Indian Journal of Science and Technology, Vol 8(20), DOI: 10.17485/ijst/2015/v8i20/78374, August 2015
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Integration of Near Infrared Image and Probabilistic Classifier to Increase the Classification Accuracy of Point Clouds

Dongmoon Kim*

Department of GIS Engineering, Namseoul University, Republic of Korea;
david@nsu.ac.kr

Abstract

Essential to the establishment of such 3D spatial information are the laser scanning technology to obtain high-precision 3D point clouds and the photography-metric camera to obtain high-resolution multispectral image information. TLS (Terrestrial laser scanner) is a high precision positioning technique to monitor the behavior and change of structures and natural topography. 3D point clouds of natural environments relevant to problems in geomorphology (rivers, coastal environments, cliffs, ...) often require classification of the data into elementary relevant classes[12]. There are such old techniques to classify massive point clouds as NDVI and PCA (Principal Component Analysis). Although they have many different advantages, it is extremely difficult for them to maximize their advantages to function as stand-alone techniques and overcome their disadvantages. This study thus set out to investigate the integration of NDVI and a probabilistic classifier PCA, SVM (Support Vector Machine), LDA (Linear Discriminant Analysis))[1] to obtain land cover information in a natural state through TLS and improve the classification accuracy of resulting massive point clouds.

Keywords: NDVI (Normalized Difference Vegetation), PCA (Principal Component Analysis), TLS (Terrestrial Laser Scanner)

1. Introduction

Terrestrial Laser Scanner (TLS) is now frequently used in earth sciences studies to achieve greater precision and completeness in surveying natural environments than what was feasible a few years ago[2,8]. TLS features a high-precision positioning system to monitor and analyze the structures and topography of various forms in an objective and accurate manner[14]. The information obtained with it, however, contains the information about the various objects in the surroundings as well as the information of the desired objects, which makes it essential to classify and extract by item[6]. The image data-based NDVI technique classifies the vegetation indexes of plants by acquiring the near-infrared image information of objects. Since the technique can generate final analysis results by fusing image data with three-dimensional point-based data, it should be accompanied by research on ways to fuse images with three-dimensional point-based data[4,5,13]. One of its disadvantages is the difficulty with obtaining the three-dimensional location accuracy of objects to be classified because it is demanding to fuse images with three-dimensional point-based data[3,15,16]. It is also considerably hard to secure matching accuracy between the spatial resolution of images and three-dimensional point-based data with the technique[10,11,17]. Making use of location and spatial relation characteristics, a statistical probability classification technique can be applied to classify point data in order to solve those problems. There are several types of probabilistic classifiers and one of them is a classification technique like 3D PCA, SVM, LDA[1]. Such a classification technique records high classification accuracy in certain environments but struggles to secure accuracy in environments where there are various covers. The old point group data classification technique has been restricted to certain areas in most

* Author for correspondence

cases, which calls for research on how to increase accuracy by integrating NDVI with a probabilistic classifier-based classification technique[7,9]. The present study thus made an attempt at the integration of NDVI and a probabilistic classifier-based classification technique to acquire land cover information in a natural state through TLS and increase the classification accuracy of resulting massive point clouds. As for methodology, the study examined the old analysis techniques for their advantages and disadvantages, devised an integration technique, and reviewed analysis techniques, thus producing final results.

2. Data Classification Techniques

2.1 Probabilistic Classifier

LDA can reduce dimensions to the minimum while protecting class information as much as possible, but there should be much more additional data to assess classification errors. If the distributions are significantly non-Gaussian, the LDA projections may not preserve complex structure in the data needed for classification. LDA will also fail if discriminatory information is not in the mean but in the variance of the data. When variables (x1, x2, x3,…xp) related to one another are observed, PCA can produce a small number of new variables to secure the maximum amount of information found in those variables. In other words, one major interest in the mutually related variables in the number of p (x1, x2, x3,…xp) is measuring the variation of those variables. Variation refers to the amount of information found in the variables. Principle component analysis presents a process of producing new variables through the linear combinations of the original ones in the order of volume of variation found in them. SVM classifies two groups by finding an arbitrary non-linear boundary in vector space and is known to find mathematically optimal answers under the given conditions, but it is difficult to implement it because expressions of terms in the number of N^2 or more should be calculated. The general involved principle is as follows: when two groups of data are provided, one can learn to divide them by measuring distance between data in the two groups, identifying their centers and obtaining an optimal hyper plane at the centers. Using SVM, one can learn how to discriminate two types of data appropriately on a computer and make predictions for new data. Even when soft margins are applied, however, SVM cannot perform effectively for a real issue of non-linear classification. The theory employed a process of entering given data, setting a group of identification criteria through certain algorithm-based learning and predicting the type of new given data. It has a couple of advantages including being able to solve various problems, increasing accuracy according to the accuracy of learning data and being effective for non-linear identification problems (three-dimensional). It also has some disadvantages, as well, such as having a need for data for learning, making it difficult to set criteria for boundaries between groups and having a hard time effectively distinguishing input data of various characteristics.

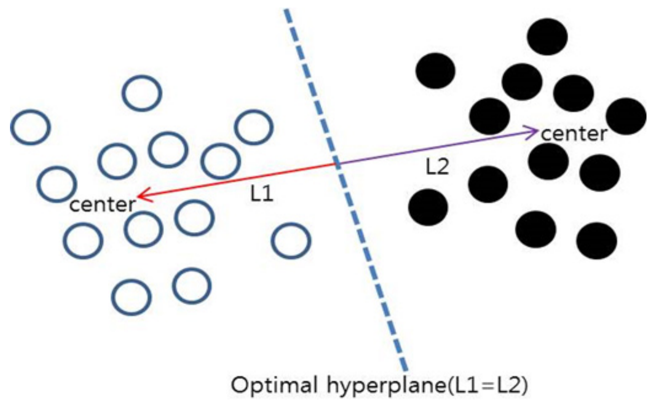

Figure 1. The basic principle of probabilistic classifier.

The above-mentioned statistics-based probability sampling technique has a disadvantage in needing various criteria for accurate learning data and boundary division Figure 1. When the technique is used by itself, it is extremely difficult to produce classification results while maintaining the desired level of accuracy.

2.2 Analysis of Near-infrared Imaging Data

Near-infrared imaging data can help to extract vegetation with a vegetation index. There are a couple of vegetation indexes including the normalized difference vegetation index, infrared index, and soil adjusted vegetation index. The technique can identify vegetation (Equation 1) by analyzing only the images of objects and present grid-based results, thus boasting a high level of utilization. However, it is difficult to obtain the three-dimensional location accuracy of objects to be classified with the technique due to the difficulty of fusing images with three-dimensional point-based data. The technique also makes it extremely difficult to secure matching accuracy between the spatial resolution of images and three-dimensional point-based data.

NDVI=[(Near_Infrared-Red_ray)/(Near_Infrared+Red_ray)] (1)

3. Integration Technique Application and Review

There is a need for an integration technique to maximize the merits of the old ones in order to classify large-volume point data effectively. By analyzing the advantages and disadvantages of NDVI and a probabilistic classifier-based classification technique, the study devised the following integration technique Figure 2.

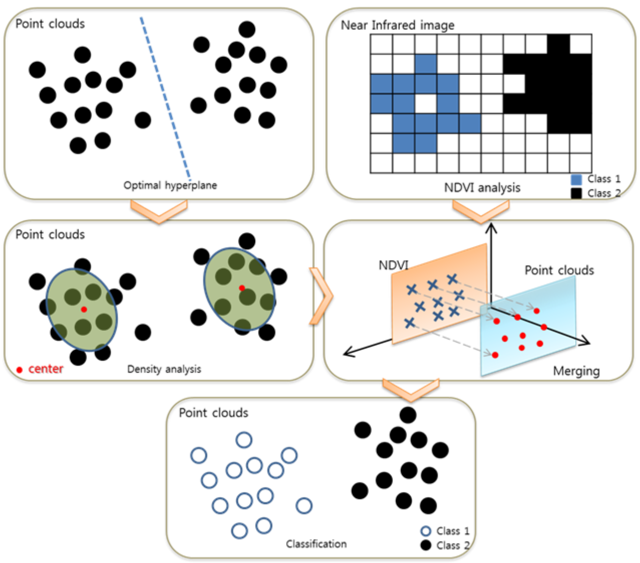

Figure 2. An integration technique of probabilistic classifiers and NDVI.

The investigator thus devised one as follows: this integration technique can help to extract the surface of the earth (Digital Surface Model) to classify large-volume point data obtained from natural topography and group point cloud data with point data other than the surface points by using a probability classification technique. Grouping falls in the stage of objectifying the object data. When objectified, data are used to obtain near-infrared imaging data for cover classification, NDVI analysis, and matching between point data and raster data.

3.1 Data Collection and Arrangement

The research data included terrestrial LiDAR data and NIR (near-infrared) imaging data. As seen in Figure 3, the former provide high-density point data scanned from the general natural topography with a combination of artificial structures and vegetation, whereas the latter provide multispectral images shot at the same time as terrestrial LiDAR data with a near-infrared camera

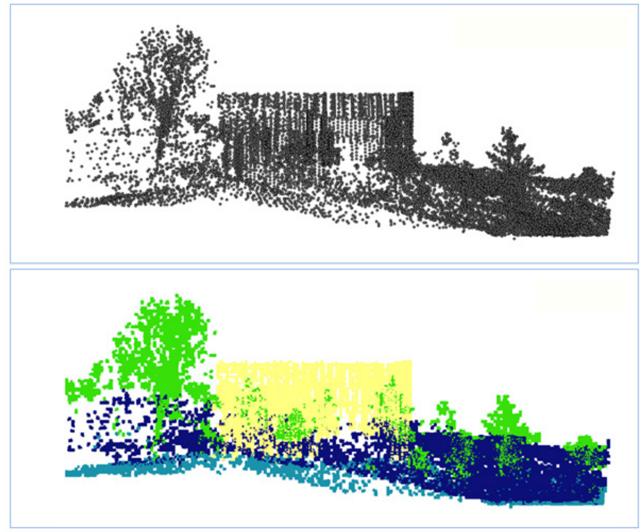

Figure 3. The research data (Terrestrial LiDAR data and NIR data).

3.2 Classification and Extraction of DSM (Digital Surface Model)

This is a process of extracting the surfaces from the large-volume point data obtained. Natural topography contains many different elements including the surface, vegetation and structure, which means that the surface extraction process should come before accurate classification. DSM classification with point data requires a normalization process of selecting only altitude data and converting them to a certain height. As seen in Figures 4 and 5, the investigator selected the minimum values among the point data within a certain range and normalized them to the standard height.

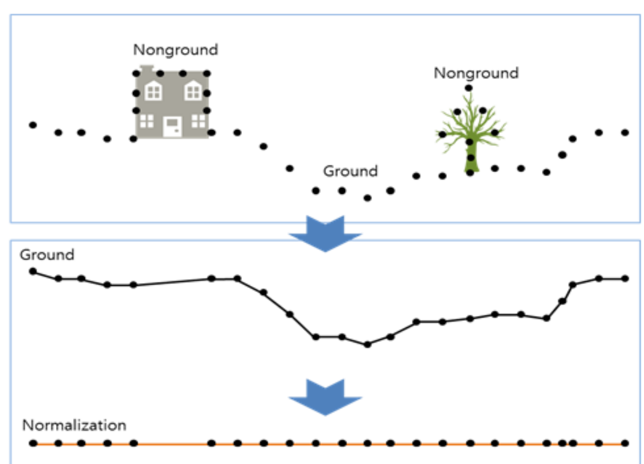

Figure 4. Surface and non-surface.

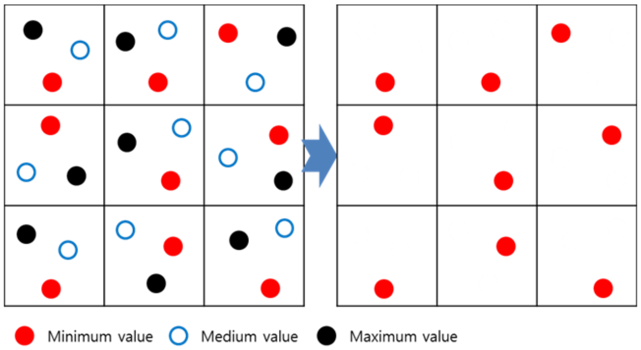

Figure 5. Classification standard of surface and non-surface data.

There are a variety of statistics-based probability classification techniques including PCA, LDA, and SVM. Since they are two-dimensional techniques, they need another technique applicable to 3D point data. Several studies have recently been conducted to classify 3D data with PCA. Such a probability classification technique can be expanded to 3D for application and used in complex 3D multi-class classification Figure 6. In the present study, the investigator set neighborhood balls as the basic analysis unit as seen in Figure 7, selected the points inside the balls, and assigned a class to them. When classifying the classes, the investigator calculated 3D distance with the standard of ball center and judged whether it was within the radius of a ball.

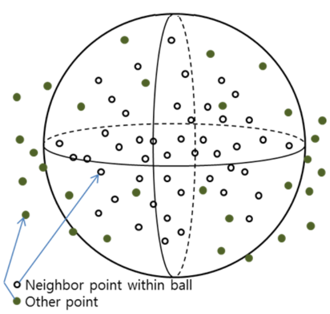

Figure 6. Point data in and outside neighborhood balls.

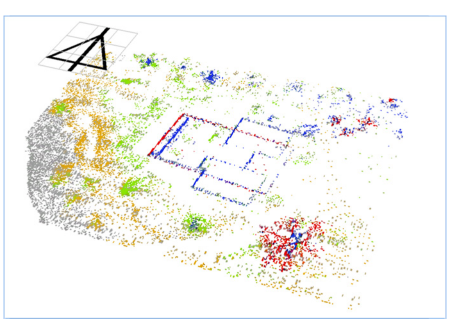

Figure 7. Grouping results of point data.

Classified in the previous stage, the classes are based on the location correlations in 3D space and thus have a difficult time with reflecting the actual surface features into class classification (each class cannot be defined as a surface, structure, or vegetation). Such a problem can be resolved by assigning the surface features to each segment classified into a class. In the present study, the first stage of assigning the surface features involved the selection of the center of each segment by calculating density and geometric center based on three-dimensional location correlativity Figure 8. In addition, the number and size of adjacent points or density were also calculated for each unit area in certain domains, the function of which can be used to calculate the density surface at locations where points are concentrated (Equation 2). It is also possible to select a center by calculating one in 3D space and considering the point density.

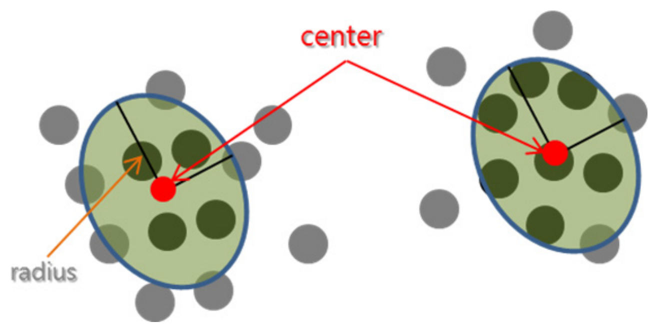

Figure 8. Selection of a center in each segment.

$$Density = (\sum_{i=1}^{n}(P_i \times w_i)) / E_{x,y} \tag{2}$$

p refers to the number of point data for each class; w refers to the weighted value to be applied to the density analysis of each point data; and $E_{x,y}$ refers to the domain of each segment.

NDVI is an index to show the distribution and activity of vegetation and can be calculated with reflectivity differences among green plants. The calculation results are in the range of +1 ~ -1. Their usability can be raised through normalization in the stages of 0~255 in the integer unit with the Equation 3.

$$NDVI = [\frac{(Near\_Infrared - Red\_ray)}{(Near\_Infrared + Red\_ray)} + 1] \times 128 \tag{3}$$

When the calculated value is close to 255, it means the plant has a high level of vitality. When it is close to 0, it means that the plant has a low level of vitality. Figure 9 shows the NDVI classification results for the research scope.

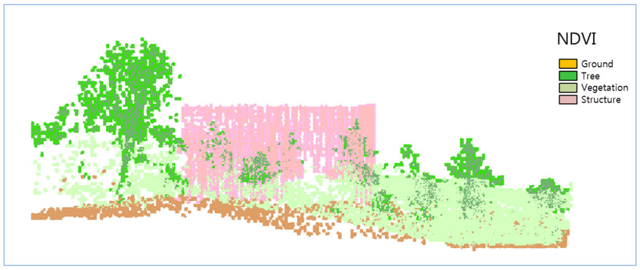

Figure 9. NDVI classification results.

3.3 Assignment of Cover Characteristics to the Center of an Object

After classifying point data in the object unit with a probability classification technique, it is time to define the detailed characteristics of each object Figure 10. In the present study, the investigator analyzed the detailed characteristics of a location that corresponded to the center of each object through NDVI analysis and then defined the cover characteristics of each point data as attributes by matching and assigning the results based on location relations.

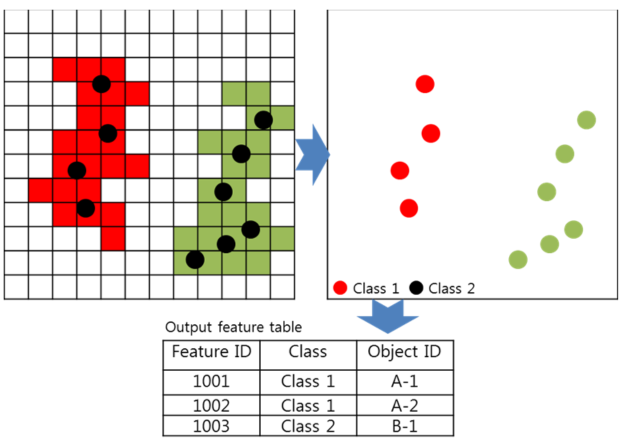

Figure 10. The assignment of cover characteristics to the center of an object.

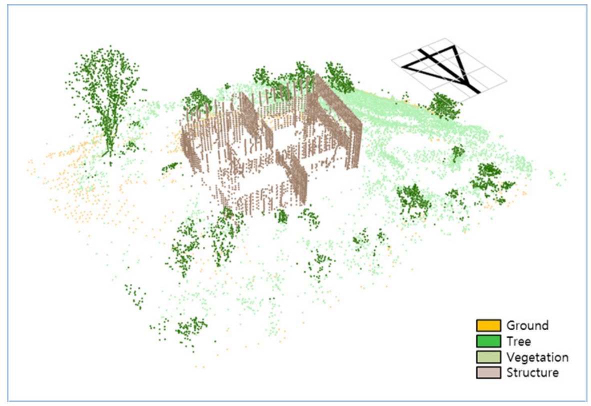

Figure 11. Final classification results.

Figure 11 presents the final classification results when an integration technique was applied.

4. Analysis and Consideration

The investigator compared point data and actual topography and obtained standard deviation to assess the classification results by the application of an integration technique for accuracy Figure 12. As seen in Table 1, the results of classification based on an integration technique were within RMSE (Root Mean Square Error) .001m, which implies that an integration technique can be applied to the classification of high-density point cloud data such as terrestrial LiDAR data.

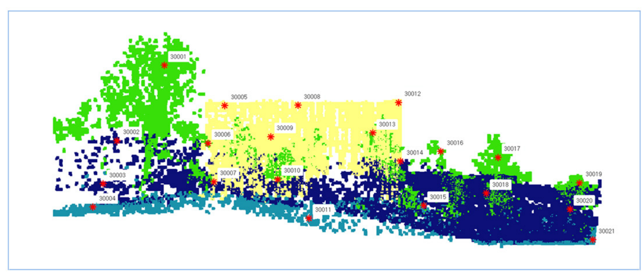

Figure 12. The POI (Points of Interest).

Table 1. Accuracy of classification based on an integration technique

Coordinate	POI (Point of Interest)											RMSE /3D
	3001	*3002*	*3003*	*3004*	*3005*	*3006*	*3007*	*3008*	*3009*	*3010*	*3011*	
X	0.029	0.035	0.038	0.036	0.034	0.038	0.031	0.029	0.028	0.030	0.033	
Y	0.037	0.034	0.035	0.028	0.039	0.031	0.027	0.031	0.036	0.032	0.029	
Z	0.041	0.040	0.045	0.048	0.043	0.049	0.041	0.043	0.046	0.042	0.048	
Coordinate	*3012*	*3013*	*3014*	*3015*	*3016*	*3017*	*3018*	*3019*	*3020*	*3021*	*RMSE/m*	
X	0.031	0.036	0.031	0.029	0.028	0.032	0.028	0.036	0.032	0.029	0.0007	
Y	0.032	0.029	0.032	0.031	0.031	0.029	0.032	0.029	0.036	0.031	0.0006	0.001
Z	0.036	0.038	0.039	0.040	0.037	0.040	0.041	0.042	0.043	0.041	0.0008	

With a relatively large RMS value, "A" covers the boundary between vegetation and surface and between artificial structure and surface. It is apparent that classification errors often take place at class boundaries, which is attributed to the characteristics of a probability classification technique based on three-dimensional spatial relations. The current situation calls for attempts at research to increase the classification accuracy of boundaries. That is, there is a need for studies on evaluation algorithms for classification and factor selection and application according to the three-dimensional space distribution characteristics of classification objects. Studies such as the present one can help to figure out the applicability and usability of an integration technique.

5. Results

The present study investigated an integration technique to classify topographical covers with TLS data and reached the following conclusions: first, it analyzed NDVI and a probabilistic classifier-based classification technique for advantages and disadvantages and identified what should be supplemented and reinforced; secondly, it devised a technique to integrate each analysis technique and predicted the effective classification of land covers with the technique. The class classification accuracy of the proposed technique is within .001m, which means that the technique holds excellent usability as a classification technique for point cloud data; and finally, the proposed integration technique is expected to increase the classification objectivity and accuracy of objects, develop an array of application areas with massive point clouds, and make a contribution to the improved accuracy of cover classification in a natural environment in follow-up study.

6. Acknowledgment

Funding for this paper was provided by Namseoul university.

7. References

1. Brodu N, Lague D. 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: applications in geomorphology. ISPRS Journal of Photogrammetry and Remote Sensing. 2012 Mar; 68:121–34.
2. Kim DM. Construction of multi-dimensional ortho-images with a digital camera the multi-image connection method. Journal of Digital Convergence. 2014; 12(8):295–302.
3. Kim DM. Development of an ST3E model for the 3D modeling of the seafloor space. Journal of Next Generation Information Technology. 2013 Oct; 4(8):257–63.
4. Kim DM, Kim SH. Behavior analysis of coastal structure using terrestrial LiDAR data. Journal of Next Generation Information Technology. 2013 Oct; 4(8):456–62.
5. Kim DM, Park JK. Inclination analysis of structure using terrestrial LiDAR data. Journal of Next Generation Information Technology. 2013 Oct; 4(8):463–9.
6. Shin ES, Kim DM. Extraction of vegetation group through the fusion of multi-spectral images and terrestrial LiDAR data. International Journal of Applied Engineering Research. 2014; 9(22):14597–608.
7. Franke R. Smooth interpolation of scattered data by local thin plate splines. Comp & Maths with Appls. 1982; 8(4):237–81.
8. Hatfield JL, Asrar G, Kanemasu ET. Intercepted photosynthetically active radiation estimated by spectral reflectance. Remote Sensing of Environment. Science Direct. 1984; 14(1-3):65–75.
9. Huete AR, Dian K, Miura T, Rodriguez EP, Gao X, Ferreira G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment. Science Direct. 2002; 83(1-2):195–213.
10. Jonathan PS, Hu J, Robert EK, Arthur JL, Brian DC, Craig AD, Thomas DO'R. Use of airborne and terrestrial LiDAR to detect ground displacement hazards to water systems. Journal of Surveying Engineering. ASCE. 2009 Aug; 135(3):113–24.
11. Chuang K-C, Ma C-C. Multidimensional dynamic displacement and strain measurement using an intensity demodulation-based fiber bragg grating sensing system. Journal of Lightwave Technology. 2010; 28(13):1897–905.
12. Brodu N, Lague D. 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: applications in geomorphology. ISPRS. 2012; 68:121–34.
13. Verbree E. Piece wise linear complex representation through conforming delaunay tetrahedronization. GIScience 2006. LNCS. 2006; 4197:385–8.
14. Wang Y, Liang X, Flener C, Kukko A, Kaartinen H, Kurkela M, Vaaja M, Hyyppa H, Alho P. 3D Modeling of coarse fluvial sediments based on mobile laser scanning data. Remote Sensing. 2013; 5(9):4571–92. ISSN 2072-4292.
15. Yanmina W, Guolib W. Integrated registration of range images from terrestrial LiDAR. The International Archives of the Photogrammetry. Remote Sensing and Spatial Information Sciences. 2008; 37(B3b):361–6.
16. Zhang B. Automatic terrain extraction using multiple image pair and back matching. ASPRS 2006 Annual Conference; Reno. Nevada. 2006 May 1-5. p. 1–12.
17. Zhu J, Shimizu K. The effect of pfl gene knockout on the metabolism for optically pure D-lactate production by Escherichia coli. Appl Microbiol Biotechnol. 2004 Apr; 64(3):367–75.