A TIN-based classification approach for buildings or vegetation extraction

Shijun Tang
School of Engineering, Computer Science and Mathematics
West Texas A&M University
Canyon, TX, USA 79016
stang@wtamu.edu

Rajan Alex
School of Engineering, Computer Science and Mathematics
West Texas A&M University
Canyon, TX, USA 79016
ralex@wtamu.edu

Abstract—In this paper, we propose a new TIN-based classification approach for feature extraction from Light Detection and Ranging (LiDAR) data point clouds. The method builds TIN first and then computes variance (standard deviation) of normal vectors at each node to classify objects from raw LiDAR data point clouds. Experimental results indicate that our method can effectively classify buildings or vegetation via choosing different threshold values of variance (or standard deviation) of normal vectors.

Keywords—LiDAR, variance of normal vector, triangulated irregular network, classification

I. INTRODUCTION

Classification and detection of vegetation or buildings are important in land use planning as well as environment monitoring. Feature extraction from LiDAR point clouds has been playing the key role in classification and detection of vegetation or buildings. It is convenient that the Triangulated Irregular Network (TIN) model has been employed to describe and operate the 3D sub-randomly spatial distributed LiDAR points and neighboring relation of spatial discrete points. A TIN model has been constructed via the planar Delaunay triangulation network. The Triangulated Irregular Network model (TIN) can directly reflect and consider the points’ neighboring relation in 3D LiDAR point clouds.

For filtering and classifying the LiDAR data, many methods based on the TIN model have been proposed. The main approaches include the region growing [1], and the least square interpolation method [2]. Akel et al (2003) [3] provided that, for each triangle in the TIN model, the normal direction and mean height of the three points that compose it are calculated. Then, for each two triangles, if they have similar normal directions and heights, these triangles are merged into a region using the region growing approach. Although this method is for dense raw LiDAR data, there exists many misclassifications because they determine if two neighboring triangles are similar only by computing their normal direction and the height of the neighboring triangle.

Zeng et al [4] designed an assistant plane and made classifications according to the neighbor’s number and height difference of every point based on a TIN model. But, the height difference of every point on TIN doesn’t completely reflect the characteristics of objects (vegetation and buildings) from the physical nature.

Belkhouche et al [5] proposed a surface flatness method based on a TIN model for detection of vegetation from raw LiDAR data. The flatness of a 3D object modeled by a set of points is defined as its volumetric surface divided by its surface projected on 2D. Vegetation tends to have higher flatness values than other objects. However, some objects (buildings with no flat tops) also have higher flatness values.

The above methods have their own advantages and disadvantages. In order to make classification better, we combine the variance of normal vectors distribution with height interval [7, 8, 9] and propose a novel method to perform classification of aerial LiDAR data into buildings or vegetation based on a TIN model, which effectively uses the information of LiDAR data cloud points.

II. METHODOLOGY

A. Triangulated Irregular Network (TIN)

Triangulated Irregular Network (TIN) comprises a triangular network of vertices with associated coordinates in three dimensions connected by edges. Three-dimensional visualizations are readily created by rendering of the triangular facets. The TIN model has typically been constructed based on the Delaunay triangulation network. The TIN is widely used for representation of the physical land surface or sea bottom, made up of irregularly distributed nodes and lines with three dimensional coordinates (x, y, and z). The TIN model has often been employed to operate the 3D spatial distributed LiDAR points.

B. Digital Terrain Model

Since LiDAR data is collected from measuring the time delay between transmission of a pulse and detection of the reflected signal, the elevation and surface information of objects might be extracted from LiDAR data. Digital
Elevation Model (DEM) is created using elevation of bare earth points, and Digital Surface Model (DSM) is based on the actual surface, including vegetation and buildings. A normalized digital surface model (nDSM) represents the height of ground features. The nDSM is obtained from the aerial LiDAR data (DSM) subtracting standard (DEM), that is, nDSM = DSM - DTM.

C. LiDAR Data Set

Light Detection and Ranging (LiDAR) is an optical remote sensing technology that measures properties of scattered light to find the range and/or other information of a distant target. Our method does not require other information but LiDAR data. In this paper we use raw LiDAR data provided by the state of Louisiana [6]. The dataset was collected at a high emission rate of about 15,000 to 30,000 pulses per second. The resolution of the LiDAR data set is about 0.18 pt/m².

D. Variance of Normal Vectors Based on TIN Model

We use the new TIN-based classification approach—the variance of normal vectors at each node on TIN (Triangulated Irregular Network) to find the characteristics and detect vegetation or buildings from LiDAR raw data at urban areas. The below algorithms were implemented in Matlab.

Algorithm

Step01: Input the preprocessed LiDAR data (i.e., nDSM = DSM - DTM)

Step02: Use function Delaunary from Matlab Library to obtain a set of triangles

Step03: Find all triangles associated with the vertex \( p \) for each vertex \( p \)

Step04: Calculate and normalize each normal vector at the vertex \( p \) by using two vectors at two sides of each triangle

Step05: Count the number of normal vectors at the vertex \( p \)

Step06: Get the average normal vector \( N_p \) at each vertex

Step07: Calculate the distribution of normal vectors (Standard Deviation \( Std \) or Variance \( Var \)) at the selected point \( p \)

Step08: Apply the criterion of \( Std \) or \( Var \) to determine if the point belongs to a building or vegetation

Step09: If \( Std \) (or \( Var \)) < threshold value

   Classify flat, e.g. building roof

else

   Classify uneven, e.g. vegetation

End if

Figure 1 shows the diagram for representation of a surface TIN and normal vectors distribution at node \( p \) on a surface TIN. The vectors \( A \) (or \( B \)) are from two edges of a triangle starting from node \( p \). \( N \) is the normal vector to the plane formed by vectors \( A \) and \( B \). The vectors \( N \) can be expressed as:

\[
 N = A \times B = N_x i + N_y j + N_z k
\]

As shown in Figure 1, three components of variance of normals are \( \text{var}_x = \frac{\sum_{m=1}^{m} (N_x - N_{xm})^2}{m} \), \( \text{var}_y = \frac{\sum_{m=1}^{m} (N_y - N_{ym})^2}{m} \), and \( \text{var}_z = \frac{\sum_{m=1}^{m} (N_z - N_{zm})^2}{m} \), respectively. The number of normal vectors at the vertex \( p \) is expressed as \( m \). Three components of the standard deviation of normal are \( \text{std}_x = \frac{\sum_{m=1}^{m} (N_x - N_{xm})^2}{m} \), \( \text{std}_y = \frac{\sum_{m=1}^{m} (N_y - N_{ym})^2}{m} \), and \( \text{std}_z = \frac{\sum_{m=1}^{m} (N_z - N_{zm})^2}{m} \), respectively.

III. EXPERIMENTAL RESULTS AND ANALYSES

In this paper, we preprocessed the raw LiDAR data to get normalized height. Due to the characteristics of the urban region, we employed the methods of height interval (separated levels) to reduce the errors of computing normal vectors. We only took the data when height \( h > 9.14 \text{m} \) in this paper (not including classes such as cars, shrubs, grass and roads, etc. under the height \( h < 9.14 \text{m} \)).

Figures 2~3 give the classification results using our method. Here we employed the proposed normal vector variation at each node on TIN to classify objects buildings/vegetation. We took the same threshold value for three components of standard deviation of normal vectors in this paper. The threshold we selected may be adjusted according to the different density of raw LiDAR data. The blue represents buildings. The green represents vegetation.
From figures 2-3, we find that the main buildings have been recognized, although there exists some green dots on the surface of buildings. In figure 3, the green dots obviously decrease on the roof of buildings when the threshold value of \( \text{std} \) is taken as 0.782.

![Fig. 2(a)](image)

![Fig. 2(b)](image)

Fig. 2: Classification results (the yellow represents lines for Triangulated Irregular Network, the blue represents buildings, and the green represents vegetation) (a) 3D classified result using our TIN-based classification approach at \( h > 9.14 \text{m} \) and \( \text{std} < 0.482 \), (b) 2D classified result at \( h > 9.14 \text{m} \) and \( \text{std} < 0.482 \)

In our TIN-based classification approach, the category of determining each node depends on the distribution of normal vectors on the node which relates with its neighboring triangles. During the computation of variation of normal vectors, each normal vector at each node relates to two vectors on the two sides of each triangle. The distribution of triangles around the node and the relative height of the point greatly affect the classifying result. Thus, for the sparse LiDAR data at the urban areas, the height distribution of the roof of a building easily mixes with the distribution of normal vectors around the node which belongs to vegetation.

![Fig. 3(a)](image)

![Fig. 3(b)](image)

Fig.3. Classification results (the yellow represents lines for Triangulated Irregular Network, the blue represents buildings, and the green represents vegetation) (a) 3D classified result using our TIN-based classification approach at \( h > 9.14 \text{m} \) and \( \text{std} < 0.782 \), (b) 2D classified result at \( h > 9.14 \text{m} \) and \( \text{std} < 0.782 \)

Although there exists little misclassification in the classified categories, as an effective feature extraction method combining with the properties of TIN, this method can be used for vegetation or buildings extraction via choosing the proper threshold of normal distribution. This method is more suitable for raw LiDAR data at large areas due to its simple algorithm and quick computation.

IV. CONCLUSIONS

In this paper, we present a method for vegetation/buildings extraction using low resolution LiDAR raw data. The contribution is to correctly extract vegetation/buildings from the raw LiDAR data set. We successfully used the fact that vegetation tends to have a larger derivation of normal vectors distribution than those of other objects.
Also, we have effectively classified the sparse LiDAR data point clouds using the new TIN-based method. The advantage of the proposed method includes that it can be employed for buildings or vegetation extraction via choosing the threshold value of normal vectors distribution at each node \( p \), and that the proposed method reflects the neighbors’ relationship of each node and reduces computational time since the proposed method only involves the direction (normal vector) distribution of triangular facets at each node of TIN. The further exploration and comparison of the proposed techniques will be completed in future work.

REFERENCES