LEVERAGING 3D MODEL IMAGERY TO AUTOMATICALLY ESTIMATE A NEW WINDOW VIEW INDEX

Law, Stephen[1]; Suel, Esra[2]; Stalder, Steven[3]; Takizawa, Atsushi[4]

[1] Department of Geography, University College London;
[2] Centre for Advance Spatial Analysis, University College London;
[3] Swiss Data Science Center, EPFL/ETH;
[4] Department of Living Environment Design, Osaka Metropolitan University

Introduction

Extensive research has been conducted in environmental psychology that points to the benefits of high-quality views which are mentally restorative and provide opportunities for outlook, refuge, and solace [1]. Despite these inclinations, existing window view indices are often time-consuming to collect and over-simplified in their construction. The limited research can be attributed to the lack of data and computational methods for analysing vistas from properties. To address this gap, this study developed a novel data collection and resampling procedure that leverages the newly accessible photo-realistic 3D modelled imagery from Google Maps, along with existing machine learning techniques, to improve housing price/rent prediction and to establish a novel window view index for automatically assessing the appeal of views at home.

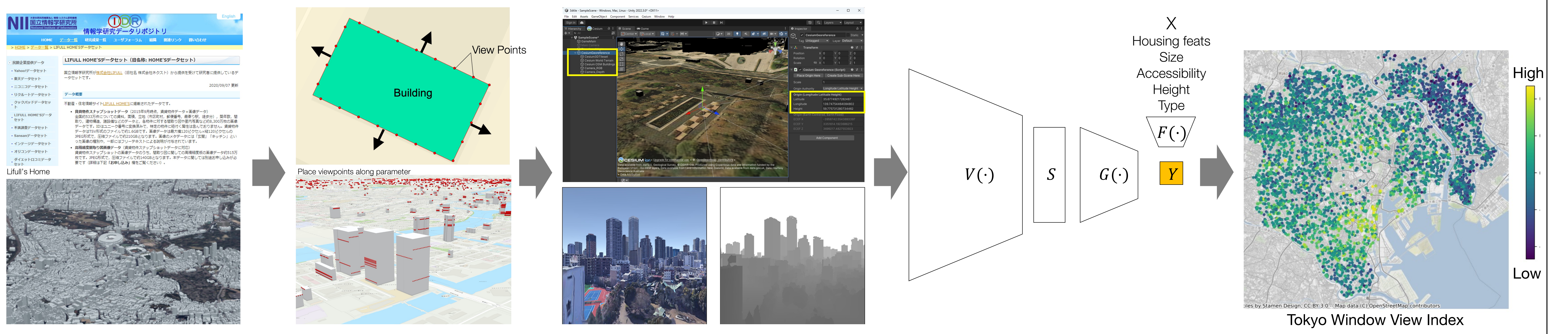

Fig1. Data Collection and Model pipeline. (left) Combine Lifull Property data and Plateau 3D data to retrieve high resolution coloured and depth window view image dataset from Google maps 3D API. (right) Leverage the 3D model imagery to predict Tokyo rental prices and extract a novel window view index automatically.

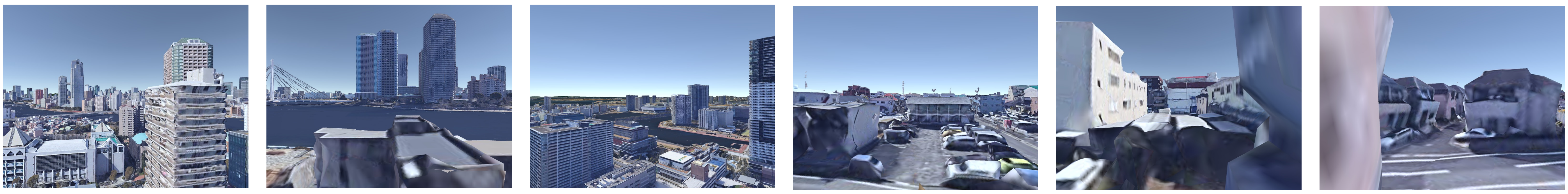

Fig2. sample of images with higher WVI offering more expansive views of skyline (left 3) and lower WVI with shallower views overlooking buildings (right 3).

© Google

Data Collection and Model Pipeline

The data collection and model pipeline can be seen in fig1. We first retrieve Tokyo's (23 wards) rental data from the Lifull property dataset between 07-2015 and 06-2017. Next, we merged the Lifull property dataset with the Plateau building dataset which contains 3D information on the floor and orientation of each flat. The merged dataset was then used as input for Google Maps photorealistic 3D tiles API via the Cesium plugin within Unity to generate high-resolution coloured and depth imagery based on the property window views. Following the model architecture introduced by [1], we then employ first a conventional hedonic regression model which learns a maping between housing characteristics X and Logged Tokyo rental prices Y. We then extract deep image features S from the modelled imagery using different pretrained/trained vision backbone models and train a second regression model that takes the deep image features to predict the difference $W = Y - \hat{Y}$ from the first stage to infer a predicted window-view index $\widehat{W}$. We minimise the mean squared error loss function using an ADAM optimiser with a learning rate of 0.01 for 5000 epochs.

References

[1] Law, S., Paige,B., Russell,C. (2019). Take a look around: Using street view and satellite images to estimate house prices. ACM TIST, 10(5).
[2] Kaplan, S. (1995). The restorative benefits of nature: Toward an integrative framework. Journal of environmental psychology, 15(3), 169-182.
[3] Dosovitskiy, A., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
[4] Stalder, S,et al. (2023). Self-supervised learning unveils change in urban housing from street-level images. *arXiv preprint arXiv:2309.11354*.
[5] Radford et al (2020). Learning transferable visual models from natural language supervision. In International conference on machine learning, pp.8748–8763. PMLR.
[6] Bastoni et al (2023). Satlaspretrain: A large-scale dataset for remote sensing image understanding. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 16772–16782, 2023

Results

We estimated a simple linear hedonic model as a baseline, incorporating standard housing attributes (R2=62.0% and rmse=0.202) and then nine other regression models building on top of the baseline one (refer to Table 1). The base + SegDepth model, which adds urban semantic classes and aggregate depth information from the window views, showed only minor improvements. The image features retrieved from large pretrained models namely ViT[3], BarlowSV[4], Clip[5] and Satlas[6] all performed better than these semantic and depth features. The best performance model came from the two Encoder-Decoder models trained on the 3D modelled imagery which boosted the out-of-sample R2 to 70.7% with an rmse of 0.178. These results show the use of semantic features and pretrained vision models on natural images are not as performant for predicting rents with synthetic 3D model data.

n	r2	rmse
Base	0.620	0.202
Base+SegDepth	0.650	0.194
ViT	0.651	0.194
ViT128	0.663	0.191
ViT256	0.663	0.190
Clip	0.672	0.189
BarlowSV	0.641	0.197
Satlas128	0.669	0.189
EncDecRGB	0.700	0.180
EncDecRGBD	**0.707**	**0.178**

Table 1. Regression results

Conclusions

In summary, this research introduces a novel data collection pipeline that uses newly accessible 3D-modelled imagery data from Google Maps and applied existing machine vision models to enhance the predictive accuracy of housing rent forecasts and in deriving a window view index WVI. This novel holistic index goes beyond the mere semantics of the window scene, providing insights into the quality of city views automatically. The results indicate that people generally prefer higher floors with expansive views overlooking the Tokyo Bay area, than lower floors with shallower views over-looking buildings. These findings underscore the need to improve the quality of window views for properties located on lower ground floors in Tokyo possibly through the incorporation of on-street greenery and art. Several limitations remain. Most importantly, the WVI needs to be validated and examined carefully through human subject surveys. Future research is also required to better understand the composition of the scenes. For example, would people prefer architecturally complex or simpler views? Despite these limitations, this research offers a new way to understand the city from the sky, demonstrating the usefulness of 3D-modelled imagery data and machine learning.

Data Sources

https://www.nii.ac.jp/dsc/idr/en/lifull/
https://www.geospatial.jp/ckan/dataset/plateau-tokyo23ku
https://developers.google.com/maps/documentation/tile/3d-tiles

Stephen.law@ucl.ac.uk

Acknowledgements
We thank the support of the Japan Society for the Promotion of Science (JSPS) and the Nikken Sekkei Research Institute which made this research and publication possible.