CANOPY HEIGHT ESTIMATION FROM LIDAR DATA USING OPEN SOURCE SOFTWARE COMPARED WITH COMMERCIAL SOFTWARE

Nurhanisah Hashim, Zulkepli Majid and Halim Setan
Faculty of Geoinformation and Real Estate
Universiti Teknologi Malaysia
Email: niesa_405@yahoo.com

ABSTRACT

The goal of this study is to analyze the performance of Open Source Software (OSS) towards the generation of Digital Terrain Model (DTM) and Digital Surface Model (DSM), further on estimates the canopy height by using Light Detection and Ranging (LIDAR) data. Generation of DTM and DSM are very important in this research to ensure that better canopy height can be modeled. DTM and DSM commonly known as a digital representation of earth surface topography where DTM only represent the ground surface while DSM represent all the features including buildings and trees. Many software that have a function to generate DTM and DSM were developed recently. However, most software has been commercialized; therefore it requires a high expenditure to own the software. Advanced technology has lead to the emergence of the growing OSS. OSS is software that can be downloaded for free via the internet. By taking the forestry area of Pekan, Pahang for this research, LIDAR data for that particular area is processed by using the OSS Geographic Resources Analysis Support System (GRASS). To determine the effectiveness and capability of GRASS in the DTM and DSM generation, the same data were processed using commercial software which is TerraScan so that the result can be compared, further on better canopy height can be modeled.

Key words: LIDAR, Digital Terrain Model, Digital Surface Model, Canopy Height, Open Source Software.

1.0   INTRODUCTION

Light Detection and Ranging which is known as LIDAR is a very advanced active technique which uses laser light as the energy sources (Jihn, 2005). It is a new high technology technique that had been used by some organization nowadays in survey and mapping. It is actually a method of measuring the height of the earth's surface using laser technology. Height is measured by measuring reflected laser light from distant object and from this information to determine position. LIDAR combines three advance technologies which are laser, Global Positioning System (GPS) and Inertial Navigation System (INS) or Inertial measurement unit (IMU) **(Figure 1)**. Interest in the use of LIDAR is particularly high because it can take 3D

measurement of objects and surface (Dubayah and Drake, 2001). With these technologies that can give better 3D results, collecting LIDAR data can generate topography maps such as Digital Terrain Model (DTM) and Digital Surface Model (DSM). By its ability to generate topographic maps, it is very useful for many applications such as forest management, environmental monitoring, flood plain mapping and many more (Li et al. , 2008) However this study only covers a small part of forest management which is estimating the canopy height in forest area.
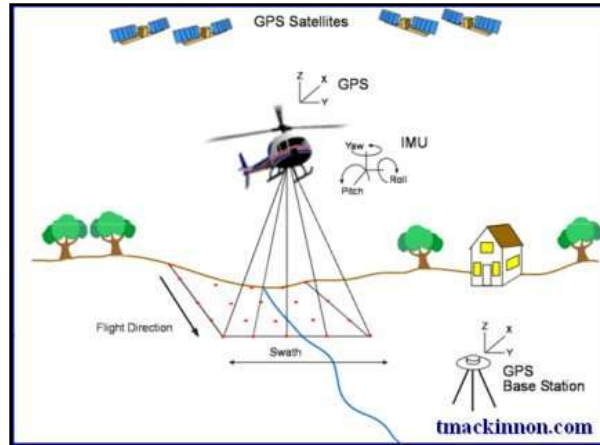

Figure 1: Three advanced technologies combines in LIDAR system

Before LIDAR is developed, maps can be produced using conventional methods such as geodetic and photogrammetric methods. It has limitations in term of their accuracy, time consumption, feasibility and applicability. The limitation has increase when it comes to produce forest mapping. Field data collection is very difficult in forest area since the conventional methods had always been used widely in Malaysia for this time period. Sometimes the data collection works were made impossible due to the existence of thick vegetation. Therefore, it will give adverse effect on the map producing. Interest in the use of LIDAR is particularly high because it promises to provide information about the three dimensional (3D) structure of forest canopy which has been particularly difficult using conventional method (Ustin, 2004). Thus, there is great hope for improving producing forest map including variables of information such as tree height, biomass, timber volume, and the vertical distribution of canopy elements. Due to the widespread use of LIDAR recently, it is found that the emerging of software with high growth potential for LIDAR data processing has increases, including whether commercial software or Open Source Software (OSS). Advanced information technology today has resulted in the emerging of OSS is increasing as well as the number of users. Commercial software has facing enormous competitor from OSS because OSS can be accessed freely via internet.

Hence, this study was carried out to investigate the performance one of the OSS which is GRASS in production of canopy height compared to TERRASCAN commercial software in the forest area of Pekan, Pahang. In order to generate DTM and DSM the best filtering method for the forest area should be identified earlier furthermore better canopy height can be estimated.

2.0   CANOPY HEIGHT

Canopy height is one of the applications in the use of LIDAR. It is modeled based on the difference in height between the surface of the ground and the tree surfaces. The height of the

ground surface is known as DTM but the surface height of the tree is known as DSM. There are many applications and information can be obtained with a canopy height data. According to Dubayah and Drake (2000), vegetation height is a function of species composition, climate and site quality, and can be used for land cover classification or in conjunction with vegetation indices from passive optical sensors. Therefore, canopy height data is very important to ensure that all the above functions can be performed.

3.0   DATA PROCESSING

Data was processed using two different software which are GRASS OSS and TERRASCAN Commercial software. The main objective of this study is to model the canopy height of Pekan Pahang forest area. Hence, the implementation of the process to be discussed referred to generate Digital Terrain Model and Digital Surface Model furthermore the estimated of the canopy height for the study area. A filtering process was performed to extract grounds points and surface points from the LIDAR data. In order to determine the potential of GRASS OSS for filtering and processing LIDAR data, TERRASCAN software was used to utilize the same processing. Hence the result can be compared.

3.1   Geographic Resources Analysis Support System (**GRASS)**

Geographic Resources Analysis Support System or better known as GRASS is general-purpose GIS software. It is actually one of the Open Source Software that had been developed for some time and can be used by anyone freely without purchasing or buying any license. GRASS was developed during the period 1982–1995 by the U.S. Army Corps of Engineers Construction Engineering Research Laboratory (CERL) in Champaign, Illinois, to support land management at military installations (Westervelt, 1991).

GRASS cannot read raw laser data format which is .LAS. If the data came in the format of .LAS, it needs to be converted into XYZ format data. The process was divided into several steps starting from the data load until the estimation of the canopy height. **Figure 2** below shows steps to process LIDAR data in GRASS.
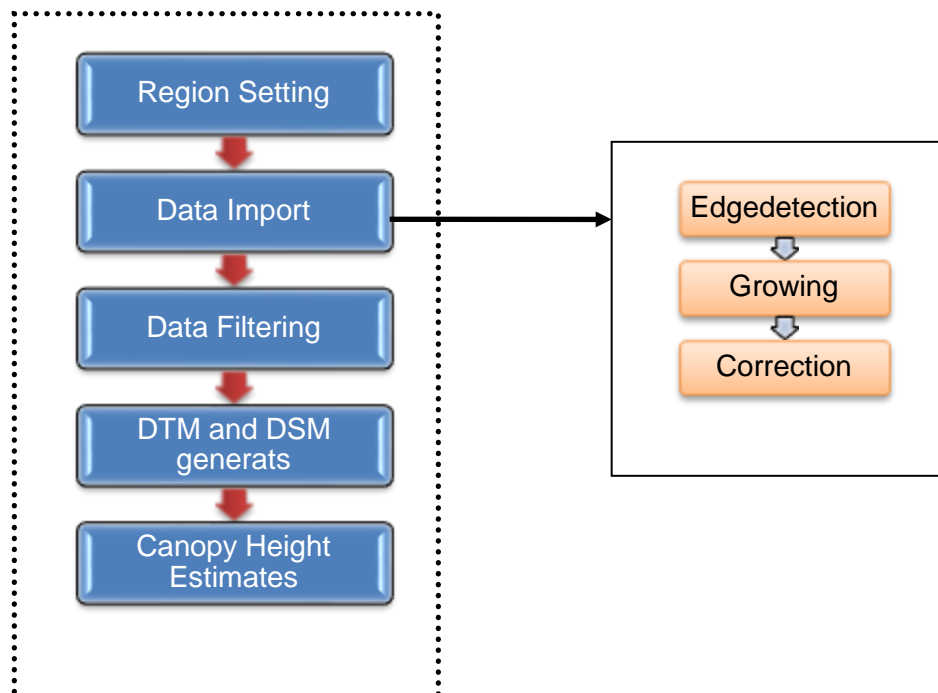
Figure 2: Steps in GRASS processing

The **Figure 3** shows the managing the region for the study area by using 'g. region' module. The module allows managing the settings of the current geographic region. The region's boundaries are given as the northernmost, southernmost, easternmost, and westernmost points that define its extent (cell edges). Data which is in XYZ format was imported in grass using ' r.in.xyz ' module which means the module will load the XYZ format data into a raster map. ' r.in.xyz ' was designed for processing massive point cloud datasets for example raw LIDAR or side scan sonar swath data. This module available statistics for populating the raster which are Variance and derivatives use the biased estimator (n) and Coefficient of variance is given in percentage and defined as (stddev/mean)*100**. Figure 3** shows the raw data for the study area in raster format after imported the XYZ data using ' r.in.xyz ' manual.
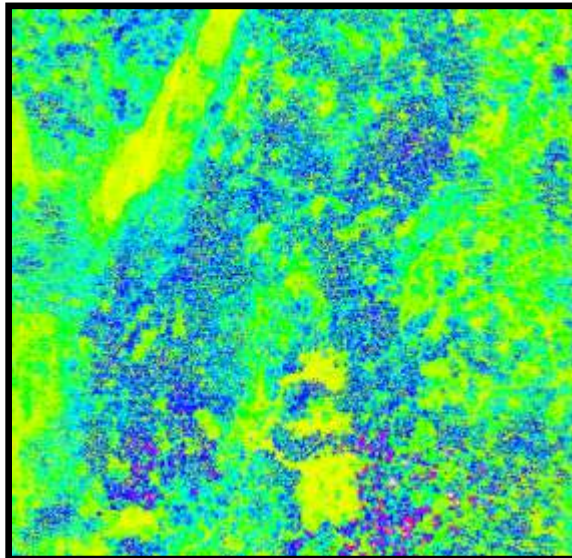

Figure 3: Raw data raster format in GRASS software

There are three main steps to utilize in order to filter LIDAR data. The three main steps were done using module 'v.lidar.edgedetection', 'v.lidar.growing' and 'v.lidar.correction'. v.lidar.edgedetection is the first of three steps to filter LiDAR data. The filter aims to recognize and extract attached and detached object such as buildings, bridges, power lines, trees and any other features in order to create a Digital Terrain Model. In particular, this module detects the edge of each single feature over the terrain surface of a LIDAR point surface. The output will be a vector map in which points has been classified as terrain, edge or unknown.

Next step is building contours for determining the areas of vegetation by using 'v.lidar.growing' module. This is the second of three steps to filter LIDAR data. The flter aims also to recognize and extract attached and detached object in order to create Digital Terrain Model. The module identifies which is the internal area of every object on a LiDAR point surface. The classification categories from v.lidar.edgedetection are now rasterized.

The final step to filter LIDAR data was done by using 'v.lidar.correction' module. This module was applied several times to the data, then a comparison between the observation and the

interpolation data can be made with Tychonov regulation parameter which only performed on Terrain Single Pulse points. From this step, it was detected that the gradient of the data had been minimized by the regularization of the parameter.

After DTM and DSM were fully generated, the next step is to determine the estimation of the canopy height. The canopy height was estimated by subtracting a DTM from the DSM. In this software, the difference between DTM and DSM was getting by using 'Raster Map Calculator' module

3.2   TERRASCAN

Terrascan is actually dedicated software for processing laser scanning points. It can easily handle millions of points with it versatile tools prove useful in many applications including survey transmission line, flood plains, proposed highways, stock piles, forest areas or for city models. In order to evaluate the quality of DTM and DSM results generate using GRASS OSS, the same LIDAR data was processed using TERRASCAN Commercial software. This software can read data either in .LAS or XYZ format. For this study, the data which is in XYZ format were in the WGS84 coordinate system where elevation representing geocentric height had been used in order to make synchronization to the data that had been used in GRASS software.

There are several steps to be completed to generate the DTM and DSM by using this software. **Figure 4** below shows the steps to process LIDAR data in Terrascan and Terramodeler Software.
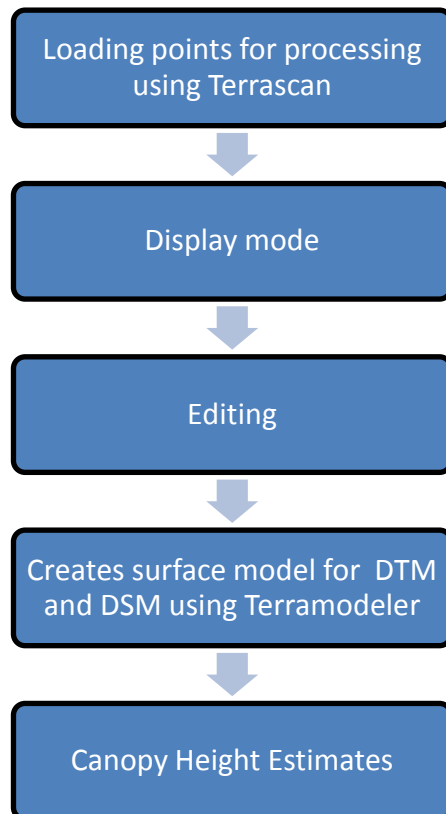
Loading points for processing using Terrascan

Display mode

Editing

Creates surface model for DTM and DSM using Terramodeler

Canopy Height Estimates

Figure 4: Processing steps in Terrascan and Terramodeler Software

Points are loaded using the Load Airborne Points Tool. Here, Terrascan will read all the selected files so the classification and manipulation of the data can be made. **Figure 5** Shows the LIDAR raw data based on point cloud view in Terrascan and the surface model raw LIDAR data 3D view in Terramodeler.
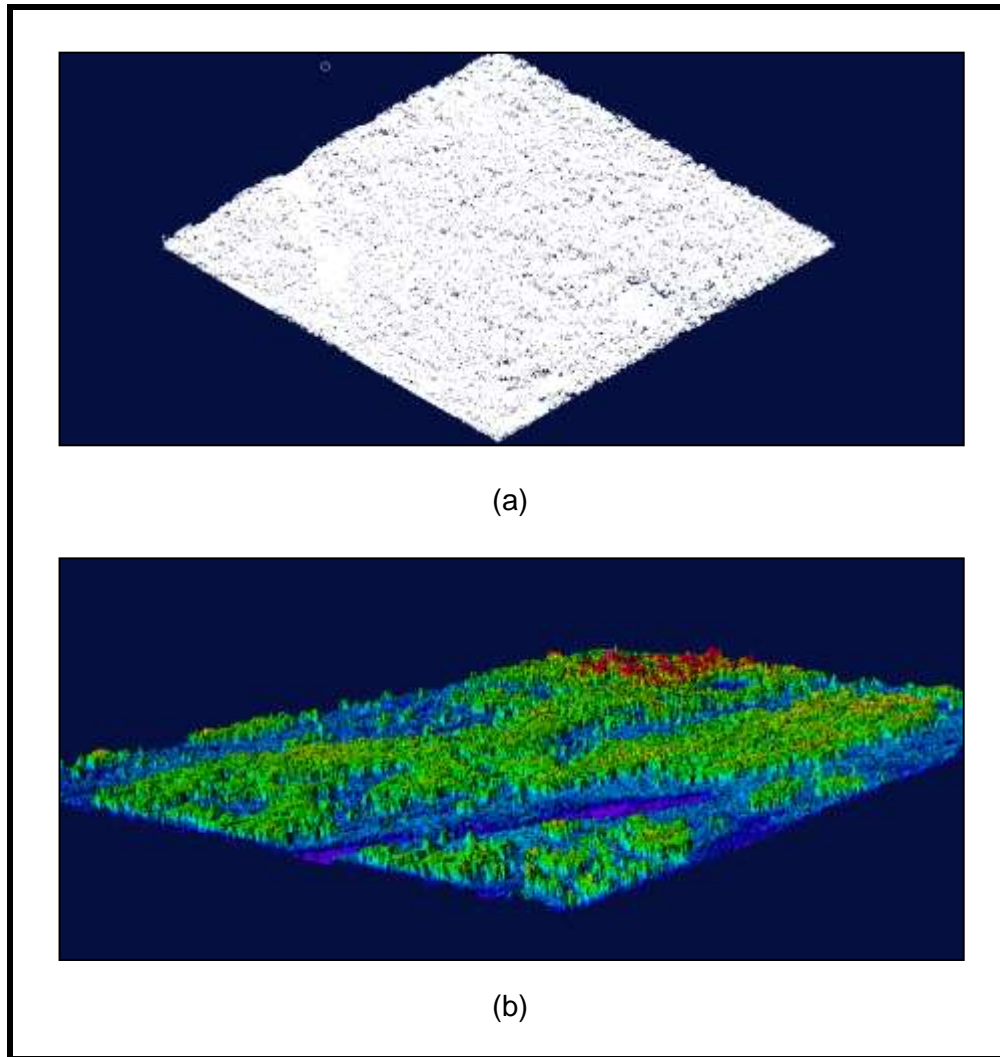

(a)

(b)

Figure 5: After Data loading shows (a) LIDAR raw data based on point cloud and (b) surface model raw LIDAR data 3D view

Editing process was followed after load the raw LIDAR data. This process include the classify points into point classes which is either in ground or vegetation class using macro tools processing. The result of the classification steps was shown as in the **Figure 6.**
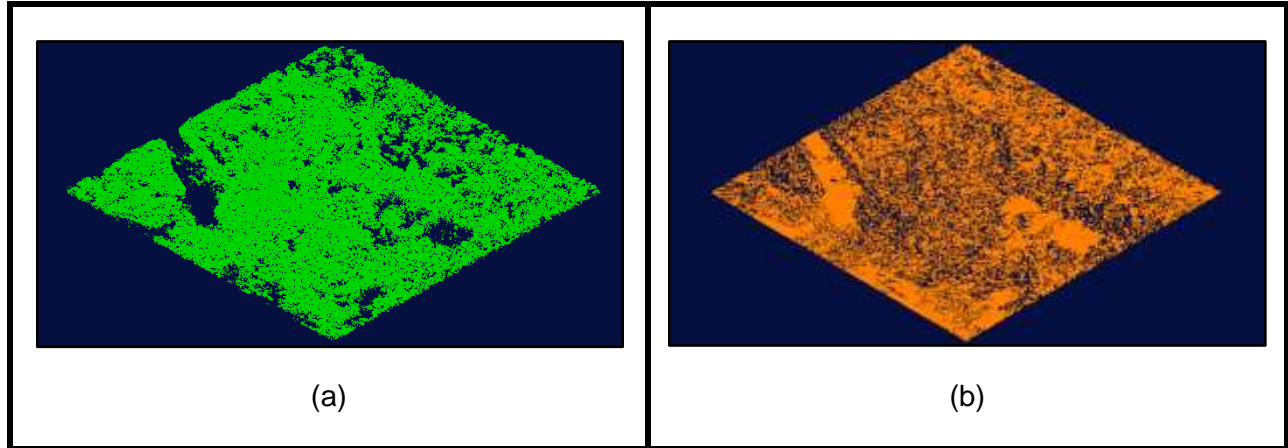
(a)

(b)

Figure 6: Visualization of (a) Vegetation points and (b) ground points

4.0   RESULT AND ANALYSIS

The DTM and DSM produced from GRASS software were in raster image. Figure 7 shows the DTM and DSM in raster image after going through filtering process using GRASS software
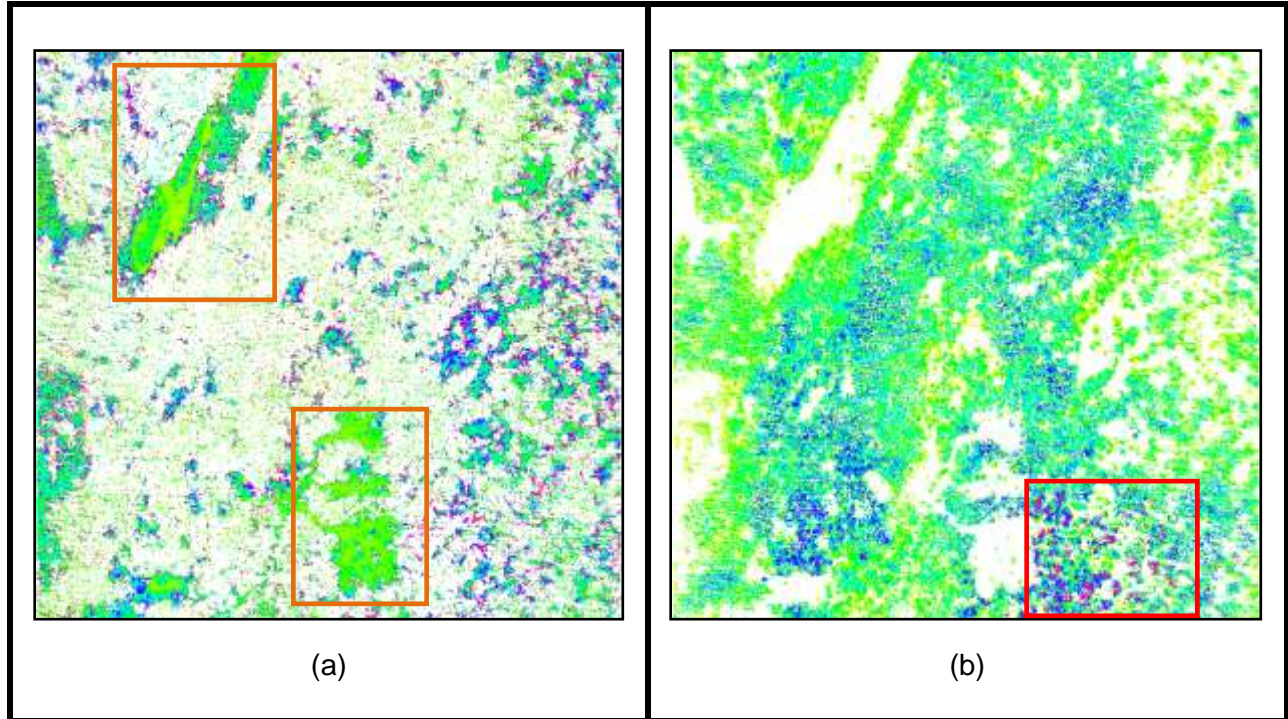.

(a)

(b)

Figure 7: Output after the filtering steps shows the (a) DTM and (b) DSM in raster image

The colors in the image represent the z value. Each color has its own value. The green color represent the lowest z value, it is followed by blue, purple and lastly red color. As we can see at

DTM image, there are clear green color areas which are in the orange box. The cleared green colored actually shows that the ground points in that area are very dense. It is because that area was a clear ground area that had not been covered from any vegetation compared to the other areas which have sparse ground points due to dense vegetation covered. In the DSM image, we can see that only a small part of the area have purple and red points which is in the red box. The purple and red points represent the high z value. It means that in this study area, there is only a very small part that has higher z value than any other parts of area.
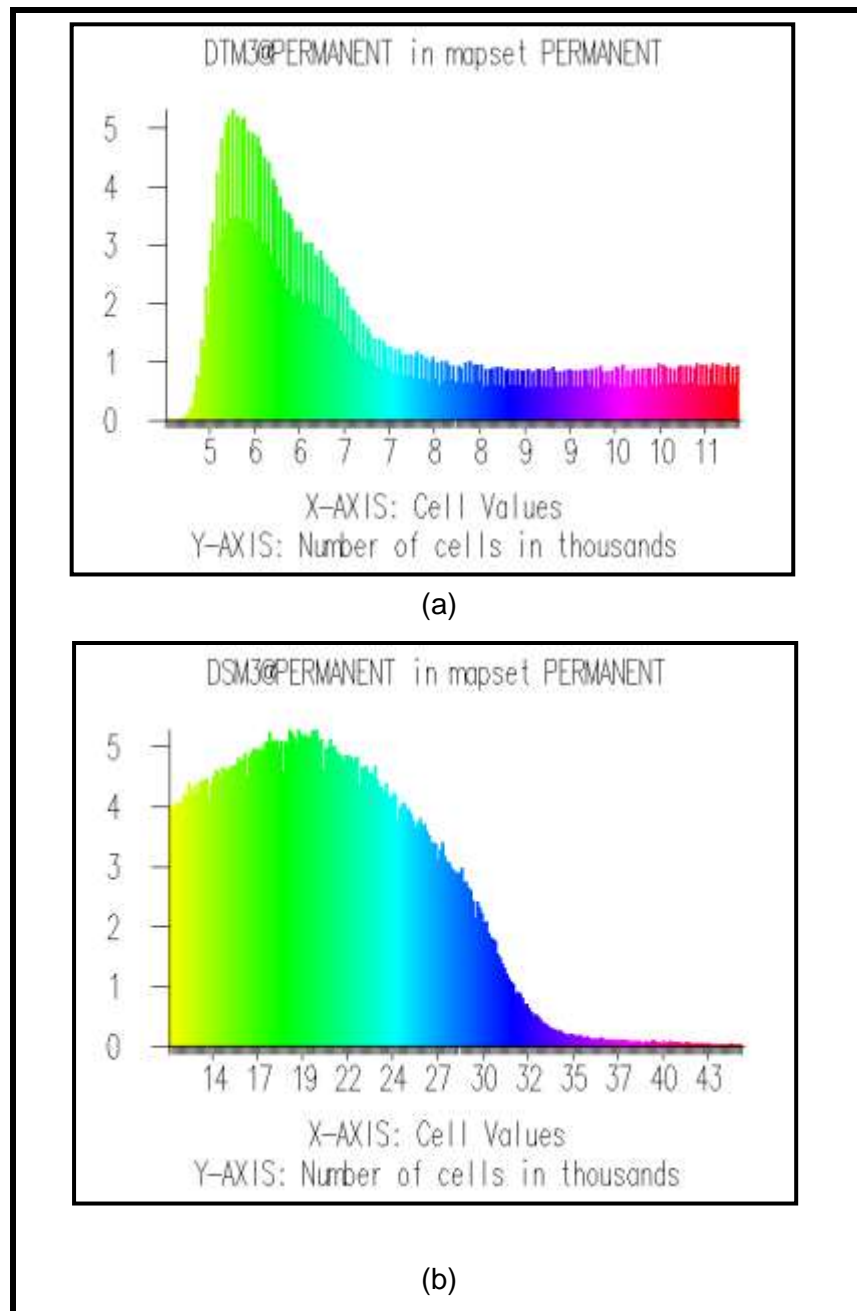

(a)

(b)

Figure 8: (a) Histogram of DTM and (b) Histogram of DSM.

Figure 8 shows the histograms of the DTM and DSM that had been produced using GRASS software. The histograms illustrate the number of cells in thousand that includes in the image in Figure 7 according to the z value.

The number of cells was decreased as the z value was increased. From the Histogram (a) it can be analyze that more than a half of the cells that was in DTM image have the z value for about five to six. However, there were also other cells that represent higher z value which are seven to eleven but only a small number compared to other lower z value.

Histogram (b) shows that the highest z value for the DSM image was fourty three and the lowest z value was fourteen while the highest number of cells was five thousand and the lowest number of cells was about less than one thousand cells. From Histogram (b) it can be analyze that most of the cell that was in DSM image have the z value for about fourteen to twenty seven. The outcome result for both software which is GRASS and Terrascan were already produced which are the DTM and DSM. However, these outcome results came in two different data format. The DTM and DSM produced from GRASS is in XYZ format while DTM and DSM produced from Terrascan is in Binary format. To ensure the advanced analysis can be conducted, these two different formats should have been synchronized which is in XYZ format. For the preliminary analysis, the DTM and DSM produced by GRASS are compared to the DTM and DSM produced by Terrascan according to their shaded relief view.
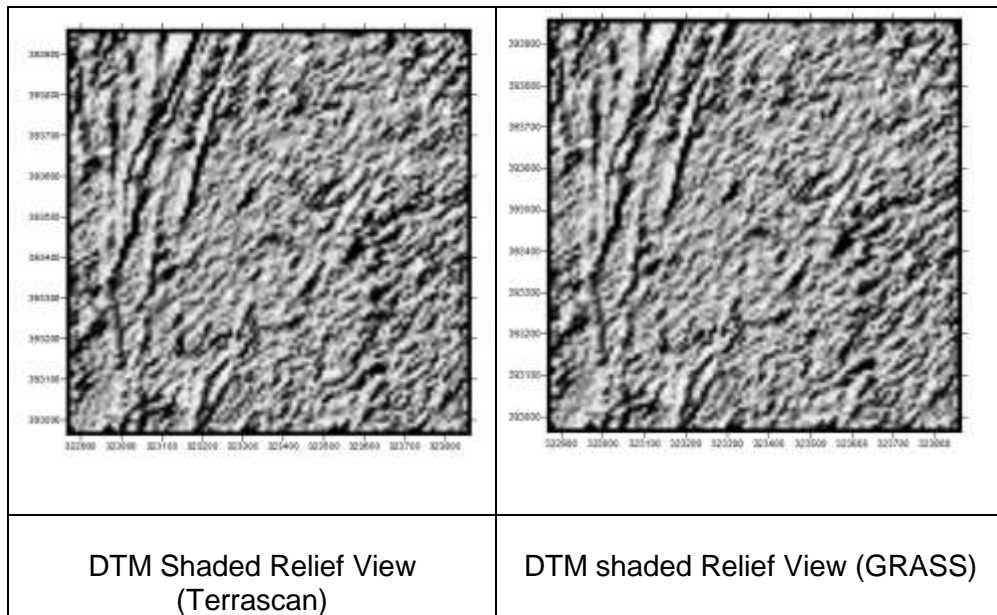

| DTM Shaded Relief View (Terrascan) | DTM shaded Relief View (GRASS) |

Figure 10: DTM Shaded Relief view

As a result, GRASS filtering has produced quite significant surface with the surface produced by Terrascan software which can see from **Figure 10**. It is clearly seen that the surface is almost the same for both DTM that had been generated. It seems like it is very difficult to detect any difference by only looking to this shaded relief view. From this result it is easily can be analyze that DTM produced using GRASS software have quite same quality level with the DTM produced using Terrascan Software.
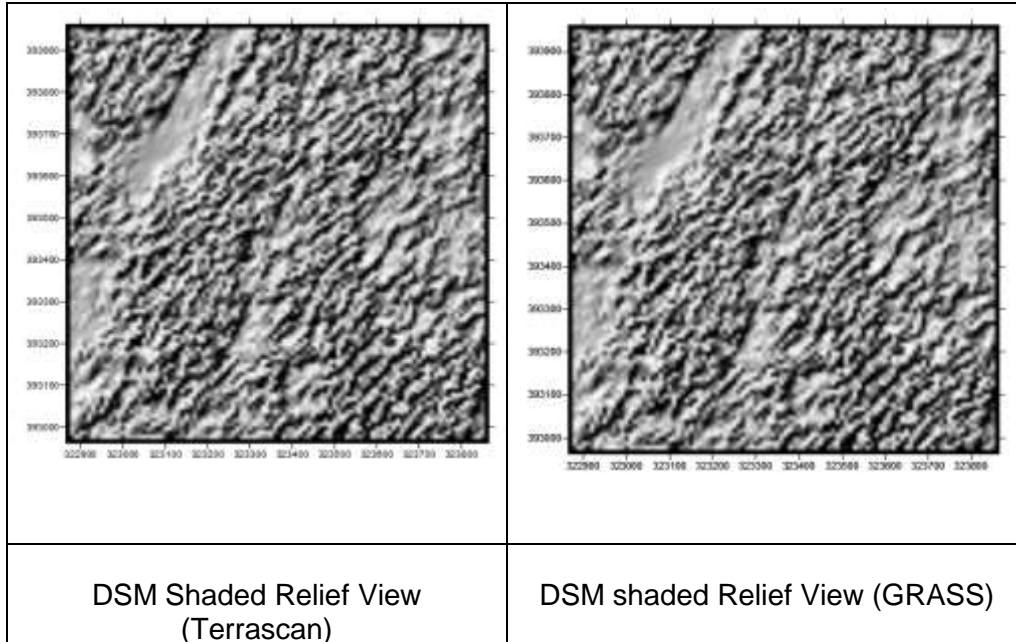
DSM Shaded Relief View (Terrascan)	DSM shaded Relief View (GRASS)

Figure 11: DSM Shaded Relief View

By look at the result in **Figure 11**, both surfaces have quite significance surface eventhough it was produced using two different softaware. It is also very difficult to detect the difference between these two DSM by just look at the Shaded Relief view. From this result it is also can be analyze that DSM produced using GRASS software have quite same quality level with the DSM produced using Terrascan Software.

Since the DTM and DSM were already produced, hence the estimation of the canopy height can also to be known. It is by distraction of the DTM from the DSM. The result for estimation of the canopy height from DSM and DTM produced by Terrascan software shows in **Figure 12** while the estimation of the canopy height from DSM and DTM produced by GRASS Software shows in **Figure 13**.
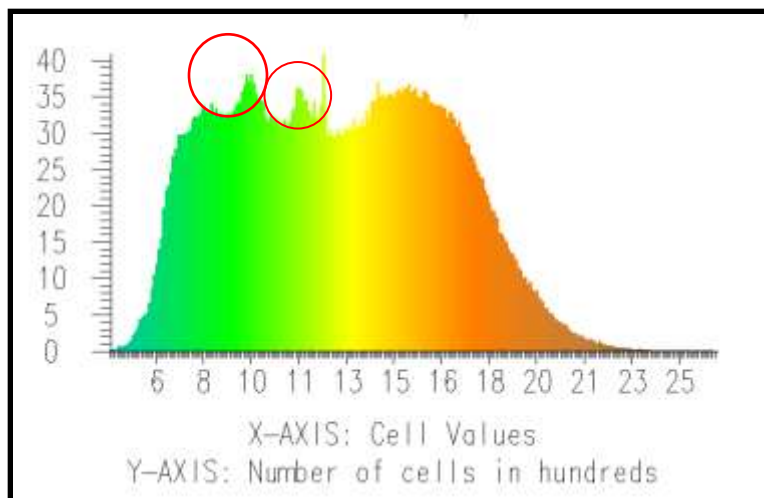

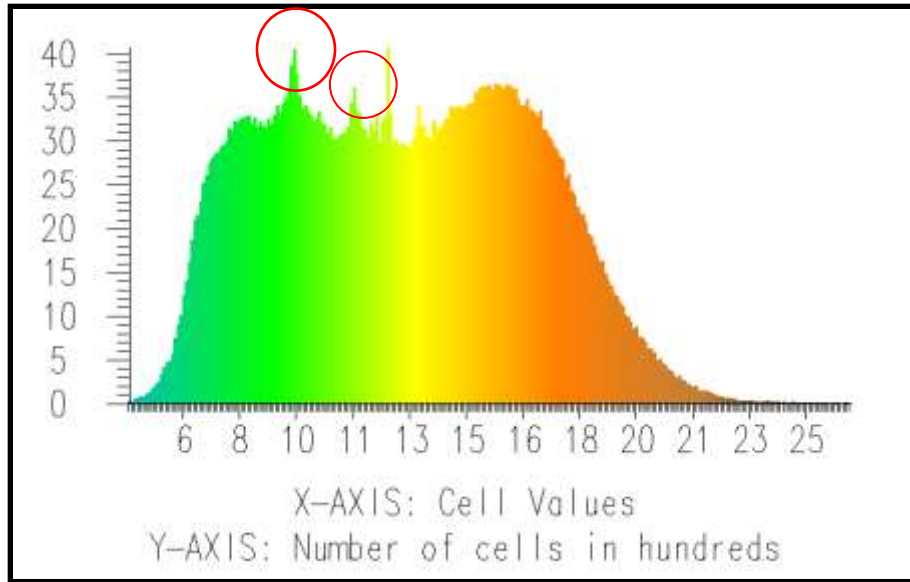**Figure 12: Canopy Height estimation from Terrascan Software**

Figure 13: Canopy Height estimation from GRASS Software

Figure 12 and **Figure 13** shows the histogram for the estimation of the canopy height. The Histograms illustrate the number of cells in thousand according to the z value. It was clearly seen that these two histograms just look alike in terms of its trends. The number of cells starts to increase rapidly at the value of six for the z value and began to move ups and downs as the value of z increase. However, the number of cells starts to decrease at the value of eighteen for the z value. Eventhough the trends for the two histograms are same, there are also slight difference between these two histograms by look at in the red circle. Histogram in **Figure 12** shows that it have the higher number of cells compared to Histogram in Figure 13. From these canopy height estimation results, it can be analyze that both canopy did not have the same number of cells.

For further analysis In order to analyze the potential of GRASS software, the minimum and maximum value for the z value of the canopy produced from Terrascan and GRASS software had been identified. The minimum and maximum value was given at the **Table 1** below. From Table 1, it shows that there was only small difference detected between the minimum z values. There is higher difference in maximum values compared to minimum values which is 0.295. However, it is still can be considered as long as the difference did not achieve 0.5.

Table 1: Comparison on minimum and maximum z value for canopy produced from GRASS and Terrascan Software.

Software	Minimum value (z)	Maximum Value (z)
Terrascan Commercial Software	5.019	26.656
GRASS OSS	5.020	26.361
Difference	0.001	0.295

4.0   CONCLUSION

The point view of the LIDAR data in **Figure 9** shows the points that already had been classified according to their class which are ground and vegetation. It shows that the ground points are seems to be in a small number compared to vegetation points. Due to dense vegetation cover, it is quite difficult actually to extract true ground points from the LIDAR data. More analysis is needed to improve the accuracy of the classification result. However due to the result obtain from the comparison between the models produced from two different software, it was proving that GRASS software has a good potential in order to produce the same quality DTM and DSM as Terrascan software eventhough it is an Open Source Software.

REFERENCES

Dubayah. R. O. and Drake. J. B (2001). *Lidar Remote Sensing for Forestry Applications*. University of Maryland.

Jihn. F. J. (2005). *Comparison of Forest Canopy Height Derived Using Lidar Data and Aerial Photos*, Taipei, Taiwan.

Li, Zhilin., Cher, J. and Baltsavias, E. (2008) Advances in Photogrammetry, Remote Sensing and Spatial Information Science: 2008 ISPRS Congress Book.

Ustin. L. S. (2004). *Remote Sensing for Natural Resouce Management and Environmental Monitoring.* California.

AUTHORS

Nurhanisah Hashim is a Bachelor Degree candidate at the Faculty of Geoinformation and Real Estate, Universiti Teknologi Malaysia taking Geomatic Engineering. She holds Diploma in Land Surveying from Universiti Teknologi Malaysia.

Dr. Zulkepli Majid is a senior lecturer at the Faculty of Geoinformation and Real Estate, Universiti Teknologi Malaysia. Currently, he is the head of Photogrammetry and Laser Scanning Research Group. He holds B.Sc.(Land Surveying) degree, M.Sc (Photogrammetry) degree and a Ph.D (Medical Photogrammetry) degree from Universiti Teknologi Malaysia, Malaysia. His research interests lie in the areas of Photogrammetry and Laser Scanning for various applications.

Dr. Halim Setan is a professor at the Faculty of Geoinformation Science and Engineering, Universiti Teknologi Malaysia. He holds B.Sc. (Hons.) in Surveying and Mapping Sciences from North East London Polytechnic (England), M.Sc. in Geodetic Science from Ohio State University (USA) and Ph.D from City University, London (England). His current research interests focus on precise 3D measurement, deformation monitoring, least squares estimation and 3D modelling.