Urban vegetation has been used to tackle architectural and urban problems by reducing urban heat islands and improving the quality of urban landscapes and biodiversity. The green view index provides end users with a metric to intuitively understand the vegetation scenarios. This study integrates a green view index estimation method and augmented reality (AR) and diminished reality (DR) scenes of future architectural and urban design simulations. We developed the AR/DR system "PhotoAR+DR2016 (photogrammetry-based augmented and diminished reality)" that simultaneously measures the green view index and simulates building, urban, and planting designs with addition, demolition, and removal of the objects such as structures. The developed system enables real-time measurement of the green view index by appropriately reducing the image size and extracting the green area. Using the developed prototype system, the on-site verification can be conducted; in addition, the processing speed and the accuracy and inaccuracy rates can be measured, and the green view index can be sufficiently measured in real time.

Keywords: Green View Index, Landscape assessment, Design support system, Diminished Reality, Augmented Reality, Image analysis

INTRODUCTION
Urban vegetation has been used to tackle architectural and urban problems by reducing urban heat islands and improving the quality of urban landscapes and biodiversity. The quantification of urban vegetation using metrics such as the green coverage ratio and green view index is important for stakeholders in order to motivate increases in vegetation and to simulate the amount of vegetation required in a design process (Shafer and Brush 1977, Aoki 1987). Specifically, due to technological developments and the diffusion of green walls, the green view index has received increasing attention (Almazan et al. 2012). Local governments exist that set green view index as one of the landscape criteria for new construction and extension or reconstruction of buildings [1].

PhotoAR+DR2016
Integrating Automatic Estimation of Green View Index and Augmented and Diminished Reality for Architectural Design Simulation

Tomohiro Fukuda¹, Kazuya Inoue², Nobuyoshi Yabuki³
¹,²,³Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University
¹,³{fukuda|yabuki}@see.eng.osaka-u.ac.jp ²inoue@it.see.eng.osaka-u.ac.jp
The green view index provides end-users with a metric with which to intuitively understand vegetation scenarios (Yang et al. 2009). One of the conventional measurement techniques is to determine the vegetation percentage per viewpoint using image processing software applications. In this technique, after taking photographs from representative viewpoints, the natural green areas in the images are manually selected. However, this technique is time consuming and is subject to large variations in the amount of time required to process different images. Therefore, researches on automatic measurement of the green view index has been reported. Komiya and Susaki (2015) proposed a methodology to estimate a green view index in urban areas using airborne LiDAR and aerial photographs. This method efficiently estimates a wide range of green view index; however, it is difficult to measure hedges and wall greening with low depth. A method to measure the green view index by analyzing landscape images acquired by Google street view (GSV) was proposed (Li et al. 2015, [2]). However, GSV images are captured at a certain point in time; therefore, it is difficult to obtain the green view index for other times. In addition, this technique is not practical for dynamic viewpoints, such as gathering continuous data on visible greenery while walking or driving. Ding et al. (2016) proposed an automatic estimation system that can exclude certain unwanted objects (such as the reflection of trees in windows) using Gaussian blur, mean shift, hue, and saturation filtering functions based on image processing technology. However, real-time processing was impossible.

This research tackles a method for real-time estimating the green view index using image processing. This research also integrates a green view index estimation method and augmented reality (AR) and diminished reality (DR) scenes of future architectural and urban design simulations. AR integrates the 3-D virtual objects of design proposals into an existing 3-D environment in real time (Milgram and Kishino 1994). Specifically, AR can help visualize full-scale design projects on a planned construction site (Gudrun et al. 2001). However, the use of only existing AR approaches cannot correctly simulate the view after the demolition and removal of the structures. If new structures are simulated while an old existing structure is still present, a 3D virtual model of the new structure will overlap the existing to-be-renewed structure. As a result, a portion, if not all, of it will still be visible and displayed. To solve this problem, DR removes the image of an existing object from a scene (Enomoto and Saito 2007, Inoue et al. 2016). In this way, possible future changes in vegetation can be visualized in an AR/DR environment by adding planting design models.

AUTOMATIC ESTIMATION OF GREEN VIEW INDEX USING IMAGE PROCESSING

Our proposed method aims to realize real-time processing for the automatic estimation of the green view index by extracting natural green pixels in landscape images (Figure 1). In this study, an algorithm with six steps is developed and implemented using OpenCV (ver. 3.0.0) on Microsoft Visual Studio 2013 (C/C++): (1) input target image; (2) reduce the image...
size using bilinear interpolation; (3) apply Gaussian blur, mean shift (Cheng 1995), hue (40, 180), and saturation filtering (0.2, 1.0) (Ding et al. 2015); (4) calculate the number of pixels extracted by filtering; (5) calculate the green view index; and (6) restore the image size using bilinear interpolation.

Important parameters that affect the accuracy rate and calculation time of the developed system are the reduction ratio of the image size (step 2), the Gaussian kernel size in Gaussian blur (step 3), and the spatial window radius and color window radius in the mean-shift filter (step 3).

In order to identify the optimal combination of these parameters, a parameter study was carried out as follows. First, a combination of arbitrary values for each parameter was studied for a representative image. Next, based on the results, the search ranges of each parameter value were further narrowed down and the optimum combination of the parameters was identified using a full search. The criteria for the optimum combination are an accuracy rate of more than 85%, a calculation time of less than 0.0667 seconds (15 fps), and a minimum inaccuracy rate.

In the parameter study, Figure 2 shows an example of the input image (1200 × 800 pixels) and the output image. With reference to Figure 2, both the accuracy and inaccuracy rates that are the setting conditions of the optimum solution search are explained. First, based on the original image (Figure 2 (a)), a correct image (Figure 2 (b)) was created via a conventional manual method using Adobe Photoshop. Second, an automatic measurement image (Figure 2 (c)) generated by combining certain parameter values and the correct image are synthesized (Figure 2 (d)). Third, pixels extracted in both the correct image and the automatic measurement image are regarded as correct pixels. Incorrect pixels obtained by combining pixels extracted only in the correct image (hereinafter referred to as unextracted pixels) and pixels extracted only in the automatic measurement image (hereinafter referred to as over-extracted pixels) are calculated. Finally, the accuracy and inaccuracy rates are calculated by dividing the correct and incorrect pixels by the correct pixels, respectively. In Figure 2 (d), the yellow pixels, the green pixels, and the red pixels correspond to the correct pixels, the unextracted pixels, and the over-extracted pixels, respectively.
Therefore, as shown in Table 1, the search range in the full search was set. As a result, the optimum combination was obtained as (reduction ratio, Gaussian kernel size, spatial window radius, color window radius) = (20, 5, 4, 20), respectively.

We applied the optimal combination to the developed automatic estimation system and measured the green view index of live video images. The image size reduction ratio was set such that the image width was 240 pixels. The experimental target was the change in green view index with respect to landscape changes during walking etc. As the important viewpoints, we selected the front of the Techno Alliance building and selected Keyakidori road to the entrance of US1 building in the Osaka University Suita Campus as a walking landscape. The measurement results are shown in Figures 3, 4 and 5 and in Tables 2 and 3.

As result of the experiment, the accuracy rate was 77 to 90% and the inaccuracy rate was 15 to 28%. The green view index by our proposed system was measured less than the green view index by manual operation using Adobe Photoshop. The reasons for these values were that (i) the extraction of thin components such as tree branches and leaves was incomplete and (ii) the shade of the crown could not be extracted due to low brightness. Compared with the results of Ding et al. (2016), the accuracy (85.6%-93.3%) and the inaccuracy rates (11.5%-21.4%) were almost similar. On the contrary, the proposed method has enabled real-time processing.
The proposed system creates DR scenes from video images acquired by a camera for a post-demolition landscape simulation. As explained in Chapter 1, DR is a technique for removing unwanted objects from a scene by overlaying an appropriate background image on the objects. DR uses computational techniques for tasks, such as estimation of the video camera’s position and orientation, computation of the background image, and recognition and tracking of the object. Our method estimates the video camera’s position and orientation based on the photogrammetry method. Based on the estimated position and orientation, our method computes the background image of the target area and recognizes and tracks the target. Our method has two main phases: pre-processing and the real-time processing.

In the pre-processing phase, several photographs of the target structure to be diminished

<table>
<thead>
<tr>
<th>Landscape from important viewpoints</th>
<th>(A) Accuracy rate [%]</th>
<th>(B) Unextracted rate [%]</th>
<th>(C) Over-extracted rate [%]</th>
<th>(D)=(B)+(C) Inaccuracy rate [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP 1</td>
<td>82.3</td>
<td>17.7</td>
<td>10.9</td>
<td>28.6</td>
</tr>
<tr>
<td>VP 2</td>
<td>80.3</td>
<td>19.7</td>
<td>9.0</td>
<td>28.7</td>
</tr>
<tr>
<td>VP 3</td>
<td>84.8</td>
<td>15.2</td>
<td>4.3</td>
<td>19.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Landscape during walking</th>
<th>(A) Accuracy rate [%]</th>
<th>(B) Unextracted rate [%]</th>
<th>(C) Over-extracted rate [%]</th>
<th>(D)=(B)+(C) Inaccuracy rate [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP 1</td>
<td>77.2</td>
<td>17.7</td>
<td>2.0</td>
<td>24.8</td>
</tr>
<tr>
<td>VP 2</td>
<td>90.1</td>
<td>30.2</td>
<td>5.4</td>
<td>15.3</td>
</tr>
<tr>
<td>VP 3</td>
<td>89.1</td>
<td>27.9</td>
<td>10.6</td>
<td>21.5</td>
</tr>
<tr>
<td>Ave.</td>
<td>84.0</td>
<td>16.0</td>
<td>7.0</td>
<td>23.1</td>
</tr>
</tbody>
</table>
and background structures that are hidden behind the target structure are necessary as input data. Photographs of the target structure are used to reconstruct point cloud data and estimate the position and orientation of the web camera. Point cloud data of the target structure are reconstructed using the SfM (structure from motion) method (Tomasi and Kanade 1992, Agarwal et al. 2009). From the point cloud data, mask polygons are made, which are used to determine the removal region. The green mesh of the mask polygon allows for the removal region to be clearly distinguished. Additionally, the local features of each photograph of the target structure are extracted and used for image matching. Moreover, point cloud data and the polygon model of background structures are reconstructed from their photographs using the SfM.

In the real-time processing phase, local features, the mask polygon, and the background structures polygon, which are calculated and created in the preprocessing phase, are used as input data. First, local features of a live video image are extracted. The extracted features are compared with the features of stored images, which are calculated in the preprocessing phase. The automatic image matching method finds the most similar image and the position and orientation of the camera for that image are chosen as the current position and orientation of the web camera. By using the estimated position and orientation of the web camera, the mask polygon and back-

Figure 7
Design simulation using PhotoAR+DR2016
ground structures polygon are rendered. The re-
dered mask polygon determines the removal area. The area of the rendered background structures poly-
gon is overlaid onto the live video image. As a result, the target structure on the live video image seems to be diminished. Then, 3D virtual models can be in-
serted using an AR function and the green view index can be measured in real time.

Figure 8
Arrangement of existing and newly constructed buildings

To verify the applicability of the developed Pho-
toAR+DR2016, a pseudo design project was con-
ducted in an outdoor environment. We used a standard spec laptop PC, Panasonic Let’s NOTE CF-
SX3TDLTC with Intel Core i7-4500U @ 2.40GHz of CPU, 4.0 GB RAM, and Microsoft Windows 7 Profes-
sional 64 bit. A Microsoft LifeCam Studio web cam-
era for Notebooks with a resolution of 1280 × 720 pixels was used. We designed to dismantle the two-
storey building of Welfare Hall on Poplar Avenue in Osaka University Suita Campus and plant vegetation around the new building. Then, the green view index of the existing building before the dismantlement and that of the new building and vegetation were compared. The arrangement of the current building and the new building is shown in Figure 7.

By applying AR/DR, it was confirmed that in land-
scape simulation, the existing structure is dismantled and removed, and that a newly designed structure including vegetation is possible. However, assum-
ing cloudiness, the sky area without a hidden back-
ground model is painted with (R, G, B) = (225, 225, 230). After adding a tree model, the visible greenery ratio is increased from 18.5% to 31.3% (Figure 8).

CONCLUSIONS
The contributions of this research are as follows:

• We developed the AR/DR system “Pho-
toAR+DR2016” which simultaneously mea-
sures the green view index and simulates building, urban, and planting designs with ad-
dition, demolition, and removal of the struc-
tures.

• The developed system enables real-time mea-
surement of the green view index by appro-
priately reducing the image size and extract-
ing the green area. Using the developed system, the on-site verification can be con-
ducted; in addition, the processing speed and the accuracy and inaccuracy rates can be mea-
sured and the green view index in real time can be sufficiently measured.

In terms of future work, the extraction accuracy of the green region that was not properly extracted, e.g., tree branches and leaves and low brightness regions such as the shade of the crown, should be improved. It is also necessary to study the algorithm to auto-
matically learn the optimum combination of many parameters such as AI (Artificial Intelligence). To pro-
vide a more user friendly development environment for users, who are not experts in computer science, it is necessary to port the development platform from Visual Studio onto a game engine such as Unity.

ACKNOWLEDGEMENT
This research has been partly supported by JSPS KAK-
ENHI Grant Number JP16K00707.

REFERENCES
72–79

Aoki, Y 1987, ‘Relationship between perceived greenery and width of visual fields’, J. Jpn. Inst. of Landscape Architects, 51(1), pp. 1-10


