and extensible building modeling from airborne lidar data,” Proceeding of the 16th ACM SIGSPATIAL International conference on Advances in Geographic Information Systems, ACM, (2008).
- [2] Liu, J., Shen, J., Zhao, R., and Xu, S., “Extraction of individual tree crowns from airborne LiDAR data in human settlements,” Mathematical and Computer Modelling, (2011).
- [3] Popescu, S.C., Wynne, R.H., Nelson, R.F., “Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass,” Canadian Journal of Remote Sensing 29(5), pp. 564-577 (2003).
- [4] Elberink, S.O., and Maas, H-G., “The use of anisotropic height texture measures for the segmentation of airborne laser scanner data,” International Archives of Photogrammetry and Remote Sensing 33, pp. 678-684 (2000).
- [5] Haala, N., Brenner, C., and Anders, K.-H., “3D Urban GIS From Laser Altimeter and 2D Map Data,” International Archives of Photogrammetry and Remote Sensing 32, pp. 339-346 (1998).
- [6] Rottensteiner, F., Trinder, J., Clode, S., and Kubik, K., “Using the Dempster Shafer method for the fusion of LiDAR data and multi-spectral images for building detection,” Information Fusion 6(4), pp. 283-300 (2005).
- [7] Secord, J., and Zakhor, A., “Tree detection in urban regions using aerial lidar and image data,” IEEE Geoscience and Remote Sensing Letters 4(2), pp. 196-200 (2007).
- [8] Labiak, R.C., van Aardt, J. A. N., Bespalov, D., Eychner, D., Wirch, E., and Bischof, H-P., “Automated method for detection and quantification of building damage and debris using post-disaster LiDAR data,” Laser Radar Technology and Applications XVI. Proc. SPIE 8037, pp. 80370F-80370F (2011).
- [9] Rabbani, T., “Automatic reconstruction of industrial installations using point clouds and images,” PhD thesis, Delft University of Technology (2007).
- [10] Gitelson, A.A., “Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation,” Journal of Plant Physiology 161(2), pp. 165 - 173 (2004).
- [11] Ok, A.Ö, “Robust Detection of Buildings from a single color aerial image,” Proceedings of GEOBIA 2008 - Pixels, Objects, and Intelligence, Calgary, Alberta, Canada (2008).
- [12] Congalton, R., “A review of assessing the accuracy of classifications of remotely sensed data,” Remote Sensing of Environment 37(1), pp. 35 - 46 (1991).
- [13] Sun, S., Salvaggio, C., “Complex Building Roof Detection and Strict Description From LIDAR Data and Orthorectified Aerial Imagery,” IEEE International Geoscience and Remote Sensing Symposium, pp. 5466 - 5469 (2012).