A REAL-TIME VISUAL ENVIRONMENTAL ESTIMATION SYSTEM USING IMAGE SEGMENTATION

A prototype for green view index

○Rui Cao[*1] Tomohiro Fukuda[*2] Nobuyoshi Yabuki[*3]

Keywords: Landscape simulation; visual environment; green view index; deep learning; semantic segmentation

1. Introduction

To reduce carbon footprint and urban heat island, and to improve the quality of urban landscape, greening has been promoted to improve architectural and urban environments.

Green view index (GVI) means the proportion of green in the perspective of people's vision, since 90% of the environment information received by people comes from vision, GVI is a measurable physical quantity that can be used as an effective green landscape estimation metric (Aoki, 1987). GVI provided a more intuitive metric which helps citizens understanding the vegetation scenarios of the city than the ratio of green space, green coverage rate or green area per capita, the most common urban greening estimation criteria (Yang et al., 2009).

With urbanization and development of cities, urban green space construction which helps creating a more comfortable and sustainable urban residential building environment to citizens has been paid more and more attention on worldwide, therefore, GVI started to be recognized to estimate the visual impact of city planning and management practices in urban construction and city planning area. For example, the Japanese local government set green view index as one of the landscape criteria for new construction and extension or reconstruction of buildings (Nishinomiya City, 2018). Therefore, efficient, low cost, high precision, easy to operate, rapid building/street-wide GVI estimation system is necessary.

To estimate the GVI, one of the conventional estimate techniques which has been put into practical use is by analyzing images taken in specific viewpoints (hereinafter "practical estimation technique") (Osaka Prefecture, 2013), it means after taking photographs in representative viewpoints, the green areas in the photographs are masked to estimate GVI by using an image processing software. This method is easy to understand and operate with an intuitive GVI estimation result in high precision. However, using the practical estimation technique in the estimation of GVI, a worker makes a masking of greening areas manually. Therefore, it takes a lot of preprocess time and high labor costs in actual production and construction. In addition, it is difficult to deal with dynamic viewpoints such as gathering continuous data on visible greenery while walking or driving.

In contrast, several automatic estimation methods of GVI has been reported. Li et al. (2015) proposed a method to estimate the GVI by analyzing landscape images acquired by Google street view (GSV). However, GSV images are captured at a static point in time, in other words, this technique is also not practical for dynamic viewpoints. Ding et al. (2015) developed an automatic GVI estimation system based on image processing technology. By using this method, Erroneous extraction of trees reflecting on the window is reduced to the utmost limit and accuracy is high at about 93.8%. However, it is difficult to process in real-time because of large processing load. Based on the above system (Ding et al., 2015), a new automatic GVI estimation system with the real-time estimation function has been developed (Inoue et al., 2018). With the aim of maintaining a certain level of accuracy while giving top priority to real-time performance, the method of image reduction processing and logical conjunctions of multiple images was used to decrease processing load. This system has high estimation accuracy at about 95.7%. However, there are still some problems, low brightness regions such as tree branches and trunk cannot be extracted correctly, the edge part of the tree is not extracted, green artifacts are extracted. In addition, it is needed to adjust the parameters manually in advance, therefore, the module in this system for estimating the green view index cannot be easily generalized.

On the other hand, in the field of image processing, semantic segmentation describes the process of associating each pixel of an image with a class label. In recent years, image semantic segmentation systems based on deep learning have attracted considerable research attention focusing on autonomous driving, industrial inspection, classification of terrain visible in satellite imagery and medical imaging analysis (Long et al.,2015). It can also be used in other area such as GVI estimation.

This study presents a new GVI estimation system by training and constructing a deep learning network for semantic segmentation of natural green pixels in landscape images.

As the result of this research, an automatic, real-time, dynamic-viewpoints-apply GVI estimation system is developed which has the similar level of estimation accuracy as the previous researches. Furthermore, this system can not only estimate the GVI of a landscape image but also the GVI of a landscape live video, there are no detailed preprocess and parameter to adjust, therefore it can be generalized easily.

2. Proposed Method

Deep learning is part of a broader family of machine learning methods based on learning data representations. It creates a more abstract high-level representation of attribute categories or features by combining low-level features to discover distributed features of data.

Semantic segmentation which is essential for image analysis tasks describes the process of associating each pixel of an image with a class label such as road, person and trees.

In recent years, significant progress of deep learning network has been made greatly to improve the accuracy of image semantic segmentation. In this research, we will introduce a new estimation system that can segment the green view from an image or a video for estimating the GVI by constructing and training a deep learning network.

This proposed system which consists of 4 steps is described in the followings as in Figure 1.

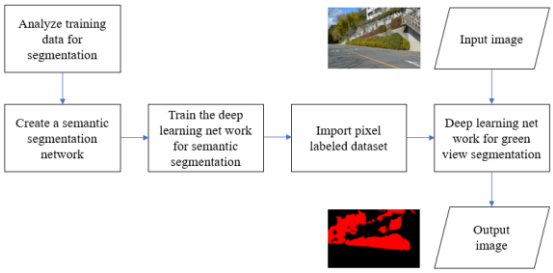

Figure 1. Proposed system flowchart.

2.1. ANALYZE TRAINING DATA FOR SEGMENTATION

To train a semantic segmentation network, a collection of images and its corresponding collection of pixels labeled images is necessary.

A pixel labeled image is an image where every pixel value represents the categorical label of that pixel. In this research, we prepared 300 original images including natural green and buildings then use label software to segment the images manually into two parts: green area and the others to get the label mask images. In addition, the original images are in three-channel and the label mask images are 8-bit grayscale in single-channel. At last of this step, both 300 original images and their 300 label mask images were prepared.

2.2. CREATE A SEMANTIC SEGMENTATION NETWORK

Various frameworks were developed to make deep learning easier to use such as Segnet (Badrinarayanan et al., 2017). Segnet is an autopilot system model network which can classify each pixel of an urban street image into one of the twelve classes (building, sky, pole, road marking, road, pavement, tree, sign symbol, fence, vehicle, pedestrian and bike) by inputting real-time images or video. We can also use this to train customized dataset to segment the elements that we want such as GVI.

Segnet network is very efficient and accurate especially applied to segment the street scene, we chose it as the base framework to train our own data for segment the green view area of images and videos to estimate GVI because GVI is also mostly estimated in the street scene.

2.3. TRAIN THE DEEP LEARNING NETWORK FOR SEMANTIC SEGMENTATION

In this step, create an original image database for loading the original images and a label database for loading label mask images. Then, we start training.

2.4. IMPORT PIXEL LABELED DATASET

In this step, after training 300 pairs of images for GVI, a pixel labeled dataset for semantic segmentation network was created installed in a text file. This file obtains the result of our training and the class-weight used for segmenting the green view of images or videos.

Up to this point our network construction and training is completed, we can use the new dataset to segment the images or videos for GVI estimation.

3. Verification Experiment

To verify the performance of the real-time estimation system for GVI developed in Chapter 2, experiments were performed as follows.

First, record landscape video including natural green and buildings uploading to the server. Then, the server starts processing that segments the transferred video into green area part and the others with the estimation result of GVI. Here are some details about the experiment:

- Date: 1st October 2018.
- Weather: Sunny.
- Location: Osaka University, Suita campus.
- Camera equipment: iPhone7.
- Video parameter: 720p HD, 30fps.
- Video length: 15 seconds.
- Amount of Video: 4.

- Recording method：Fix the phone vertical on the bike and push the bike to record slowly.

In this case, we captured 4 frames as the original images from each video and their outputted images by the proposed method to explain, the verification method is as follows.

1) Create correct images for GVI by processing the original images. Use Photoshop to mask every green view area then painted them into green color, the other area was painted into black color, at last, calculate the ratio of pixels of green region, it means the accuracy GVI.
2) Calculate the ratio of pixels of red region in the images that outputted by proposed method, it means the GVI that estimated by proposed method.
3) Combine two photos into one as follows in Figure 2.

The region where the green color of the ideal result image overlaps with the red color of the segmented image (correct region) is yellow. In the composit images, the definition of different color regions is as follows in Table 1.

Table 1. Definition of each color in composite image.

	Definition
Yellow	Accurate extracted pixel.
Black	Accurate unextracted pixel.
Red	Over-extracted pixel.
Green	Unextracted area pixel

Calculate the number of pixels in each color region separately We use Y to represent the number of pixels in the yellow region, B for black, R for red, G for green, and X for the total pixels of the images.

Then, calculate the rate of pixels of each part in the combined images to perform verification analysis, the calculation method is as follows in Table 2.

Table 2. The calculation formula of accuracy rate and inaccuracy rate.

	Calculation formula
Extract accuracy rate [%]	$\frac{Y}{X} \times 100$
Unextracted accuracy rate [%]	$\frac{B}{X} \times 100$
Accuracy rate [%]	$\left(\frac{Y}{X} + \frac{B}{X}\right) \times 100$
Over extracted rate [%]	$\frac{R}{X} \times 100$
Unextracted inaccuracy rate [%]	$\frac{G}{X} \times 100$
Inaccuracy rate [%]	$\left(\frac{R}{X} + \frac{G}{X}\right) \times 100$

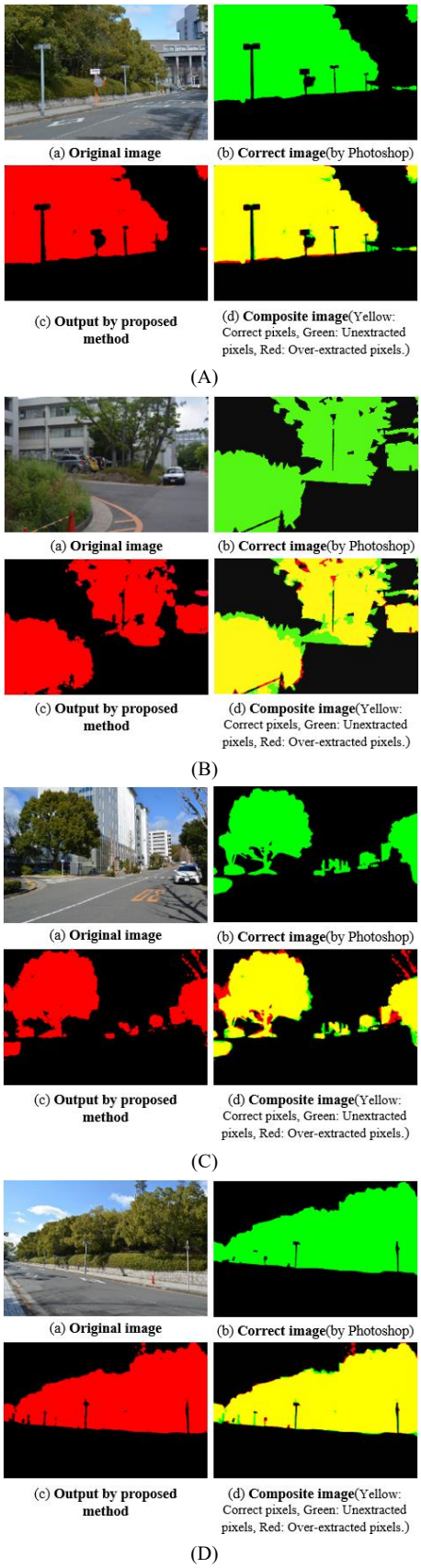

Figure 2. Comparison between correct image and the image that outputted by proposed method.

According to the Table 2, we can get the average of the accuracy rate and inaccuracy rate of these 4 images as follows in Table 3.

Table 3. The accuracy rate and inaccuracy rate of GVI estimation.

Extract accuracy rate [%]	39.1
Unextracted accuracy rate [%]	57.2
Accuracy rate [%]	96.3
Over extracted rate [%]	1.1
Unextracted inaccuracy rate [%]	2.6
Inaccuracy rate [%]	3.7

In this verification experiment, the accuracy rate of proposed system is about 96.3% which achieved the similar level the previous researches such as 93.8% (Ding et al., 2015) and 95.7% (Inoue et al., 2018). In addition, although not completely accurate, most of the tree trunks and the edge part of the tree are correctly extracted through visual observation.

4. Conclusion and Future Work

We developed a real-time visual environmental estimation system for GVI using image segmentation based on deep learning network. After constructing and training our own network, all we need to do is to input an image or video and the system will automatically segment them into green area and the others with the result of estimation for GVI, the accuracy rate of this system is about 96.3% and the inaccuracy rate is about 3.7%, we can also use a smartphone or web camera to take photographs or record live videos, remotely transferring the data to the sever and automatically estimate the GVI about the scene in front of your eyes.

Because of the use of deep learning techniques, the versatility of this system is very high, not only limited to the estimation of GVI, but also applied to other environment elements such as the sky factor ratio if we perform systematic deep learning training for it.

However, there are still some problems to perfect as follows.

- Some details such as tree branches, leaves and low brightness regions cannot be detected and segmented correctly, this is related to the accuracy of the deep learning network itself that we adapted and the number of images for training.
- There is a delay of about tens of seconds between transferring and segmenting because until this moment, we have not completed the server setup problem to make them implement at the same time, this will be the subject from now on to make56 data transferring and segmenting running simultaneously.

References

1) Aoki, Y.: 1987, Relationship between perceived greenery and width of visual fields, J. Jpn. Inst. of Landscape Architects, 51(1), 1-10.
2) Badrinarayanan, V., Kendall, A. and Cipolla, R.: 2017, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495.
3) Ding, Y., Fukuda, T., Yabuki, N. and Michikawa, T.: 2015, A Measurement Tool for Visible Greenery Ratio Derived from Gaussian Blur, Hue and Saturation Filtering, Proceedings of the Second International Conference on Civil and Building Engineering Informatics (ICCBEI2015), CD-ROM.
4) Inoue, K., Fukuda, T., Cao, R., Yabuki, N:2018, Tracking Robustness and Green View Index Estimation of Augmented and Diminished Reality for Environmental Design - PhotoAR+DR2017 project, In Proceedings of the 21st International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA 2018) pp. 339-348.
5) Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., Zhang, W.:2015, Assessing street-level urban greenery using Google Street View and a modified green view index, Urban Forestry & Urban Greening, 14, pp. 675-685.
6) Long, J., Shelhamer, E., Darrell, T.:2015, Fully Convolutional Networks for Semantic Segmentation, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
7) Nishinomiya City:2018, “Regulation of the environmental preservation and creation: planting of buildings and their sites”. Available from <https://www.nishi.or.jp/kotsu/kankyo/hanatomidori/seido/kenchikubutsu-ryokka.files/toriatsukai.pdf >
8) Osaka Prefecture :2013， “A Survey Guideline for Visible Greenery Ratio”. Available from <http://www.pref.osaka.lg.jp/attach/17426/00000000/guideline.pdf>（Accessed online 7th Oct 2015）.
9) Yang, J., Zhao, L., McBride, J., Gong, P.: 2009, Can you see green? Assessing the visibility of urban forests in cities, Landsc. Urban Plan., 91, pp. 97-104.

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