Improvement of 3D Modeling Efficiency and Accuracy of Earthwork Site by Noise Processing Using Deep Learning and Structure from Motion

N. Yabuki[a], Y. Sakamoto[a] and T. Fukuda[a]

[a]Division of Sustainable Energy and Environmental Engineering, Osaka University, Japan
E-mail: yabuki@see.eng.osaka-u.ac.jp, fukuda@see.eng.osaka-u.ac.jp

Abstract –

Nowadays, we can relatively easily create 3D models of earthwork construction sites by taking many pictures from multiple viewpoints using Unmanned Aerial Vehicles (UAV) and by applying Structure from Motion (SfM) technique to the pictures. However, since many construction machines are moving at the construction site, they become noise or disturb the model when generating a 3D model. Thus, it is necessary to take measures such as removing the noise, i.e., construction machinery after surveying, stopping construction work during the survey, or surveying on weekends when all machines are parked at proper places, which can result in reduced efficiency and productivity. On the other hand, object detection technology using deep learning is making great progress. Therefore, this research proposes an efficient and accurate UAV photogrammetry system that generates high-precision 3D models by detecting and removing construction machinery from UAV aerial images in advance. First, the object detection method by deep learning is used to detect the construction machinery. Next, the parts of the construction machines are removed from the images, and the removed parts are completed by image interpolation or compensation. Finally, a 3D model is created using SfM. We developed a system based on the proposed methodology and applied it to an actual earthwork construction site. The result showed that accuracy and efficiency have been enhanced by using this system.

Keywords –

Photogrammetry; Unmanned Aerial Vehicle; Structure from Motion, Deep Learning, Object Detection

1 Introduction

It has become easy to take multi-viewpoint images thanks to the Unmanned Aerial Vehicle (UAV) technology. In addition, photogrammetry technology that combines aerial images taken by UAV and the Structure from Motion (SfM) that generates a 3D model from multi-viewpoint images was developed. These technologies have been used in various fields, such as forest management, situational assessment at the time of disasters and construction field. Application to progress management and soil volume management using 3D models is expected to contribute greatly to productivity improvement. However, many construction machines are working at the construction site. When photogrammetry is performed by UAV and a 3D model is generated in such a situation, the working construction machine causes errors and noise. Therefore, it is necessary to take measures such as surveying with construction work stopped, surveying during weekends, or removing noise manually after surveying, resulting in low efficiency. Also, surveying with UAV takes several hours, so it is difficult to do it frequently.

On the other hand, object detection technology using deep learning is making great progress. While various high-accuracy methods have been proposed, many studies have been conducted to apply the detection results to various systems.

In this study, the object detection method using deep learning is adapted to the construction site by fine tuning, and the construction machinery is automatically detected, removed, and complemented from UAV aerial images in advance. As a result, we propose a system that can generate 3D models with high accuracy even under construction work. Also, verify the detection accuracy of construction machinery.

2 Literature Review

2.1 Research on Utilization of 3D Modeling of Construction Sites

Many methods using 3D models for construction site management and surveying have been proposed. In recent years, with the spread of SfM technology, it has become possible to easily generate a 3D model by taking a large number of pictures with a digital camera. The use of 3D models is expected to improve productivity at construction sites.

Omari et al. [1] proposed an efficient method to generate a 3D model of a construction site by combining a 3D laser scanning and photogrammetry. In surveying using a laser scanner, it is necessary to measure at many points. Efficient surveying was made possible by reducing the number of surveying points and making up for the lack of scanning by using photogrammetry. The proposed method succeeded in reducing 75% of the time required for surveying the construction site.

Gore et al. [2] proposed a photo-based 3D modeling method for space planning to improve construction site safety and productivity. With the conventional visual inspection and space planning method that relies on 2D drawings, the construction site is usually very complicated and difficult to capture, therefore we proposed a method for generating 3D models from multi-view photographs.

Yamaguchi et al. [3] proposed a system for efficient photogrammetry using UAV. The system automatically detects and removes construction machinery from UAV aerial images, and generates a highly accurate 3D model using SfM. An object detection method based on the features in the image was used to detect construction machinery. Detection was performed for each construction machinery using the Joint-HOG feature value generated by Real AdaBoost in two steps from the HOG feature value, which is a co-occurrence table of the feature value histograms of the gradient direction of the brightness of the local region. Construction machinery was classified into six classes: backhoe, wheel roller, heavy dump, hydraulic breaker, bulldozer, dump truck, and learning was conducted, accuracy was verified, the average detection rate by object class was 54.5%, which was low.

2.2 Object Detection Method

In recent years, various high-precision object detection methods have been proposed and are expected to be used in a wide range of fields such as application to automatic driving technology of automobiles. Many object detection methods using classifiers based on convolutional neural networks such as SSD [4] and YOLO [5] have been proposed, and each has high detection accuracy. The feature of SSD is that it is designed to be able to detect multi-scale from various output layers. However, at the stage of prediction at the low layer, there is a disadvantage that the accuracy is lowered because the feature quantity obtained in the subsequent layers cannot be captured. YOLO is a fast object detection algorithm. Since the object can be detected by looking at the entire image, the possibility of misrecognizing the background as an object has decreased, but it is difficult to detect when many small objects are included.

In the construction field, to improve productivity, attempts are being made to detect construction machinery using these evolving object detection algorithms and to develop systems that utilize the detection results.

Yabuki et al. [6] proposed a system for improving the efficiency of managing many photos taken at construction sites and disaster areas. The system detects various construction machines, workers, and signboards in photographs and automatically classifies them into folders. By using SSD for object detection and transfer learning, we obtained high accuracy from a few learning data sets.

Kim et al. [7] detected objects of construction machinery from photos to improve the accuracy of the vision-based construction site monitoring method to capture the situation at the construction site in real-time. R-FCN was used as the construction machinery detection method, and high accuracy was obtained from a few data sets by performing transfer learning.

While object detection based on deep learning from UAV aerial images has also been attempted. Matija et al. detected an airplane from UAV aerial images using the object detection method YOLO of a convolutional neural network. The detection accuracy is 97.5%, and the object detection method based on deep learning is useful for detecting objects from aerial images.

In this paper, an object detection method based on a convolutional neural network is used for construction machinery detection from UAV aerial images, aiming to enable more accurate detection and removal than when using Joint-HOG.

3 Proposed System

3.1 Overview of the Proposed System

The flow of the system proposed in this study is shown in Fig. 1. First, UAV is used to photograph the construction site under construction work. Next, construction machinery is detected from all photos. Next, the detected construction machinery part is masked and image complementation is performed. Finally, the goal is to create a highly accurate 3D model efficiently by creating a 3D model using SfM.

3.2 Detection of Construction Machinery

In this paper, You Only Look Once v3 (YOLOv3)

[8], which is an object detection method based on a convolutional neural network, is used to detect construction machinery. YOLOv3 has high precision and detecting small objects in images with high accuracy. In addition, YOLOv3 uses Feature Pyramid Network (FPN) to cope with the detection of different scales and aims to improve accuracy by capturing more features. For example, SSD does not use FPN, and the feature obtained in subsequent layers cannot be used in lower layers, but YOLOv3 overcomes this problem.

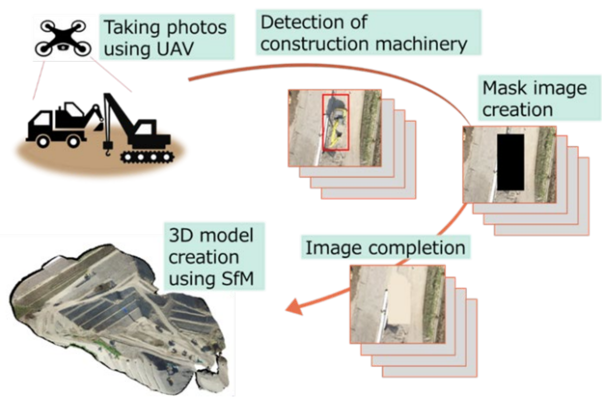


Figure 1. The Flow of the Proposed System

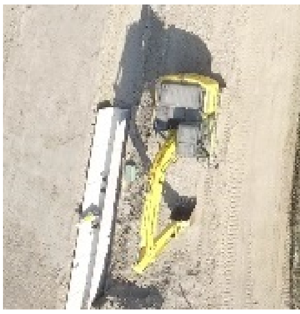
From the study of Matija et al. that YOLO is useful for detecting objects from UAV aerial images, it is expected that high detection accuracy can be obtained even in the detection of construction machinery by using YOLOv3, which was developed from YOLO. Also, because photogrammetry using UAV aerial photography is performed from a high altitude, the construction machinery that is the detection targets of this study often appear small. Therefore, YOLOv3, which is good at detecting small objects and has high detection accuracy, is considered an appropriate detection method.

In YOLOv3, learned models of general objects such as people and dogs are released. In this study, detection is performed by creating and fine-tuning construction machinery data set using the publicly learned model.

3.3 Detecting Object Types and Creating a Dataset

In this paper, 6 types of construction machinery are detected. We defined six classes: backhoe, bulldozer, road roller, dump truck, mixer truck, and car (Fig. 2). The class of ‘car’ was defined as passenger cars or light trucks moving in the construction site. By visual identification from UAV aerial images, 6 types of construction machinery were classified as shown in Table 1.

Based on the classification, a dataset was created from the UAV aerial images. There were 2,182 images included in a training dataset and 219 images included in a test dataset.



(a) backhoe

(b) bulldozer

(c) road roller

(d) concrete mixer truck

(e) dump truck

(f) car

Figure 2. Examples of construction machinery to be detected (a) backhoe (b) bulldozer (c) road roller (d) concrete mixer truck (e) dump truck (f) car

Table 1. The numbers of classified objects

Object class	Number of objects
backhoe	213
bulldozer	19
road roller	24
dump truck	185
mixer truck	18
car	350

3.4 Fine Tuning of YOLOv3

YOLOv3 was fine-tuned using the data-set described in Section 3.3. By using this result, the construction machinery is detected from the aerial image from UAV.

4 Validation

4.1 Detection accuracy verification method

An index called mean Average Precision (mAP) was used for verification. The mAP is an average value for the entire class, calculated from Average Precision (AP), which calculates the precision for each defined class. When evaluating using mAP, it is necessary to obtain Intersection over Union (IoU). IoU is the index that shows how accurately an object is detected. It can be calculated that the common part of the ground truth and the prediction area divided by the area of sum (Fig. 3).

$$IoU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \tag{1}$$

where

B_p: Predicted bounding box

B_{gt}: Ground truth bounding box

In order to calculate the mAP, the relationship between the ground truth area and the prediction area is important. These two relationships can be divided into four types True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) (Fig. 3). TP indicates a state in which an object existing in the image is correctly detected. FP indicates a state in which an object that does not exist in the image is detected (false detection). TN recognizes that the object does not exist and indicates a state in which it does not actually exist. FN indicates a state in which an object actually exists in the image but is not detected.

In this paper, when IoU ≧ 0.5, it is defined as TP, and AP and mAP are calculated. First, the precision AP for each class is calculated. Then, mAP which is the average value of the whole is calculated.

$$AP = TP/(FP + TP) \tag{2}$$

$$mAP = \frac{1}{M}\sum_{i=1}^{M} AP \tag{3}$$

where

M : The number of classes

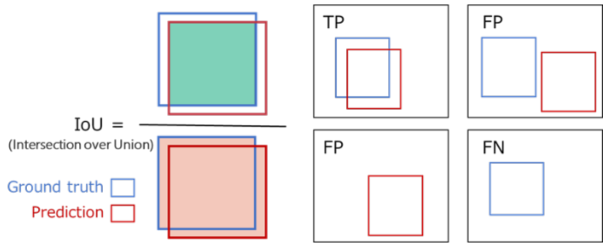


Figure 3. Image of IoU and the relationship between ground truth area and prediction area

4.2 Verification of Construction Machinery Detection Accuracy by YOLOv3

The accuracy of fine tuning was verified using mAP. The test data set created. Which is described in Section 3.3 was used for verification (Table. 2).

The mAP was 92.20%. For the five classes of the backhoe, bulldozer, road roller, dump truck, and car, high accuracy of over 90% was obtained, but for concrete mixer truck, the accuracy was relatively low at 71.43%. The road roller with the same number of objects in the data set has a detection rate of 100%. Therefore, the cause of the decrease in accuracy is considered to be due to the characteristics of the image. Many images included the road roller has a clear hue difference from the background and are considered easy to detect. On the other hand, compared to the road roller image, many of the concrete mixer trucks are not clearly different in hue from the surroundings, which may have led to a decrease in the detection rate. Fig. 4 shows examples of test data sets including concrete mixer truck and road roller. Fig. 5 shows examples of construction machinery detection based on the obtained learning results

The test data set is images that are not used for learning of YOLOv3, but it can be confirmed that construction machinery has been detected. In addition, even if the construction machinery in the image is small because of UAV aerial images, they can be detected.

Table 2. YOLOv3 construction machinery detection accuracy

Object class	Number of objects	Average Precision (%)
backhoe	187	98.40
bulldozer	16	92.73
road roller	22	100.0
dump truck	180	97.12
mixer truck	21	71.43
car	299	93.51
mAP		92.20

(a) Road roller

(b) Concrete mixer truck

Figure 4. Examples of test datasets

Figure 5. Example of construction machinery detection by YOLOv3

5 Conclusion

In order to perform photogrammetry with UAV, it is necessary to stop construction work or to remove noise manually after creating a 3D model. We proposed a system for efficient UAV photogrammetry by detecting construction machinery by deep learning method and removing them that causes noise.

For the detection of construction machinery from UAV aerial images, YOLOv3, which is an object detection method based on convolutional neural networks, is used. 2182 training data sets are created for the construction machinery to be removed and performed detection and accuracy evaluation using 219 test data sets. As a result, it was found that the detection of construction machinery with a high accuracy of 92.2% compared with previous studies. Since the construction machines shown in the UAV aerial images handled in this study are often small, the usefulness of using YOLOv3, which is good at detecting small scales, were confirmed. On the other hand, the detection rate of mixer truck was relatively low among the classes we tried to detect in this paper. This is thought to be due to the characteristics of the images used for training. Increasing the detection rate is expected by adding various mixer track images to the training data set.

As future work, the construction machinery part detected by YOLOv3 should be masked and complemented, and a 3D model should be generated by SfM, aiming at completion of a system that enables high-precision UAV photogrammetry even under construction work. We will examine an image completion method suitable for this study and verify the accuracy of noise removal after generating the 3D model. Further, the robustness of the system against environmental changes of sites should be evaluated.

References

[1] Omari S. E. and Moselhi O. Integrating 3D laser scanning and photogrammetry for progress measurement of construction work, *Automation in construction*, 18.1, 1-9, 2008.

[2] Gore S. Song L. and Eldin N. Photo-modeling for construction site space planning, In *Construction Research Congress 2012: Construction Challenges in a Flat World*, 1350-1359, 2012.

[3] Yamaguchi R. Yabuki N. and Fukuda T. Accuracy Improvement of 3D Models Generated by Structure from Motion (SfM) Using Joint Histogram of Oriented Gradients (Joint-HOG), *Internet Journal of Society for Social Management Systems*, Vol. 11, Issue 1 sms17-4917, 2017.

[4] Liu W. Anguelov D. Erhan D. Szegedy C. Reed S. Fu C. Y. and Berg A. C. Ssd: Single shot multibox detector, In *European conference on computer vision*, Springer, Cham, 21-37, 2016.

[5] Redmon J. Divvala S. Girshick R. and Farhadi A. You only look once: Unified, real-time object detection, In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779-788, 2016.

[6] Yabuki N. Nishimura N. and Fukuda T. Automatic Object Detection from Digital Images by Deep Learning with Transfer Learning, In *Workshop of the European Group for Intelligent Computing in Engineering*, Springer, Cham, 3-15, 2018.

[7] Kim H. Kim H. Hong Y. W. and Byun H. Detecting construction equipment using a region-based fully convolutional network and transfer learning, *Journal of Computing in Civil Engineering*, 32.2, 04017082, 2017.

[8] Redmon J. and Farhadi A. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.