DEM EXTRACTION FROM AIRBORNE LIDAR POINT CLOUD IN THICK-FORESTED AREAS VIA CONVOLUTIONAL NEURAL NETWORK

Yongjun Zhang[1], Sizhe Xiang[1], Yi Wan[1], Hui Cao[1], Yimin Luo[1,2], Zhi Zheng[1]

[1]School of Remote Sensing and Information Engineering
Wuhan University, Wuhan 430079, China
[2] Division of Imaging Sciences and Biomedical Engineering Research
King's College London, London SE1 7EH, U.K.
{zhangyj,xiangsizhe, yi.wan,huicao}@whu.edu.cn,
yimin.luo@kcl.ac.uk, zhengzhi@whu.edu.cn

IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9323201

ABSTRACT

Digital Elevation Model (DEM),representing the height of the earth terrain, is one of the crucial geographic information products. One of the main data source of DEM is the airborne LiDAR point cloud with its non-ground-reflections filtered out. Point cloud filtering in thick-forested areas is difficult without enough ground control points when using conventional methods. In this paper, a supervised method is proposed to handle the problem of automatic DEM extraction with little ground control points. The design of the method is inspired by the successful application of the convolutional neural networks (CNN) in the image super resolution (SR) process. First, with the given LiDAR point cloud, the digital surface model (DSM) is resampled with regular grid. Then, by learning the spatial autocorrelation between the DSM and its corresponding DEM, a robust CNN model is established.Finally, the DEM in thick-forested areas can be generated from the DSM with the trained model. Experimental results at two different mountain sites in China validate the effectiveness of the proposed method of high-precision DEM generation.

***Index Terms*—** Digital Elevation Model (DEM), Digital Surface Model (DSM), convolutional neural network (CNN), LiDAR point cloud, DEM extraction

1. INTRODUCTION

Digital Elevation Model (DEM) is a grid-wise model which expresses the height of the earth terrain in a given coordinate. DEM is widely used in various applications, such as topographic mapping, landslide monitoring, etc. Automatic acquisition of accurate DEM is one of the key problems of remote sensing [1, 2].

As a new technology of active aeronautical remote sensing, the ability of penetrating into shallow tree's canopies makes LiDAR point clouds a valuable data source for accurate DEM extraction[3], especially when dealing with complex terrains, such as shallow forested areas, coastal zones, and desert areas[4, 5, 6]. Generally, DEM in these areas are always extracted by point cloud filtering with assistance of accurate ground control points [7] and laborious manual supplement. However, due to the limited number of ground control points in thick-forested terrains, the above-mentioned methods can hardly work.

To deal with the problem of DEM retrieval in thick-forested areas, Luo et al.[8] proposed a supervised method based on deep neural network, where the DSM and its corresponding were decomposed into paired 1-dimension digital terrain signals and their latent relationship were studied by a stacked autoencoder. In consequence, with the trained deep neural network architecture and the DSM data, the corresponding DEM can be retrieved automatically. Nevertheless, since the method decomposed the DSM/DEM pairs into 1-dimension terrain signals, which neglected the correlation information in the point cloud, the retrieved DEM results could not fit the terrain perfectly, especially over the protrusive terrains.

Considering that the low resolution image and high resolution image share the same low-frequency information, convolutional neural networks are utilized to learn the nonlinear degradation relationship between the high/low resolution image pairs in the task of image super resolution. Similar to the high/low images, the DSM and DEM are also highly relevant. Therefore, in this paper, we rethought the problem of DEM extraction from the point of view of correlation learning rather than focusing on directly non-ground-reflection removal from the LiDAR point cloud and propose a **C**onvolutional neural network for **D**EM **E**xtraction (CDE). More precisely, the main contributions of this work are listed as follows:

1. a simple but effective method, containing a convolutional neural network, is designed to extract DEM from DSM.

2. The proposed method CDE takes the spatial correlation between DSM and DEM into consideration, which outperforms the traditional method in thick-forested areas.

2. METHODOLOGY

In this section, the data preprocessing procedure was introduced, the proposed model was detailed and the utilized loss function was presented.

2.1. Data Preprocessing

The proposed method CDE aims at thick-forested terrains, so the accident cases of barren land and buildings in LiDAR point clouds are unexpected and need to be removed from the training samples. Therefore, a data preprocessing procedure is added to select suitable training samples.

As an apriori assumption, the fluctuations of DSM and DEM in thick-forested terrains are extremely similar and the deviation between the DSM and the DEM floats within a certain range. However, the DSMs and DEMs over the barren lands and the residential areas do not follow this assumption. The DSM and the DEM share the same elevation over barren lands and have constant deviations over

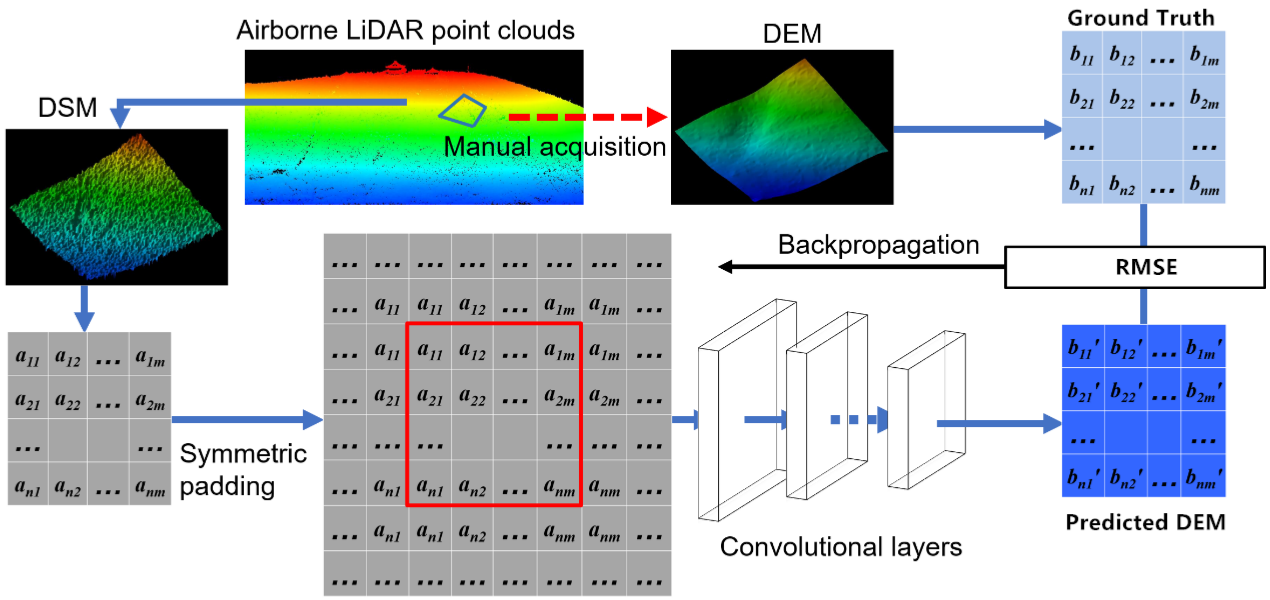


Fig. 1. The flowchart of the proposed CNN-based method

flat building roofs. ON the contrary, in thick-forested areas, the DSM contains lots of sharp features and have variable elevation deviation compared with the corresponding DEM.

Considering the above-mentioned priori, a moving-window based optimal sample selection method is applied to filter out barren land and buildings from the LiDAR point clouds to obtain training data. The window size is set as 20×20 pixels, and the moving speed is set as 4 pixels in this paper. The DSM samples are re-sampled from the selected LiDAR point clouds with regular grid. Meanwhile, with the assistance of given ground control points, the high-precision DEM labels are manually edited at some reachable locations, where measurements can be implemented. In addition, to filter out the LiDAR point cloud over barren lands and residential areas, the following two assumptions are followed:

a. To filter out barren land, a DSM/DEM deviation matrix is generated by subtracting DEM from its DSM counterpart.If the median DSM/DEM deviation within the moving-window is smaller than 0.5m, the area is judged as barren land and the DSM and DEM pairs located within the window is removed from the training data.

b. If half of the adjacent deviation values within the window is lower than 0.1m, the area is considered as containing building roofs and the corresponding DSM and DEM pairs are removed from point cloud data.

It needs to be mentioned that, theoretically, the elevation difference of DSM and DEM in barren land should be 0, as well as the elevation deviation of adjacent positions in building areas. In this paper, to avoid the error brought by original LiDAR data acquisition and point cloud regularization, the thresholds of filtering barren land and buildings are set as 0.5m and 0.1m rather than 0.

After the above-mentioned data preprocessing step, a much better training dataset is established to serve the subsequent model training.

2.2. Model Design

The flowchart of the proposed CDE method is illustrated in Fig.1. The convolution layers used for DEM extraction is modified from those in image super-resolution models. Generally, a 7-layers CNN structure equipped with PReLU[12] is designed to learn the spatial correlation between DSM and DEM pairs. On one hand, the spatial correlation between DSM and DEM is relative simple compared to that between HR images and LR images in image super-resolution tasks. Therefore, too deep structure will result in overfitting problem (i.e, 20-layers VDSR[9]) and is unnecessary in the DEM prediction task. On the other hand, too shallow structures (i.e, 3-layers SR-CNN) cannot fully represent the features. Thus, we design a 7-layers model that balance the model complexity and the ability of feature representation.

As shown in Table 1, the model is composed of three parts, namely, image-to-feature transform, feature extraction, and terrain reconstruc-tion. The image-to-feature transform part contains a single convolution operation, followed by batch normalization and PReLU[12]. This part transforms DSM samples into feature domain for further feature ex-traction. In the feature extraction parts, five tandem convolution units are stacked to learn the spatial correlation between the DSM and DEM pairs. It should be noted that the up-sampling operations are not in-cluded in the modified model, since the DSM and the DEM share the same size. After the feature extraction, a convolution operation trans-forms the features back to the image domain and reconstruct the terrain. With the above three parts, a predicted DEM can be reconstructed from the model.

Table 1. Parameter Details of the Proposed Method

Layer	Structure	Details
Image-to-feature transform	Convolution	kernel size=3
	Batchnorm	
PReLU	activation function	
Feature extraction	5 conv units	kernel size=3
PReLU	activation function	
Terrain reconstruction	Convolution	kernel size=3

2.3. Loss Function

The root-mean-square error (RMSE), which can express both the systematic and the stochastic characteristics of the data, is selected as the loss function to evaluate the model performance rather than the Euclidean distance used in Luo et al.[8]. The formulation of RMSE is denoted as (1).

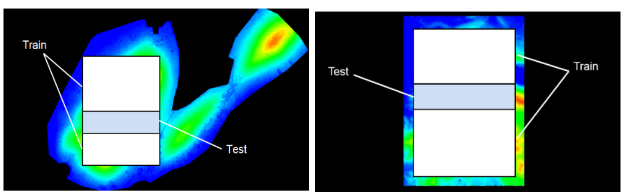
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \frac{1}{width \times height} \left\| x_i - x_i^{'} \right\|^2}{n}} \tag{1}$$

Where x_i denotes the ground truth DEM, and $x_i^{'}$ denotes the predicted DEM.

3. EXPERIMENTS

3.1. Data and Settings

In this paper, two airborne LiDAR point cloud datasets, acquired over the thick-forested areas in Fujian province, China and Hainan province China, were used for experiments. The DEMs with 5m spatial resolution were used as the ground truth. To test the performance of the proposed model, accurate artificial ground control points in both of the two test areas were provided for verification. As shown in Fig.2, the two datasets were both randomly divided into independent training set and testing set with 9:1 ratio.



(a) Fujian Dataset (b) Hainan Dataset

Fig. 2. Two datasets of thick-forested areas

All the training process was implemented on Caffe Library [11]. During the training process, the initial learning rate was set to 0.01, and the batch size was 4. After 10,000 iterations, the learning rate decreased to 0.008. The training process ended at the iteration 100,000. To train the model, the DSM/DEM pairs were normalized and the symmetric padding was utilized to offset the convolution-caused information loss and the extra-introduced mistakes.

3.2. Results and Analysis

This section evaluated the performance of CDE. To the best of our knowledge, CNN-based methods have rarely been employed for DEM extraction in thick-forested terrains. Therefore, spatial interpolation method and the most related work [8] were taken as comparison methods.

In order to show the effects of the proposed method intuitively, the DEM extraction results were visualized with a commercial software named LiDAR-Suite. This software has been widely used for generating interpolation DEM results from original DSM data due to its superior filtering performance.

As shown in Fig.3, the Fig.3(a) and Fig.3(d) were spatial interpolation results of Fujian and Hainan datasets. The spatial interpolation method filtered the DEM out from the original DSM with assistance of some ground control points, and then interpolated it by the traditional inverse distance weight interpolation method. It is obvious that the results of the interpolation-based method were like a combination of numerous finely shredded planes and could not meet the visual requirements. Moreover, most of the key information at the mountain peaks and ridge lines were lost during the interpolation process, which heavily harmed the global reconstruction accuracy.

Table 2. Quantitative evaluation of the DEM extraction results (meter)

Method	DNN method [8]	CDE
Experiment 1: Fujian Dataset		
MAE	5.6924	**4.8718**
AVE	0.7289	**0.6969**
RMSE	0.6278	**0.5077**
Experiment 2: Hainan Dataset		
MAE	**5.0431**	5.0736
AVE	0.9575	**0.9045**
RMSE	**0.7358**	0.7454

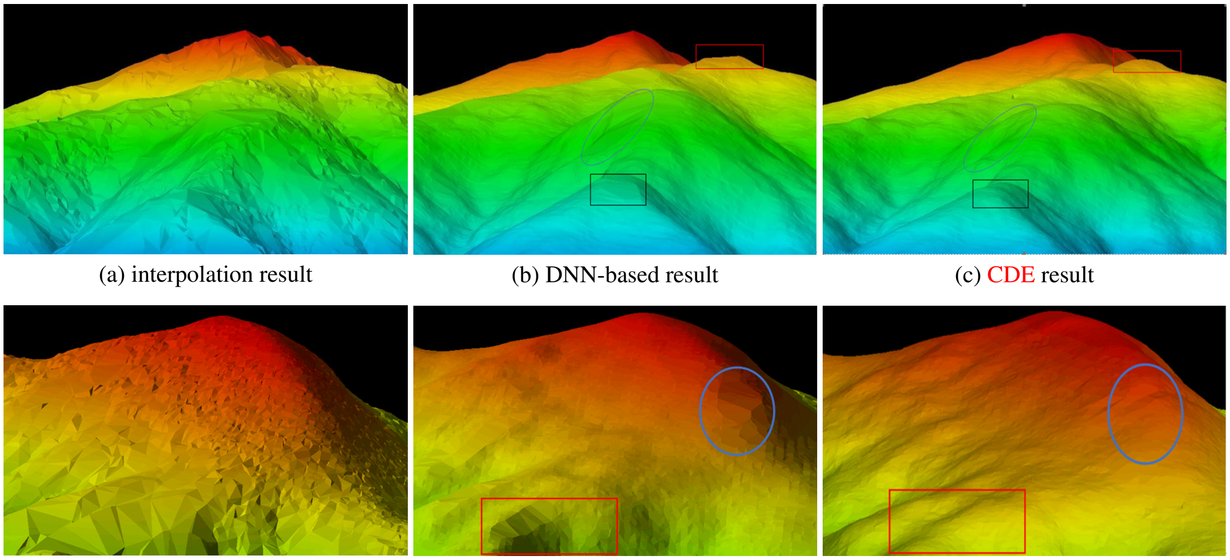
Compared to the interpolation results, as depicted in Fig.3(b) and Fig.3(e), the DNN-based method could improve the results to a large extent. The slope of DNN-based result was smoother, and reconstruction results at ridge lines were also closer to the real terrain. However, terrain protrusions in the two mountain areas were still difficult to be correctly reconstructed, such as the ellipse areas in Fig.3(b) and Fig.3(e). Besides, from the rectangle regions in Fig.3(b) and Fig.3(e), it could be seen that there were still some small retrieval mistakes on the mountain peaks. Furthermore, as the terrain of Hainan data was sharper than Fujian data, the mistakes became more obvious on some sudden scarps.

Fig.3(c) and Fig.3(f) displayed the results of CDE. Generally,CDE reached a better visualization results compared to its competitors. Since the proposed method resolved the problem from the idea of image SR process, which considered the spatial correlation between the DSM/DEM pairs, the results of CDE were closer to the real terrain, especially at finely terrain protrusions. More precisely, compared to the DNN-based method, this method obtained much more accurate results and at ridge lines and the results on mountain peaks were also sharper. For example, it could be seen from the ellipse area of the Fujian dataset that the DNN-based method could not fit the terrain well, while the result of CDE was much better. Meanwhile, the results of rectangle regions in Fig.3(c) and Fig.3(f) demonstrated that CDE extracted the terrain information better and was less likely to make mistakes.

To test the superiority of the proposed model, quantitative results were also taken into analysis. The maximum error (MAE), average error (AVE), and root-mean-square error(RMSE) were selected as evaluation criterion. The evaluation results were depicted in Table 2.

According to Table 2, CDE outperformed DNN-based method with a large margin on the Fujian test area while performed similar on the Hainan test area. Combining the visualization results in Fig.3, the experimental results were explicable. Since the reconstruction superiority of CDE mainly lay in complex protrusions, while the terrain of Hainan dataset was relatively smoother and had little protrusions, therefore, the difference of the two methods at Hainan dataset were not as obvious as the results in Fujian dataset.

Concretely, the MAE of CDE were namely 4.8718m and 5.0736m at the two datasets, which was equal to the resolution of the LiDAR point clouds.The results were acceptable considering the complexity of the thick-forested terrains. Besides, the AVE and RMSE were under 1m with both of the two sets of the test data, which indicated that the reconstructed DEM can fit the real terrain well.



(a) interpolation result (b) DNN-based result (c) CDE result

(d) interpolation result (e) DNN-based result (f) CDE result

Fig. 3. Visualization of the DEM Extraction Results: (a)(b)(c) were the results of Fujian dataset; (d)(e)(f) were the results of Hainan dataset

4. CONCLUSION

In previous research works, DEM extraction in thick-forested areas highly relied on spatial interpolation and manual supplement, which needed numerous ground points as assistance. To address the problem and improve the automation of DEM extraction over such terrains, a novel CNN-based method was proposed in this paper.The proposed solves the problem of DEM extraction in thick-forested areas within the idea of image super resolution, which works well without the need of ground control points.

Future research will focus on accuracy improvement and the design of more comprehensive and robust models for DEM extraction from LiDAR point clouds.

5. ACKNOWLEDGEMENTS

This work was supported in part by the National Key Research and Development Program of China with project number 2017YFB0503004, and the National Natural Science Foundation of China with project number 41871368.

6. REFERENCES

[1] Y. Chen, L. Cheng, M. Li, J. Wang, L. Tong, and K. Yang, "Multiscale grid method for detection and reconstruction of building roofs from airborne LiDAR data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014.

[2] B. Sirmacek, H. Taubenböck, P. Reinartz, and M. Ehlers, "Performance evaluation for 3-D city model generation of six different DSMs from air- and spaceborne sensors," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2012.

[3] B. Bigdeli, H. Amini Amirkolaee, and P. Pahlavani, "DTM extraction under forest canopy using LiDAR data and a modified invasive weed optimization algorithm," *Remote Sensing of Environment*, 2018.

[4] S. A. White and Y. Wang, "Utilizing DEMs derived from LIDAR data to analyze morphologic change in the North Carolina coastline," *Remote Sensing of Environment*, 2003.

[5] C. Hladik and M. Alber, "Accuracy assessment and correction of a LIDAR-derived salt marsh digital elevation model," *Remote Sensing of Environment*, 2012.

[6] P. Tompalski, N. C. Coops, J. C. White, M. A. Wulder, and A. Yuill, "Characterizing streams and riparian areas with airborne laser scanning data," *Remote Sensing of Environment*, 2017.

[7] H. Xiang, C. Cao, H. Jia, M. Xu, and R. B. Myneni, "The analysis on the accuracy of DEM retrieval by the ground lidar point cloud data extraction methods in mountain forest areas," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2012.

[8] Y. Luo, H. Ma, and L. Zhou, "DEM Retrieval from Airborne LiDAR Point Clouds in Mountain Areas via Deep Neural Networks," *IEEE Geoscience and Remote Sensing Letters*, 2017.

[9] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.

[10] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[11] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A Discriminative Feature Learning Approach for Deep Face Recognition," Tech. Rep.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015.