Automatic Generation Method of Building Mask Images by Using the 3D Model with Aerial Photograph for Deep Learning

Toward authentic optimal placement system of outdoor loudspeaker for communication of disaster prevention information

○ Kazunosuke Ikeno[*1], Tomohiro Fukuda[*2] and Nobuyoshi Yabuki[*3]

*1 Div. of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University

*2 Assoc. Prof., Div. of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Ph.D.

*3 Prof., Div. of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Ph.D.

Keywords: Disaster prevention; Disaster administrative radio; Machine learning;
Automatic design; Semantic segmentation; Mask image.

1. Introduction

The importance of the administrative authorities communicates information about disaster prevention is increased. When a natural disaster occurs in Japan, administrative authorities communicate this information using Common Antenna TeleVision (CATV), emergency alert email and disaster administrative radio[1)].

CATV and emergency alert email can convey the optical information clearly, however only many people can get information. On the other hand, disaster administrative radio system has been developed in roughly 82% of the municipalities in Japan[2)]. There are two ways, individual house receivers and outdoor loudspeakers for transmission methods. Ministry of Internal Affairs and Communications recommends distribution individual house receivers to all family because the transmission method using individual house receivers hardly receive influence from the weather, however, many municipalities use not individual house receivers but outdoor loudspeakers due to financial matters and limited cost-effectiveness. However, there is a problem that some regions don't receive the information from outdoor loudspeakers[3)]. Municipalities need to reconsider the placement of outdoor loudspeakers.

Municipalities request experts to consider the optimal placement of outdoor loudspeakers when they prepare the update plan. The experts make an investigation into the land use from Geographic Information System (GIS) and map owned by each municipality. However there are some cases that the information provided by the local government don't correspond to present state in provincial cities with a large population size and regions that land use changes frequently. The information on local cities in GIS is updated only occasionally and it is expensive. In some cases, there is no information about the building placement in GIS. It is difficult to confirm the present state. There is a method to confirm land use from aerial photographs. This method enables to get the current information easily and inexpensively.

Recently, the method of object detection by using machine learning is proposed. We can detect the target objects on images quickly using this method. It is also possible to detect buildings on aerial photographs using this method[4)]. The accuracy of building detection is greatly influenced by the quantity and features of the dataset. It is necessary to train the model adequately for each target area. However, it takes much time to generate mask images from aerial photographs for training models because there are many buildings in one aerial photograph and many sets of aerial photographs and mask images are needed to train the model.

The method to generate mask images automatically by using Virtual Reality (VR) models for machine learning is proposed[5)]. We can generate mask images easily by using 3D virtual models that are divided into each object such as buildings, roads, and terrain. However, the representation of the VR models is sometimes deficient to be used as datasets for training the model. In this research, by pasting aerial photographs on 3D models as materials , representation of the VR models expects to be improved. We can generate mask images automatically by using these VR models and prepare proper datasets.

This study aims to develop the automatic optimal placement system of outdoor loudspeakers by analyzing land use from aerial photographs. This article describes the method to detect buildings on aerial photographs by the model trained by using the existing dataset and generate mask images automatically from VR models for training models.

2. System development

2.1. PROPOSED SYSTEM

Our proposed system detects buildings on aerial photographs using semantic segmentation, understand land usage pattern and provide optimal placement by using geometric analysis. Semantic segmentation is one of the object detection methods using machine learning. Conceptual diagram of our proposed system is presented in Figure 1.

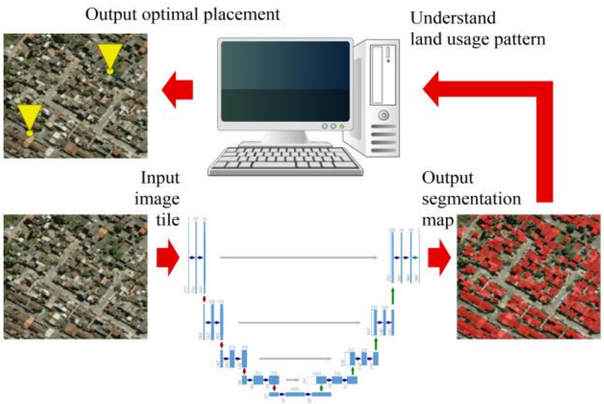


Figure 1. Conceptual diagram of our proposed system

The system for generating mask images to train the model is developed. This system loads 3D models include terrain and building objects, classify the building class and others class, switch between the model include all objects and the model include only building and generate two upper view images of the models. This system can generate plural sets of mask images and aerial photographs from one 3D model. Conceptual diagram and flowchart of generating the dataset is presented in Figure 2 and Figure 3.

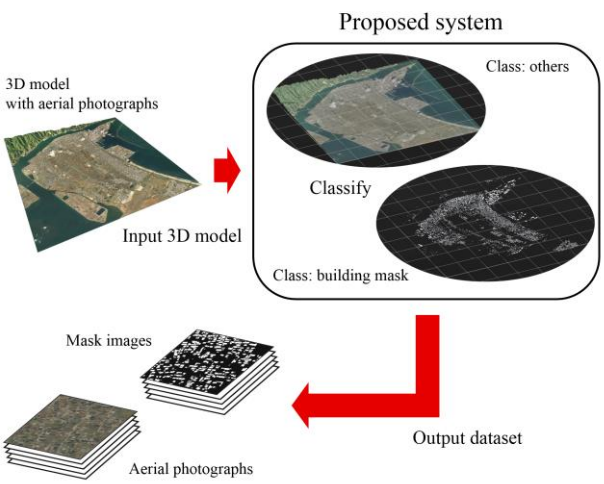


Figure 2. Conceptual diagram of generating mask images

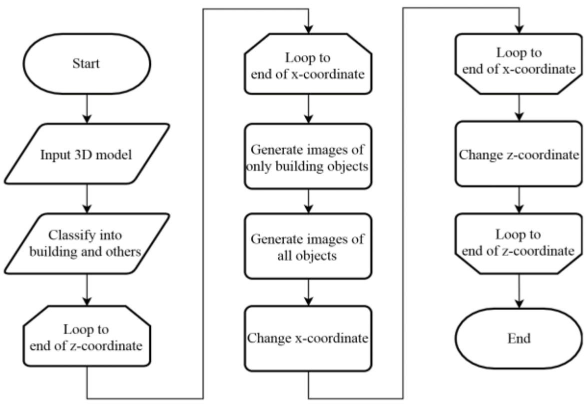


Figure 3. Flowchart of generating mask images

2.2. SYSTEM IMPLEMENTATION

2.2.1. DEVELOPMENT ENVIRONMENT

The libraries and software for detecting buildings on aerial photographs using machine learning are shown in A) and B). The OS of the PC is Ubuntu 16.04 LTS.

A) TensorboardX 1.6

TensorboardX can show loss output during training in real-time using Application Programming Interface (API). It was used to confirm the training state of the model.

B) Jupyter Notebook

Jupyter Notebook is one of the analytical tools. It can manage programs, explanations and execution results at once. It was used to output segmentation results.

The libraries and software for generating sets of mask images and aerial photographs automatically are shown in C) and D). The OS of the PC is Windows 10 Education 64bit.

C) Autodesk InfraWorks 2019

Autodesk InfraWorks is one of the infrastructure design software. It was used to create 3D models of the target area.

D) Unity 2017.3.1f1 (64-bit)

Unity is one of the game engines that can load 3D models. It was used to build systems.

2.2.2. PRELIMINARY EXPERIMENT

As preliminary experiment, we train the model by using existing dataset. Our proposed system used U-Net[6)] because it is a type of methods to detect objects in pixel units for segmentation of buildings on aerial photographs. The dataset for training was SpaceNet[7)] include satellite images and ground-truth building footprints in Rio de Janeiro, Vegas, Paris, Shanghai, Khartoum, and Atlanta. Our proposed system used the dataset of Rio de Janeiro because they have most building labels (polygons) in the six datasets. The satellite images of SpaceNet are too large (19584 px×19584 px) to be used for training the model. Then we divided the original images into 6940 small images (438 px×406 px) and classified them into three classes. The model was trained on U-Net by using the images of ground-truth building and satellite images. After that, we evaluated the accuracy of the model. For verification, Intersection over Union (IoU)[8] which is one of the metrics that evaluate how similar predicted area is to the ground truth area was used. IoU is given as Eq. (1).

$$IoU = TP / (TP + FN + FP) \quad (1)$$

TP (True Positive), FN (False Negative), FP (False Positive)

2.2.3. GENERATING MASK IMAGES AUTOMATICALLY

The 3D model of target area was created. by using Autodesk InfraWorks. The building placement were determined according to Fundamental Geospatial Data[9] (FGD) provided the Geospatial Information Authority of Japan (GSI). The aerial photographs are pasted on objects of terrain.

The objects of the created model were classified into building objects and others on Unity. We defined the camera to take images on the 3D model. The x-coordinates and z-coordinates of the camera were changed by 100 in 5600×5200 (world coordinates of Unity). The camera outputted images include only aerial photographs and images only building masks as png files (400 px×400 px). Conceptual diagram of the system to generate sets of aerial photographs and mask images automatically is presented in Figure 4.

3. Results

The results of detecting buildings on aerial photographs in the test class by the trained model are presented in Figure 5. The red area is the predicted and true area of buildings. IoU that calculated using sets of aerial photographs and mask images in verification class is shown in Table.1. The results of detecting buildings on aerial photographs in Sakaiminato by the trained model are presented in Figure 6.

Table 1. IoU of the trained model

IoU	0.602

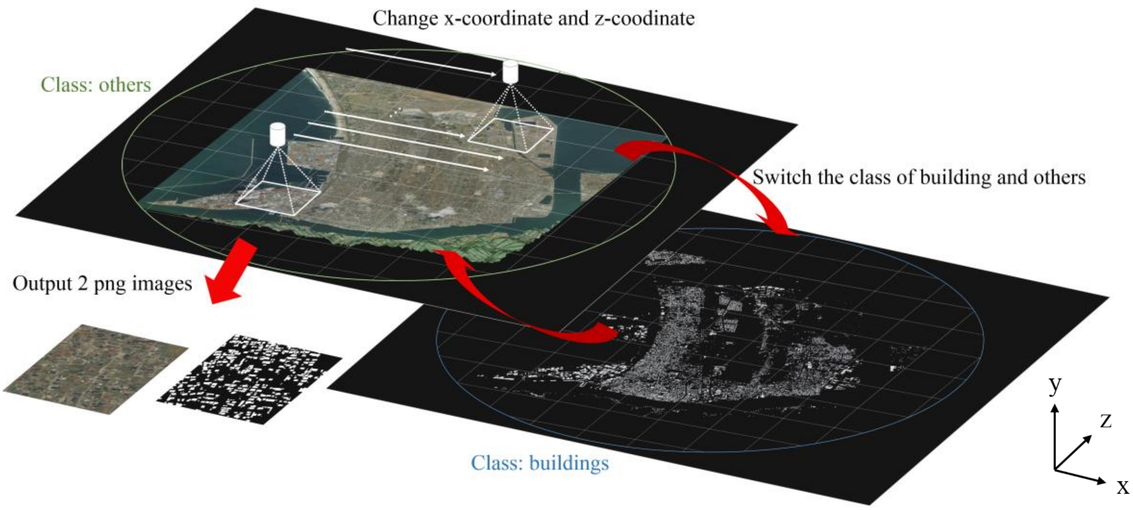
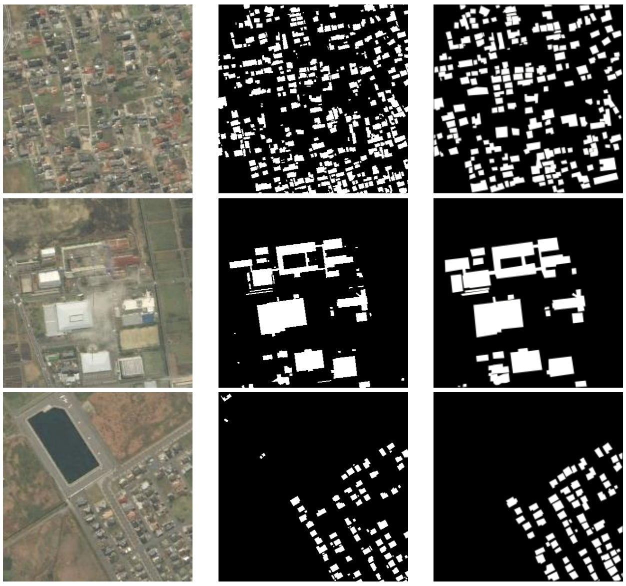


Figure 4. System to generate sets of aerial photographs and mask images

Figure 5. Results of detecting buildings on test photographs

Figure 6. Results of detecting buildings on Sakaiminato photographs

The sets of aerial photographs and mask images that generated automatically by our proposed system are presented in Figure 7. Our proposed system generated 2912 sets in 190 seconds.



Photographs Mask images Mask images

Automatically generated Manually generated

Figure 7. Generated aerial photographs and mask images

4. Discussion

Our proposed system detected buildings on aerial photographs in the existing dataset with the accuracy that IoU is 0.602. IoU of this system exceeds the threshold value (IoU = 0.5)[10]. This system can detect buildings sufficiently. However, there are some misdetected and undetected buildings in the results of detecting buildings on target aerial photographs. There are two reasons for these problems. First, the color tone of aerial photographs in dataset used to train the model significantly different to target aerial photographs. Second, there are few vacant lots and farmlands in the aerial photographs in the dataset of Rio de Janeiro. It is necessary to consider these things for the building segmentation on local cities that have many vacant lots and farmland.

The time to generate mask images is reduced by generating them from 3D models automatically. The mask images generated by our proposed system are almost the same as the mask images generated manually. This system can generate mask images in detail shapes. However, it cannot generate mask images of small warehouses. Some aerial photographs that include thin lines on boundaries between each object of the 3D model. It is necessary to prescreen generated mask images. This system can be used for other detected targets that location information is provided.

5. Conclusion

The conclusions of the present study are shown below.

- Proposed the system to generate sets of aerial photographs and mask images from 3D models.
- Compared mask images generated automatically and manually.

The future works is to adjust the color tone of target aerial photographs to the photographs in the dataset. In the next step, we train the model using the mask images generated by our proposed system and evaluate the accuracy of the trained model.

References

1） Fire and Disaster Management Agency: 2017, Chiho bosai gyoseino genkyo [Present situation of the local disaster prevention administration] <http://www.fdma.go.jp/disaster/chihoubousai/pdf/27/genkyo.pdf> (accessed 24 January 2019).

2） Ministry of Internal Affairs and Communications: 2019, Shichoson bosai gyosei musen tou seibi jokyo [States of development of disaster administrative radio system] <https://www.tele.soumu.go.jp/j/adm/system/trunk/disaster/change/index.htm?print> (accessed 3 February 2019).

3） Ministry of Internal Affairs and Communications: 2017, Saigai joho dentatsu syudan no saikin no doko [Recent trend of the disaster communication means] <http://www.soumu.go.jp/main_content/000525713.pdf> (accessed 9 October 2018).

4） Adachi, K. and Iwata, T.: 2019, A Study on Recognition of Buildings in Aerial Photograph Based on Machine Learning, Summaries of technical papers of Annual Meeting Architectural Institute of Japan, Information systems technology, 57-58.

5） Fukuda, T., Novak, M. and Fujii, H.: 2019, Development of Segmentation-Rendering on Virtual Reality for Training Deep-learning, Simulating Landscapes and Advanced User Experience, Proceeding of the 37th eCAADe and 23th SIGraDi Conference, Volume 2, 433‐440.

6） Ronneberger, O., Fischer, P. and Brox, T.: 2015, “Unet: Convolutional networks for biomedical image segmentation”. In International Conference on Medical Image Computing and Computer-Assisted Intervention, 234‐241. Springer.

7） DigitaalGlobe: 2019, SpaceNet <https://explore.digitalglobe.com/SpaceNet-Thank-You.html> (accessed 9 September 2019).

8） Everingham, M., Ali Eslami, S. M., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A.: 2015, The Pascal Visual Object Classes Challenge: A Retrospective, International Journal of Computer Vision, 111(1), 98‐136.

9） Geospatial Information Authority of Japan: 2019, Kibanchizu johosaito [Site of basic map information] <https://www.gsi.go.jp/kiban/> (accessed 10 September 2019).

10） Jabbar, A., Farrawell, L., Fountain, J. and Chalup, S. K.: 2017, Training Deep Neural Networks for Detecting Drinking Glasses Using Synthetic Images, Neural Information Processing, Part 2, 354‐363.