ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/383621216

EXPLORING THE LINKAGE BETWEEN SPACE SYNTAX AND VISUAL IMPRESSION OF ARCHITECTURAL SPACES: A case study of the Japanese streetscapes during moments of "Oku"

Conference Paper · June 2024

CITATIONS
0

READS
262

11 authors, including:

Yuval Kahlon
Institute of Science Tokyo
30 PUBLICATIONS **55** CITATIONS
SEE PROFILE

Zicheng Fan
National University of Singapore
10 PUBLICATIONS **67** CITATIONS
SEE PROFILE

Michal Gath Morad
University of Cambridge
29 PUBLICATIONS **279** CITATIONS
SEE PROFILE

Yao Shen
Tongji University
36 PUBLICATIONS **846** CITATIONS
SEE PROFILE

All content following this page was uploaded by Yuval Kahlon on 01 September 2024.

The user has requested enhancement of the downloaded file.

182

EXPLORING THE LINKAGE BETWEEN SPACE SYNTAX AND VISUAL IMPRESSION OF ARCHITECTURAL SPACES:

A case study of the Japanese streetscapes during moments of “Oku”

Stephen Law, University College London Geography, stephen.law@ucl.ac.uk

Yuval Kahlon, Tokyo Institute of Technology Architecture, yuvkah@gmail.com

Zicheng Fan, National University Singapore Architecture, zicheng.f@u.nus.edu

Michal Gath-Morad, Cambridge University Architecture, mg2068@cam.ac.uk

Demin Hu, University College London Bartlett School of Architecture, demin.hu.22@ucl.ac.uk

Yao Shen, Tongji University College of Architecture and Urban Planning, y.shen.12@ucl.ac.uk

Takuya Oki, Tokyo Institute of Technology Architecture, oki.t.ab@m.titech.ac.jp

Yen-Khang Nguyen-Tran, Tokyo Institute of Technology Architecture, khang.ntr@gmail.com

Ryo Murata, Tokyo Institute of Technology Architecture, murata.r.ac@m.titech.ac.jp

Naya Marsatyasti, Tokyo Institute of Technology Architecture, nayamarsatyasti@gmail.com

Haruyuki Fujii, Tokyo Institute of Technology Architecture, fujii.h.aa@m.titech.ac.jp

ABSTRACT

Recent research has shown increasing use of computational methods in the built environment for studying the visual impression of architectural spaces (VIAS) by either; (i)linking visual stimuli from imagery data with linguistic descriptions, (ii)analysing visual attention data from eye tracking devices, or (iii)a combination of these. Despite advancements in capturing multi-modal subjective impressions of architectural space, limited research links the geometry of the environment in relation to its impression. Comprehending how the geometry of the built environment affects our perception is essential for translating perceptual insights into designs. Our aim is thus to leverage on space syntax methodology, VIAS research and eye-tracking technology to explore the linkage between the spatial configuration of an urban space and its visual impression during moments of “Oku”, an abstract architectural concept in Japan to describe a sense of curiosity and depth. In particular, we examined subjective perception and conjecture that spaces with longer isovists(greater depth), higher occlusivity(curiosity) and visibility can potentially stimulate greater visual attention during moments of “Oku.” To trace the impression of Oku to the spatial configuration of the space, we: (1)collected gaze from *human subjects; (2)calculated visibility graph analytical measures, and (3)associated the two. Our exploratory results indicate a weak association between VIAS and SST visibility graph metrics. These findings highlight both the potential and the complexities of connecting our visual perceptions with the spatial layout of our built environment. Moving forward, we aim to assess gaze attention dynamically to uncover how the shape of our environment influences our senses within it.*

KEYWORDS

Visibility graph analysis, gaze analysis, visual impression, space syntax, streetviews

1 INTRODUCTION

Due to the increasing availability of computational methods, tools and data, there has been a growing use of quantitative techniques to investigate the visual impression of architectural spaces (VIAS) (Oki & Kizawa 2021, NGUYENTRAN et al. 2022). Two common approaches have been adopted in VIAS; first in linking visual stimuli from images with linguistic descriptions of the architectural experience (e.g. how we describe it) (Wada & Kishimoto 2011), and second by relating visual attention data with visual stimuli of the architectural space, via eye tracking technology (e.g. what we see) (Oki & Kizawa 2021, Kiefer et al. 2017). In a recent study (Nguyen-Tran et al. 2022), both of these approaches were applied in an attempt to shed light on the workings of VIAS. Despite these advancements, there has been limited research that connects the spatial configuration of the built environment with visual impression of architecture and urban scenes. Such connection is valuable as examining how the geometry of the built environment influences our visual impression can help us design spaces that induce more positive impressions and thus afford well-being for their users. Motivated by research in space syntax (Hillier & Hanson 1989, Hillier 2007), we aim to explore the linkage between the syntactic properties (SP) of urban spaces and VIAS. Rather than relating SPs with human behaviour (e.g. navigation) which has been looked at more extensively, we examine subjective perception, by studying how spaces are associated with abstract concepts experienced by human subjects. We focus on the Japanese architectural concept of “Oku”, which is used to describe a sense of curiosity and visual depth (Maki et al. 2018, Jonas 2011). We begin by providing background to these research domains followed by outlining our objectives and approach.

2 RELATED WORKS

2.1 CAPTURING THE VISUAL IMPRESSION OF ARCHITECTURAL SPACES

Visual impressions of architectural space (VIAS) is an emerging research discipline that refers to people's subjective impression or opinion of an architectural space. Two common approaches have been adopted in VIAS research. The first links visual stimuli with linguistic descriptions of the architectural experience (Wada & Kishimoto 2011) by correlating a subject's opinion and stimuli in photographs or actual physical spaces. This method has been applied in studies focusing on the perception of streetscapes, using indicators such as the sky-view factor (Nishio & Ito 2015), the green ratio (Nakamura et al. 2010), or the perceived attractiveness of urban spaces (Wada & Kishimoto 2011). The second connects visual attention data from visual stimuli with the architectural space (Oki & Kizawa 2021) with eye-tracking technology (Kiefer et al. 2017). For example, one study investigated the connection between the arrangement of landscapes, gaze patterns, and impression factors (Ohno 2018). Another study delved into the correlation between the gazing patterns of participants and their assessments of streetscape attractiveness (Oki & Kizawa 2021). In a recent study (Nguyen-Tran et al. 2022), both of these approaches were applied in an attempt to shed further light on the workings of VIAS.

2.2 SPATIAL COGNITION RELATION TO AFFECTIVE APPRAISAL

Furthermore, in the field of spatial cognition, visual impressions play a crucial role in how individuals perceive and respond to architectural spaces. This aspect is commonly referred to as affective appraisal (Russell 2003), a multifaceted process involving the dimensions of pleasure (positive or negative experience) and arousal (intensity of emotional response). To measure affective appraisal in architectural and urban environments, researchers use a combination of methods from reported preferences using the affect grid to skin conductance, EEG, and fMRI (Presti et al. 2022, Gregorians et al. 2022, Mavros et al. 2023). For example, a study by Gregorians et al. (2022) focusing on affective appraisal across a diverse range of architectural environments found that Arousal is associated with fascination, spatial complexity, and unusualness which affects our emotional and aesthetic experience. While arousal doesn't necessarily affect how coherent or comfortable a space feels, it does have a moderate impact on how fascinating we find it. How complex the layout is (spatial complexity), and how different or unique the space is (unusualness), are intricately connected to our emotional and aesthetic experience of a place. A study by Hollander et al. (2018) explored how edges, shapes, patterns, narrative, and biophilia, referred to as principles of Cognitive Architecture (CA), affect gaze behaviour and emotional experience. The allocation of gaze on a stimulus, and thus visual engagement with it, tends to reflect overt attention, and from this, the experiential significance of certain aspects of a scene can be inferred. Their findings from an eye-tracking study suggest that Urban environments exemplifying Cognitive Architecture (CA) principles were associated with more positive reactions regarding the likelihood of spending time in these places and feeling relaxed. Qualitative assessments of eye movement showed more pronounced patterns of fixation in CA urban environments compared to non-CA environments. This suggests that participants fixated more explicitly on face-like facades and well-defined edges in CA environments. In another study (Simpson et al. 2019) using eye tracking, the authors measured visual engagement with urban street features. Their findings suggest that urban street edges were the most visually engaged-with components across streets walked. Pedestrians' visual engagement was also affected by the walking task they performed. While existing research has connected architectural features with visual impressions and affective appraisal, there is a crucial gap. We need a more explicit approach that describe the spatial layout and geometry of our built environments and how it directly impacts our overall experiences.

2.3 SPACE SYNTAX AND ITS LIMITED CONNECTIONS WITH VISUAL IMPRESSION

As a result, we will leverage space syntax theories and methodologies to examine the influence of spatial geometry have on visual impression, extending from previous research (Hillier & Hanson 1989). Specifically, we will use visibility graph analysis (Turner et al. 2001), a technique that measures intervisibility in navigable space. Such analysis is often applied to capture wayfinding behaviour in complex buildings (Weisman 1981, Peponis et al. 1990, Wiener et al. 2009, Gath-Morad et al. 2021). However, limited research has been conducted on quantitatively linking human perception, such as gaze behaviour, to the geometry and spatial configuration of the built environment as capture by VGA analysis. One notable exception is Eloy et al. (2015), who linked gaze analysis with space syntax theory (Hillier & Hanson 1989, Hillier 2007) for studying navigation. In this research, the author acquired gaze information using eye tracking device and found that landmarks with higher visual integration correspond to greater counts of gaze fixations and saccades. Another research that looked at the intersection of perception and space syntax is Emo (2012). For this research, the author used eye-tracking with photographs in the real world to study objectively and quantitatively, where people looked during wayfinding. The research found that navigation decision strongly favours streets that have higher axial integration and that participants on average have more frequent and longer fixation on paths they eventually will choose. The author also highlighted people and traffic as attractors to attention and that depth of view, amount of visible space and vertical edges are possibly related to gaze fixations. These latter findings suggest there is indicative relationship between our perceptual attention and its spatial geometry which will be further examined in our study quantitatively.

3 AIM, OBJECTIVES AND SIGNIFICANCE

We aim to ground subjective VIAS with quantitative measurements of architectural space using Space Syntax theory. Our proposed interdisciplinary research hopes to bridge these related disciplines thereby enriching our understanding of architecture and human perception.

Understanding how we measure spaces and its effect on our impression is important for the progress of Computer-aided architectural design (CAAD) systems. At the moment, we do not have a well-established way of relating spatial configuration with VIAS resulting in the gap between the designer and human experience. Bridging this gap may enhance our ability to digitally design spaces which possess desirable experiential features from the users' perspective. We examine the concept of "OKU" in Japanese architecture (Maki et al. 2018), which concerns a sense of curiosity and visual depth, in relation to our Visual impression (VIAS) and their connection to the geometric and configuration of space. Specifically, we hypothesize that during moments of "OKU" (characterised by visual depth and curiosity), spaces exhibiting the greatest depth (visual connectivity), interest (visual occlusivity) and highest visual integration might overlap with our visual attention, as captured by eye-tracking devices.

4 METHOD

The study focused on a specific site in Daikanyama, Tokyo, Japan which is an exemplar of the sense of Oku. The neighbourhood contains buildings designed by Fumihiko Maki, the architect who defined the concept of Oku. We link VIAS to SPs in three phases:

A. collect gaze data from human subjects (10 subjects trained in architecture with 9 images)

B. compute visibility graphs analyses VGA in SST (Turner et al. 2001).

C. integrate the data from A and B, to relate SP with VIAS, in three steps:

C1. Visibility Graph Analysis: VGA measures are calculated for each space, and then projected into the same perspective view as the gaze analysis.

C2. Statistical Analysis: an Intersection Over Union (IOU) experiment is conducted where the top 10% of the VGA measures (in the perspective) and the 10% of the Gaze attention are intersected to measure the amount of overlaps between the two types of analysis.

C3. Comparative Analysis: experiment is conducted where the top 10% of the gaze attention are intersected with the scene semantic segmentation and depth of the same street scene.

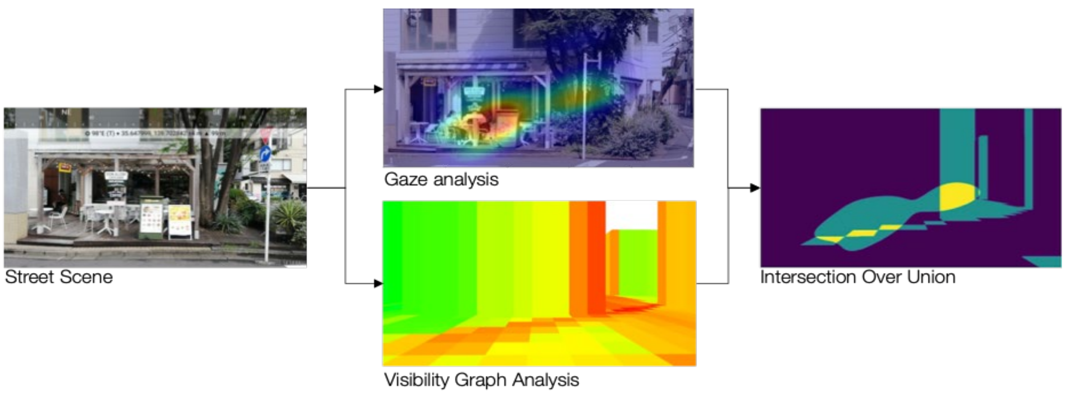

Figure 1: General pipeline for research method

4.1 GAZE ANALYSIS

Gaze data was collected, processed and analyzed, as follows. First, the site was photographically documented using a digital camera, focusing on several locations which are acknowledged as an embodiment of the concept of Oku.

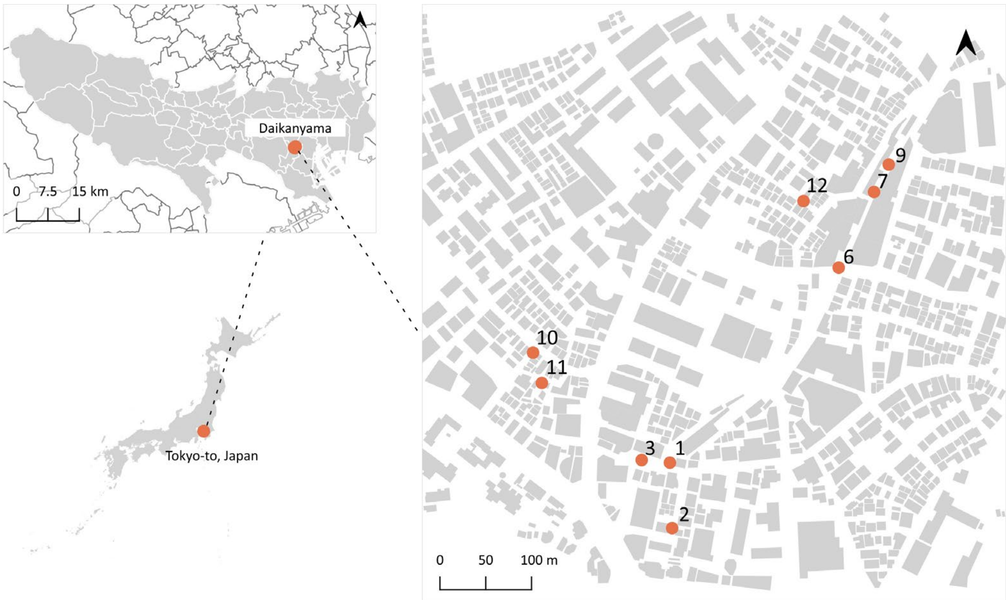

Figure 2: Case study area and the locations of street scenes

The photos were then projected on a wall in lab conditions to human subjects, who were interviewed regarding the degree to which the locations evoke a sense of “Okuness”, in their view. More specifically, subjects saw the photos one by one, and were asked to comment on their experience of Oku with respect to each (weak/strong etc.). Nine subjects were interviewed in total. During the interviews, gaze data was collected using a small screen-based static eye tracker (Tobii Pro nano), in the form of pupil diameter and gaze coordinates. Additionally, the conversation between the subject and the moderator was recorded using a standard commercial recording device, to enable inquiry into their personal perspective regarding the presence/absence of a sense of Oku in each photo. The collected gaze data was visualized as heatmaps, focusing on the specific time durations prior to and after discussing the notion of Oku, to reflect the subjects’ attention at these times. Given the speed s of a subject’s eye movements at a certain moment, gaze attention was defined as states where s¡ 10 deg/sec for longer than 100 msec, according to (Fukuda et al. 1996). The audio recordings were fully transcribed manually, and were segmented according to the photos under discussion in each segment. The heatmaps were then integrated with the verbal protocols into diagrams which contain attention and utterance data, and enable to compare them. By examining these, we identified relationships (matches/discrepancies) between mentions of Oku and the objects focused on in the physical environment. For additional details regarding the collection and analysis methodology, please refer to Nguyen-Tran et al. (2022).

Figure 3: Street scenes case study

4.2 VISIBILITY GRAPH ANALYSIS

In the next stage, we calculate the visibility graph analysis (Turner et al. 2001) for the same urban scenes. The procedure as shown in fig5 consists of the following steps: i. matching street scenes between real space and model space, ii. construct visibility graph, iii. calculate visual connectivity, occlusivity, and integration, iv. project the VGA results into the 3D model and v. visualise the results. The visibility graph analysis was calculated and visualised using the Python NetworkX Library, Decoding Space Library[1] and Grasshopper, a visual programming language available with Rhinoceros 7[2].

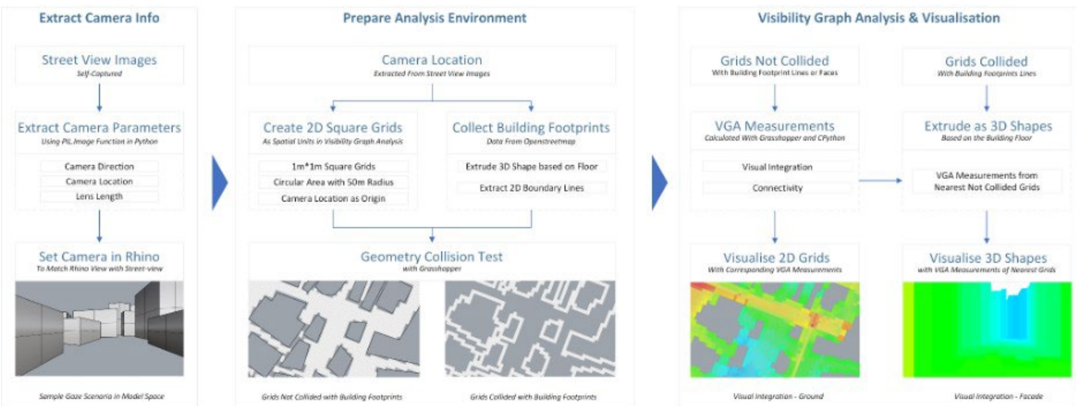

Figure 5: Visibility Graph Analysis pipeline

1 https://toolbox.decodingspaces.net
2 https://www.rhino3d.com/

4.2.1 MATCH VISUAL SCENES IN EXPERIMENTAL PHOTOS AND IN MODEL SPACE

The study first matches the street scenes from the photo captured in real space with the same view in the 3D modelling space. To begin, we download the building footprint data of the case study area in Central Tokyo with Open Street Maps. We then construct a 3D building model by extruding the building footprints to an assumed height. We then extract the photo meta information namely location, direction, and lens length and use these as input for the camera control parameters in Rhinoceros7 to simulate the pedestrian's eye-level view on the street. Additional manual calibration of the camera parameters were necessary to ensure the visual scenes in real space and model space matches.

4.2.2 CONSTRUCT VISIBILITY GRAPH

The study then constructs the Visibility Graph on the open street spaces around each photo from the experiment. Specifically, we centred on the ground projection of each case photo's location, compute a circular buffer with a radius of 50m and overlay a square grid of 1m x 1m cells. Then, based on the geometric relationship between the building footprints and the 2D grids, we extract: (1) the Ground-plane grid (Type A) where we retain the cells that do not collide with the building footprints; and (2) the Facade grid (Type B) where we only retain the cells that intersect with building boundaries. The former corresponds to the spaces between buildings; while the latter corresponds to the grid formed by the building boundaries. The undirected visibility graph is then constructed where each cell represents an abstract unit of space and as a node in the graph and link or an edge is created when two cells/nodes are mutually visible. The study then calculates three visibility graph measures namely visual connectivity, visual integration and visual occlusivity from the visibility graph of the ground plane grid. The ground-plane analysis for the nine areas can be found in the appendix.

Figure 4: Samples gaze attention for street scenes in Tokyo

4.2.3 CALCULATING VISUAL CONNECTIVITY, INTEGRATION, OCCLUSIVITY

Based on the visibility graph, we first calculate visual connectivity which characterizes the amount of space directly visible to a certain space, that is, the degree measurement for single a node in the graph (equation 1). The measure approximates the area of the isovists (Benedikt 1979).

$$Conn_i = Deg_i \tag{1}$$

Next, we calculate visual occlusivity which is another local metric that describes the potential of exploration or perceptual uncertainty (Benedikt 1979; Gibson 1977). This metric is calculated by measuring the length (number of cells) of the occluded boundary R_i from cell i (equation 2);

$$Occ_i = R_i \tag{2}$$

Lastly, we calculate visual integration which is a global measure that indicates the visual nearness between one space to all other spaces in the system. Generally, the calculation of visual integration is closely related with Closeness Centrality (CC_i) in graph theory, or the reciprocal of Mean Depth (MD_i) of the graph (equation 3).

$$CC_i = \frac{1}{MD_i} = \frac{n-1}{\sum_j d_{ij}} \tag{3}$$

where CC_i represents the closeness centrality of node i, MD_i represents the mean depth for node i, d_{ij} represents the depth of the shortest path between i and j in the visibility graph and n represents the total number of nodes in the graph. Considering the symmetry and scale difference between different spatial systems, Hillier & Hanson (1989) proposed the Relativized Asymmetry value (RA) (see equation 4), and the RA of Diamond graph(D) (equation 5), to normalise Closeness Centrality (equation 6). Thus, visual integration is calculated with;

$$RA_i = \frac{MD_i - 1}{\frac{n}{2} - 1} \tag{4}$$

Where RA represents Relativised Assymetry of any node I in the graph;

$$D = \frac{n(log_2(n/3-1))+1}{(n-1)(n-2)/2} \tag{5}$$

D represents Relativised Asymmetry of the diamond graph corresponding to the total no. of nodes;

$$Int_i = D/RA_i \tag{6}$$

and Int_i as the normalised closeness of the visibility graph.

4.2.4 PROJECTING VGA RESULTS INTO 3D SPACE

Normal VGA simplifies the 3D space into a 2D grid where the results cannot be directly mapped into a 3D perspective necessary to match the gaze attention. For this reason, this study resolves this by projecting 2D VGA results onto the ground and the building facades for analysis and visualization. This is done by assigning the values from the cells of the ground plane grid to the nearest/intersecting cell of the facade cell which then gets extruded in representing the facade visibility of the buildings. By combining the 2D VGA values calculated on the ground plane grid and the facade visibility from the extruded cells, we are then able to project the visibility graph analysis in 3D perspective that corresponds with the street scenes from the experiment. Future research will use more advance 3D visibility graph analysis instead (Varoudis & Psarra 2014).

4.3 VISIBILITY GRAPH ANALYSIS VISUALISATION

The VGA results were remapped within a range of 0-0.7, and converted into HSL Color for visualization as seen below where red indicates higher VGA values and blue indicates lower VGA values. Figure 6a,b,c show the visual integration, visual connectivity and visual occlusivity of the nine urban scenes respectively. The three VGA metrics show some similarity, for example in scene 3 and 12, where all three metrics seem to have higher values in similar areas. However, the three VGA metrics also show dissimilarity, for example scene 11 is highly homogeneous in terms of visual connectivity but more heterogeneous in terms of visual integration and visual occlusivity. Scene 3,6, 9 which faces a street/path also seem to have more varied VGA values. While scene 1,2 which faces a street facade seems to have higher VGA values along the border near the street space. And finally, the VGA visualisations of scene 11 does not capture the visibility of the highly occluded/cluttered space due to the lack of street furniture. Future research should try to capture the complex urban environment in measuring visibility.

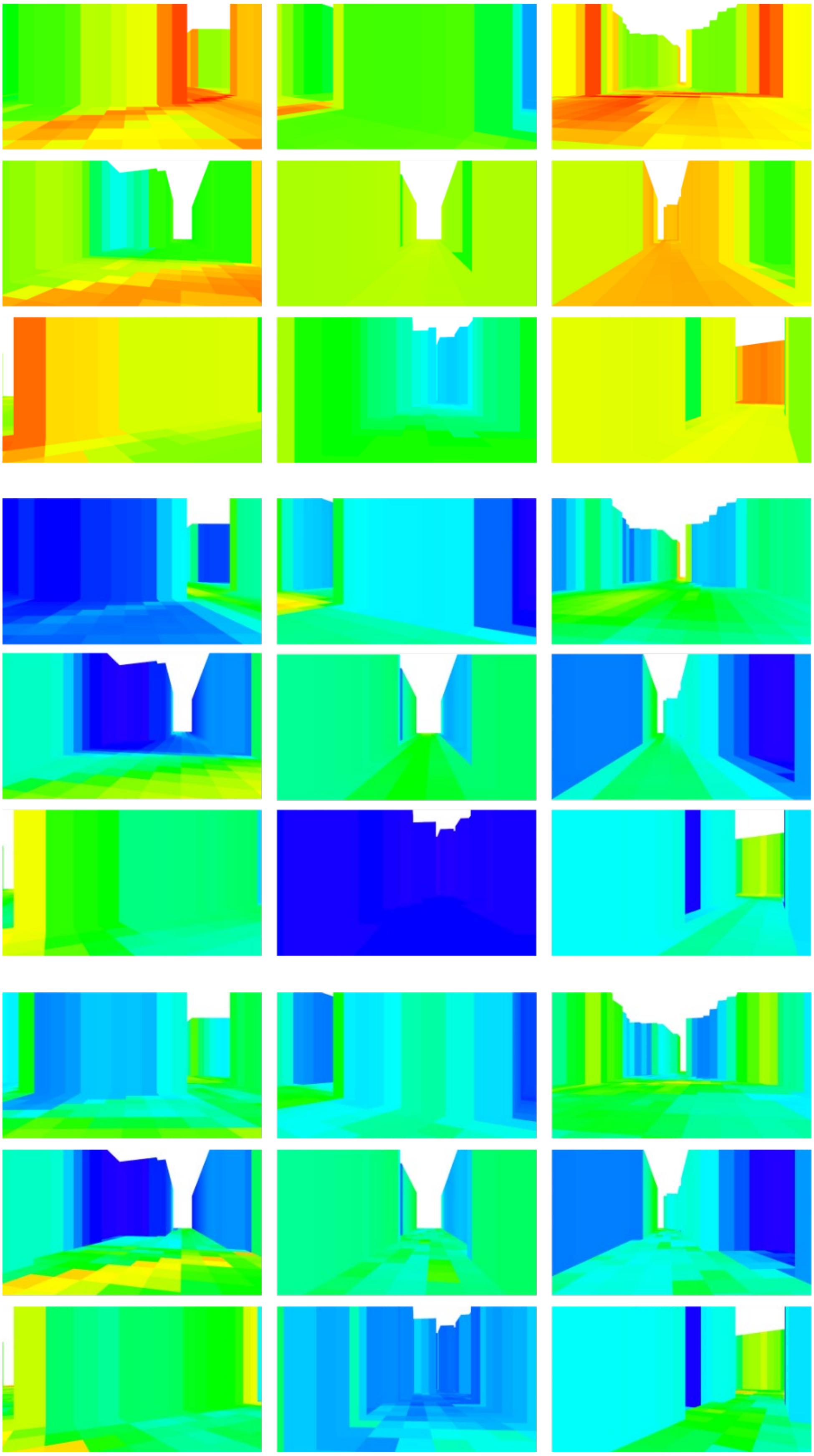

Figure 6: (Top) a. Visual Integration, (Middle) b. Visual Connectivity, (Bottom) c. Visual occlusivity for the case studies in Tokyo.

4.4 STATISTICAL ANALYSIS - INTERSECTION OVER UNION EXPERIMENT

To study the extent gaze attention associate with the visibility graph analysis metrics, we conducted a simple Intersection Over Union experiment shown in fig7. A standard metric in computer vision, specifically object detection, is to measure the extent two boundaries overlap through its Intersection Over Union (IoU). IoU measures the extent the predicted bounding box (bbox) of the object matches with the ground truth bbox of the object. Adapted for our study, we instead measure the overlap between the top 10% of gaze attention and the top 10% of the in-scene visibility graph analysis measures for each view. The first step in the experiment is to threshold the top 10% of the gaze attention and the top 10% of the VGA measures in the scene. The second step is cleaning where we remove the sample if the top 10% of gaze value is 0 or if the variance of the visibility metric is less than 5% to remove erroneous gazes in the first condition and scenes with a lack of visibility variability in the second condition. In the third step we measure the Intersection Over Union which is simply the area of overlap divided by the area of union between the top 10% gaze attention and the top 10% of the in-scene VGA values. Finally, we repeat this for every single scene, every single gaze from each participant and every single visibility graph measures. We then report the average IoU metric for each scene and each SST measure where the higher the metric the more overlap there is between the gaze attention and the visibility graph metric.

Figure 7: Intersection Over Union

4.5 COMPARATIVE ANALYSIS EXPERIMENTS

Lastly, two comparative experiments are conducted, building on previous studies that linked the built environment features to eye fixation (Hollander et al. 2018; Simpson et al. 2019). The first studies the extent the top 10% of the gaze attention is intersected with the scene semantic such as wall, building, sky, tree, road, pavement, people and car class (mIOU) to better understand where our participants are looking. For this analysis, we leverage on the widely used Segformer model which comprised of a hierarchical Transformer encoder and a MLP decoder that aggregates both local and global attention for semantic segmentation (Xie et al. 2021). The second is where the top 10% of the gaze attention is intersected with the scene depth of the same street scene (mIOU). For this analysis, the relative scene depth is estimated from a standard monocular depth estimation model (Ranftl et al. 2020). Figure 8 shows from top left to bottom right, a street scene in Daikanyama in Tokyo, its gaze attention, scene depth, and scene semantic segmentation where the different colours represent the different semantic classes.

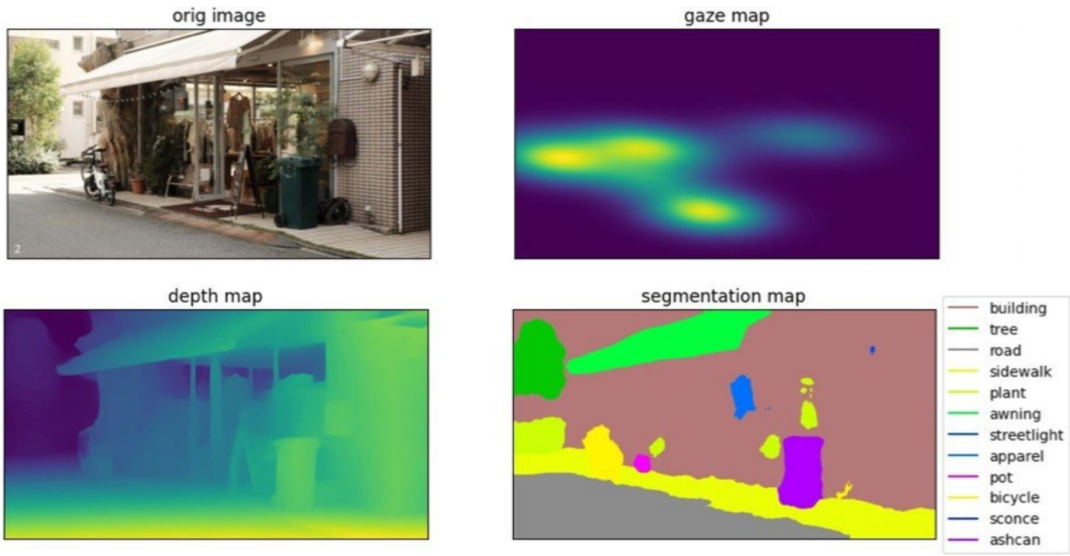

Figure 8: Left to right: street scene in Tokyo, its gaze attention, scene depth and scene semantics.

5 RESULTS

5.1 QUALITATIVE RESULTS

Figure 9 shows some qualitative results examining the overlap between the top 10% of the gaze attention (in green) and the three syntactic indices (in green) and their overlaps (in yellow). The sample result indicates two main findings: firstly, the results show weak overlap between visual attention and visual integration where the result seems to be quite heterogeneous. For example, high visual attention coincides with high visual integration of the shop entrance and trees in the first scene, the public space with seating in the third scene, and to some extent the public space in the sixth scene. While, there are also scenes where the overlap is not observable such as in scene ten. Secondly, the result seems to be slightly better with visual integration when compared with visual connectivity and visual occlusivity qualitatively.

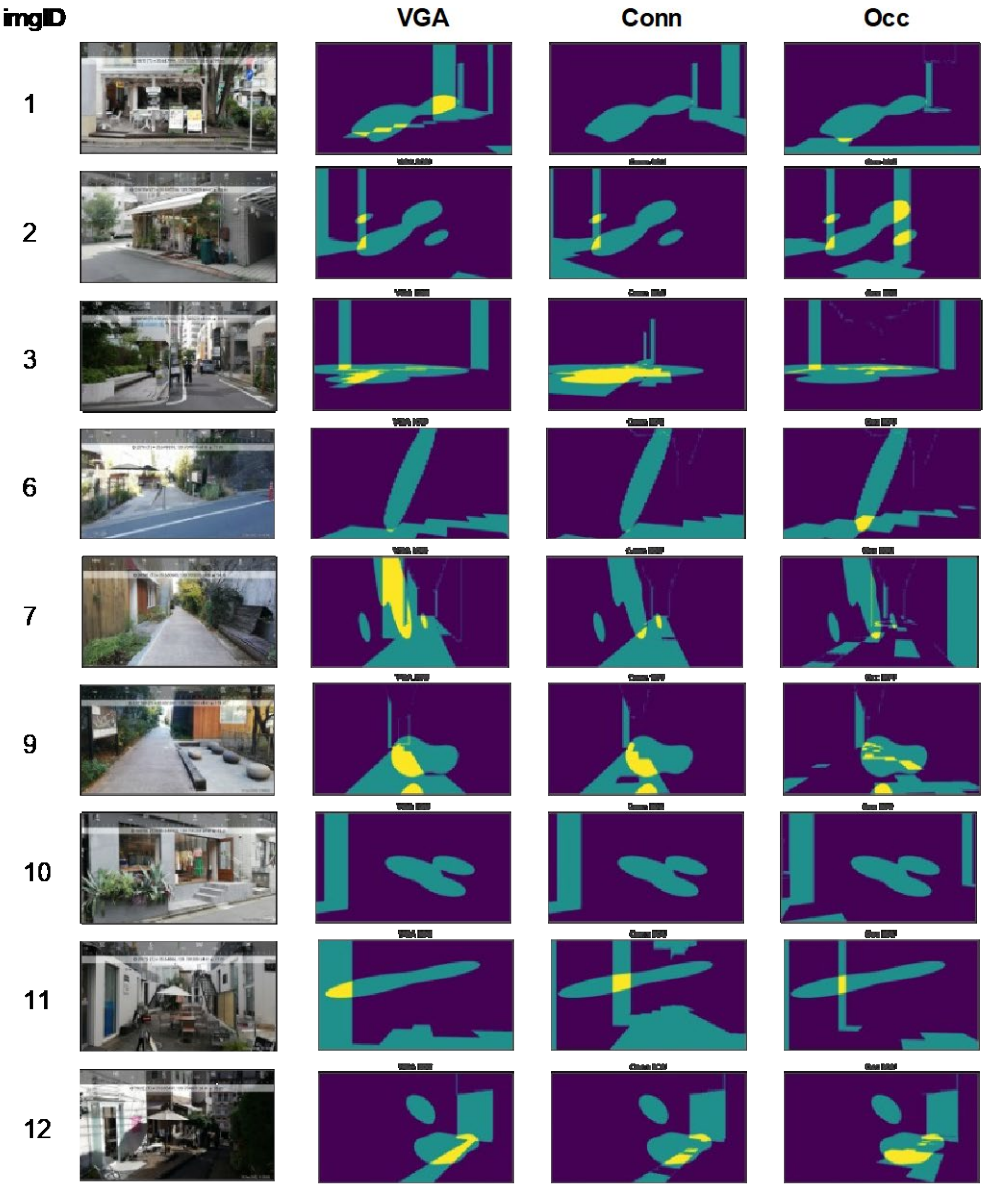

Figure 9: Sample showing overlap between the top 10% of the gaze attention (in green) and the syntactic indices (in green) and their overlaps (in yellow).

5.2 QUANTITATIVE RESULTS

To further validate whether the overlaps between visual attention and visibility metrics are significant, we computed these overlaps across eight street scenes, each containing a range of 7 to 15 participants per scene. We have excluded scene 11 from the statistical analysis as the scene contains street furniture that significantly occludes the visibility of the urban space. In total we have 79 samples for the eight scenes. Table 1 shows the mean Intersection Over Union for the eight scenes and each of the visibility graph metrics. The outcome indicates that among the three metrics considered, visual integration performed the most effectively, with 6 out of 8 urban scenes exhibiting over 5% alignment between visual attention and visual integration. Notably, scene 6 and scene 10 had little to no overlap. In scene 6, the highest visibility is observed along the street, contrasting with the most attended area, which is along the path. In scene 10, the highest visibility is the area to the left of the shop, which differs from the area with the highest attention, encompassing the shop and staircase. Excluding the two scenes with no overlap, the mean Intersection Over Union (IOU) is calculated to be 12.16 for all scenes and all participants. Visual connectivity exhibited the lowest performance, with only 4 out of 8 urban scenes displaying over 5% overlap. On the other hand, visual occlusivity showed slightly improved results, with 5 out of 8 urban scenes displaying more than 5% overlap. These findings generally indicate a significant but weak overlap between visual attention and visual integration, and very little to no alignment for the other two visibility measures confirming the qualitative observations.

ImgID	N	VGA mIOU	Conn mIOU	Occ mIOU
1	15	**6.26%**	0.84%	**5.60%**
2	8	**7.90%**	**7.87%**	**7.30%**
3	11	**9.38%**	**24.12%**	**6.67%**
6	9	0.02%	0.02%	0.09%
7	8	**24.12%**	0.61%	1.54%
9	11	**10.20%**	**9.56%**	**10.50%**
10	7	0.00%	0.00%	1.24%
12	10	**15.11%**	**12.83%**	**25.45%**

Table 1: Summary table for the IoU experiment between visual attention and visibility graph metrics

5.3 COMPARATIVE ANALYSIS RESULTS - URBAN SCENE SEMANTICS

To compare and contrast with other baseline scene measures, we study the extent the visual gaze attention intersects with common urban scene semantics firstly. Table 2 shows the mean Intersection Over Union (mIOU) between gaze attention and each of the urban scene semantics (e.g. wall, building, sky, tree, road, pavement, people and car class). The result shows participants on avearge had consistently higher eye fixation on buildings in images where 8/8 scenes had a mIOU of over 5%. This is followed by the pavement where 5/8 scenes had a mIOU of over 5% and the people class where 3/8 scenes had a mIOU of over 5%.

ImgID	N	bldg	sky	tree	road	pave	people	car
1	15	**6.76%**	0.04%	1.55%	3.09%	**12.14%**	**8.08%**	0.00%
2	8	**10.98%**	0.00%	0.09%	1.39%	**5.11%**	0.00%	0.00%
3	11	**9.33%**	0.02%	0.68%	1.03%	4.00%	**11.29%**	0.83%
6	9	**10.36%**	0.05%	**8.93%**	0.56%	**9.18%**	0.62%	0.00%
7	8	**13.22%**	0.08%	**10.23%**	0.00%	1.60%	0.64%	0.00%
9	11	**6.26%**	1.06%	4.02%	0.02%	**8.02%**	**5.39%**	0.00%
10	7	**12.54%**	0.00%	0.00%	0.80%	0.47%	0.00%	0.00%
12	10	**5.48%**	0.02%	1.01%	2.97%	21.33%	0.11%	0.38%

Table 2: Summary table for the IoU experiment between visual attention and scene semantics.

We then also study the extent the visual gaze attention intersects with the scene depth of the image. Table 3 shows the mean Intersection Over Union (mIOU) between the gaze attention and the relative depth for each urban scene. The result shows 6/8 scenes had a mIOU of over 5%. The mean IOU for all scenes above 5% was slightly higher with scene depth (12.53%) compared to Visual integration (12.16%). The highest mIOU is for scene 6,7 and 9. These results indicate that on average participants had greater visual attention in spaces that had higher scene depth.

ImgID	N	Scene depth mIOU
1	15	0.70%
2	8	**7.62%**
3	11	**5.48%**
6	9	**17.39%**
7	8	**21.86%**
9	11	**15.43%**
10	7	**7.27%**
12	10	3.92%

Table 3: Summary for the IOU experiment between visual attention and scene depth.

6 CONCLUSION, LIMITATIONS AND FUTURE WORK

This is to our knowledge, the first exploratory study that tries to investigate the linkage between the spatial configuration of the urban space and human perceptual visual attention of it during moments of “oku”, which is an abstract Japanese concept related to inner depth and curiosity. The results show that there is a weak overlap between the most visually attentive space as captured by the eye tracker during moments of Oku and the highest visual integration space. One explanation for the overlap with visual integration could be related to the fact that the most integrated spaces induce a sense of Oku or inner depth. This is related with the results from the second comparative experiment where gaze attention overlaps with scene depth in 6/8 scenes. Specifically, the results are the highest for scene 6, 7 and 9 where the street scene has longer isovists and greater scene depth in the view. Another explanation is that these spaces also contain or correlate with more interesting urban elements that induce the same sense of “Oku”. For example, the location of public spaces in scene 3 and scene 9 is also in an area with higher visual integration. While in scene 1, the most integrated space also has the most seats which the shop owner might have purposedly placed.

The relatively weaker result specially with the other two visibility metrics could be related to the fact that the urban scenes are highly complex while the configuration is relatively simple with homogeneous visibility graph metric values for some of the scenes. In particular, scene 1, 10 and 11 faces the building facades with relatively homogeneous visibility graph metric values. This is supported with the results from the first comparative experiment where visual attention most aligns with the building semantic class. This also supports the argument that in urban scenes, the richness of the urban facade is possibly capturing the visual attention of the individual rather than as afforded by the geometry. Further research is necessary to assess whether specific semantic arrangements and configuration are more conducive to visual attention during these “Oku” moments.

These results are not conclusive as many limitations remain for this exploratory study. This includes fundamentally the complexity of the urban environments. For example, are we capturing attention on highly integrated space or the urban elements that correlate with highly integrated spaces. Future research should consider the application of quantitative methods to either statistically control or remove urban furniture from the urban scenes without losing its realism and calculating more advance 3D geometrical/configuration analysis (3D VGA) in relating to visual attention. Another direction is to explore beyond “Oku”, for example do specific spatial configuration or semantic arrangement induces other urban impression such as excitement or boredom, calmness or stress. Another limitation is the measurement of visual attention under the impression of “Oku” using static imagery within a lab setting. As humans fundamentally experience urban space dynamically and in-the-wild (e.g. moving from a low depth space to a high depth space) rather than statically with images projected in a computer lab, future research should try to measure in-the-wild gaze attention along a pedestrian path to reveal further insights between the geometry of the urban space and its visual impression whilst moving. Some of these future works can be summarised into a pipeline as shown in fig10.

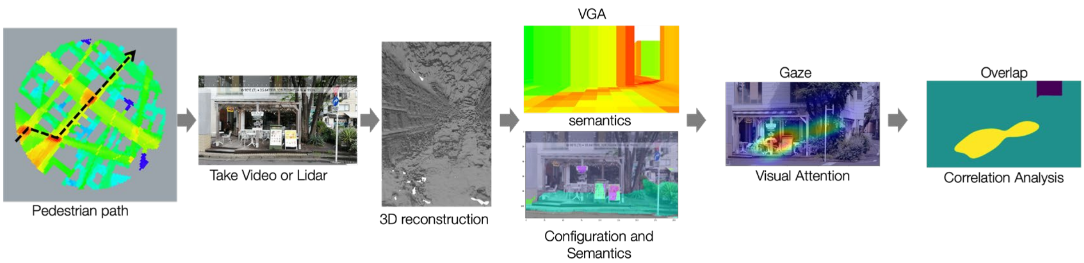

Figure 10: Future works linking dynamic visual attention with advance syntactic and semantic analysis.

This exploratory study between visual impression and spatial configuration shows some initial evidence albeit weakly between the geometry of the urban space and its perceptual attention during moments of “Oku” (sense of curiosity and depth). Due to the complexity of urban environment and individual perceptions, further research is necessary to confirm the existence or non-existence of this relationship controlling for the many urban factors. Despite the difficulty in disentangling such relationships, this research is highly valuable for urban designers and architecture researchers. It represents the first step into building a richer understanding on how the geometry of the built environment and its visual impression relate with each other which is fundamental into translating the perceptual impression of urban space into actionable urban design.

REFERENCES

Benedikt, M. L. (1979), 'To take hold of space: isovists and isovist fields', Environment and Planning B: Planning and design 6(1), 47–65.

Eloy, S., Ourique, F., Resende, R., Dias, J., Freitas, J. & Pedro, T. (2015), 'Analysing people's movement in the built environment via space syntax, objective tracking and gaze data', eCAADe pp. 341–350.

Emo, B. (2012), Wayfinding in real cities: Experiments at street corners, in 'Spatial Cognition VIII: International Conference, Spatial Cognition 2012, Kloster Seeon, Germany, August 31–September 3, 2012. Proceedings 8', Springer, pp. 461–477.

Fukuda, R., Sakuma, M., Nakamura, E. & Fukuda, T. (1996), 'An experimental study on the definition of gazing point', Ergonomics 32(17), 197–204.

Gath-Morad, M., Thrash, T., Schicker, J., H¨olscher, C., Helbing, D. & Aguilar Melgar, L. E. (2021), 'Visibility matters during wayfinding in the vertical', Scientific reports 11(1), 18980.

Gibson, J. J. (1977), 'The theory of affordances', Hilldale, USA 1(2), 67–82.

Gregorians, L., Velasco, P. F., Zisch, F. & Spiers, H. J. (2022), 'Architectural experience: Clarifying its central components and their relation to core affect with a set of first-person-view videos', Journal of Environmental Psychology 82, 101841.

Hillier, B. (2007), Space is the machine: a configurational theory of architecture, Space Syntax.

Hillier, B. & Hanson, J. (1989), The social logic of space, Cambridge university press.

Hollander, J. B., Purdy, A., Wiley, A., Foster, V., Jacob, R. J., Taylor, H. A. & Bruny´e, T. T. (2018), 'Seeing the city: Using eye-tracking technology to explore cognitive responses to the built environment', Journal of Urbanism: International Research on Placemaking and Urban Sustainability .

Jonas, M. (2011), Oku: the notion of interior in tokyo's urban landscape, in 'Urban interior: informal explorations, interventions and occupations', Spurbuchverlag, pp. 99–110.

Kiefer, P., Giannopoulos, I., Raubal, M. & Duchowski, A. (2017), 'Eye tracking for spatial research: Cognition, computation, challenges', Spatial Cognition & Computation 17(1-2), 1–19.

Maki, F., Wakatsuki, Y., Ono, H., Takatani, T. & Pollock, N. R. (2018),¯ City with a hidden past, Kajima Institute Publishing Company, Limited.

Mavros, P., Liang, N. Z., Kirk, S., Te, A., Gru¨bel, J., Melgar, L. A., Guizzo, A. O. & Makowski, D. (2023), 'Attenuating subjective crowding through beauty: an online study on the interaction between environment aesthetics, typology and crowdedness'.

Nakamura, K., Kita, M. & Matsumoto, K. (2010), ‘Evaluating impression of vertical greenery using sd method’, Journal of Architecture and Planning (Transactions of AIJ) 75(654), 1943–1951.

Nguyen-Tran, K., Kahlon, Y., Oki, T., Marsatyasti, N., Murata, R. & Fujii, H. (2022), ‘Exploring users’ visual impression of a japanese streetscape by correlating attention with speech’, eCAADe .

Nishio, S. & Ito, F. (2015), ‘The effect of sky factor and the change on impression of townscape, with the experiment for evaluation impression while walking’, Journal of Architecture and Planning (Transactions of AIJ) 80(710), 907– 914.

Ohno, R. (2018), ‘Studies on environmental perception during locomotion—a review of empirical studies by the ohno laboratory’, Japan Architectural Review 1(2), 194–206.

Oki, T. & Kizawa, S. (2021), ‘Evaluating visual impressions based on gaze analysis and deep learning: A case study of attractiveness evaluation of streets in densely built-up wooden residential area’, The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 43, 887– 894.

Peponis, J., Zimring, C. & Choi, Y. K. (1990), ‘Finding the building in wayfinding’, Environment and behavior 22(5), 555–590.

Presti, P., Ruzzon, D., Avanzini, P., Caruana, F., Rizzolatti, G. & Vecchiato, G. (2022), ‘Measuring arousal and valence generated by the dynamic experience of architectural forms in virtual environments’, Scientific reports 12(1), 13376.

Ranftl, R., Lasinger, K., Hafner, D., Schindler, K. & Koltun, V. (2020), ‘Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer’, IEEE transactions on pattern analysis and machine intelligence 44(3), 1623–1637.

Russell, J. A. (2003), ‘Core affect and the psychological construction of emotion.’, Psychological review 110(1), 145.

Simpson, J., Freeth, M., Simpson, K. J. & Thwaites, K. (2019), ‘Visual engagement with urban street edges: Insights using mobile eye-tracking’, Journal of Urbanism: International Research on Placemaking and Urban Sustainability 12(3), 259–278.

Turner, A., Doxa, M., O’sullivan, D. & Penn, A. (2001), ‘From isovists to visibility graphs: a methodology for the analysis of architectural space’, Environment and Planning B: Planning and design 28(1), 103–121.

Varoudis, T. & Psarra, S. (2014), ‘Beyond two dimensions: architecture through three dimensional visibility graph analysis’, The Journal of Space Syntax 5(1), 91–108.

Wada, S. & Kishimoto, T. (2011), ‘Impressions and oppressed feelings according to the distance from a skyscraper’, Journal of Architecture and Planning (Transactions of AIJ) 76(667), 1651–1657.

Weisman, J. (1981), 'Evaluating architectural legibility: Way-finding in the built environment', Environment and behavior 13(2), 189–204.

Wiener, J. M., Buchner, S. J. & Holscher, C. (2009), 'Taxonomy of human wayfinding tasks: A knowledge-based approach', Spatial Cognition & Computation 9(2), 152–165.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M. & Luo, P. (2021), 'Segformer: Simple and efficient design for semantic segmentation with transformers', Advances in Neural Information Processing Systems 34, 12077– 12090.

APPENDIX

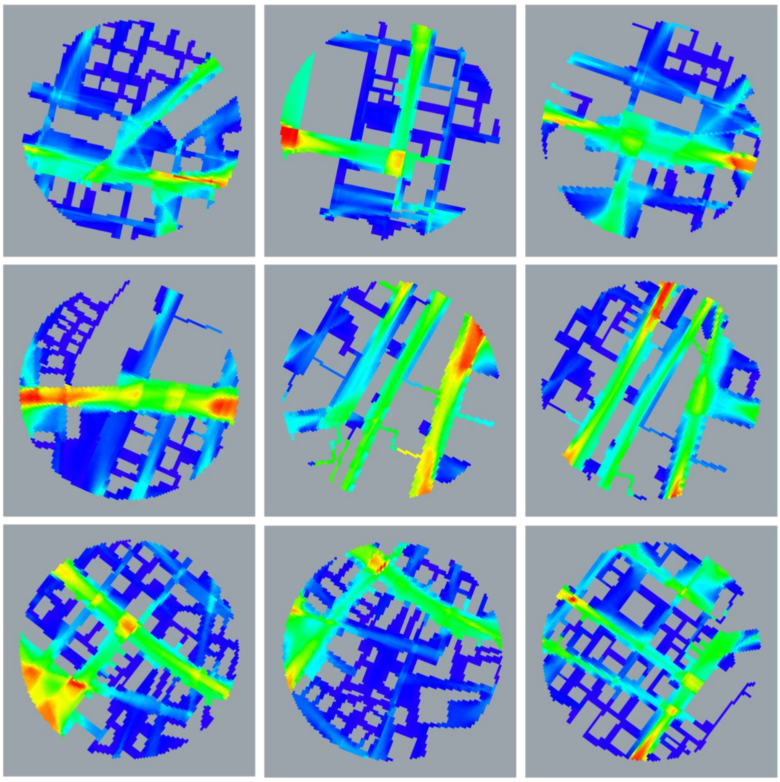

Figure A-1. Visual connectivity for the case studies

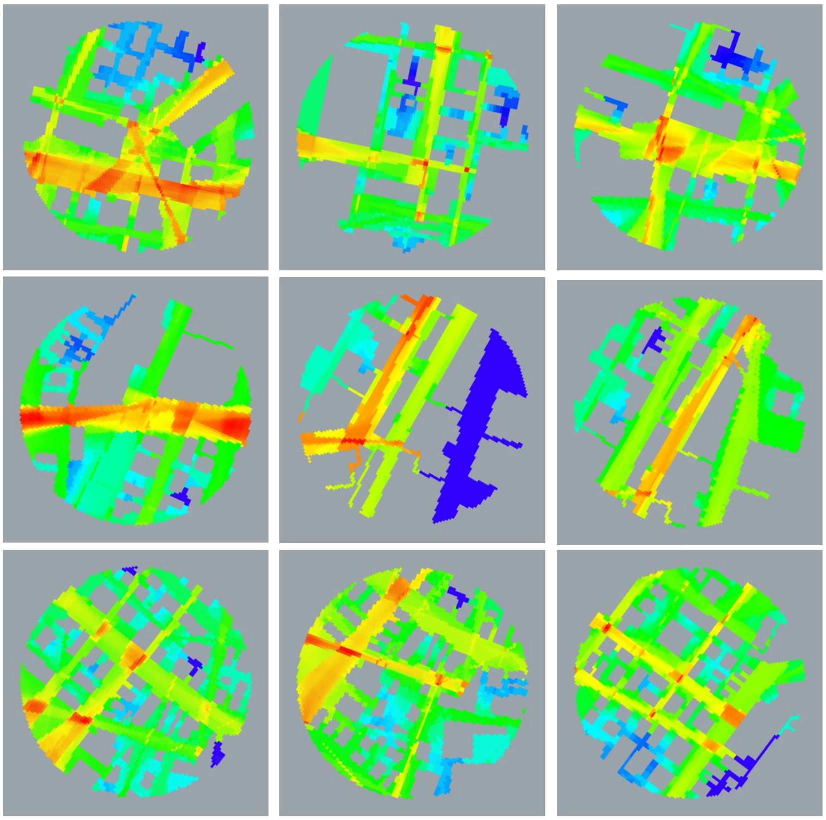

Figure A-2. Visual integration for the case studies

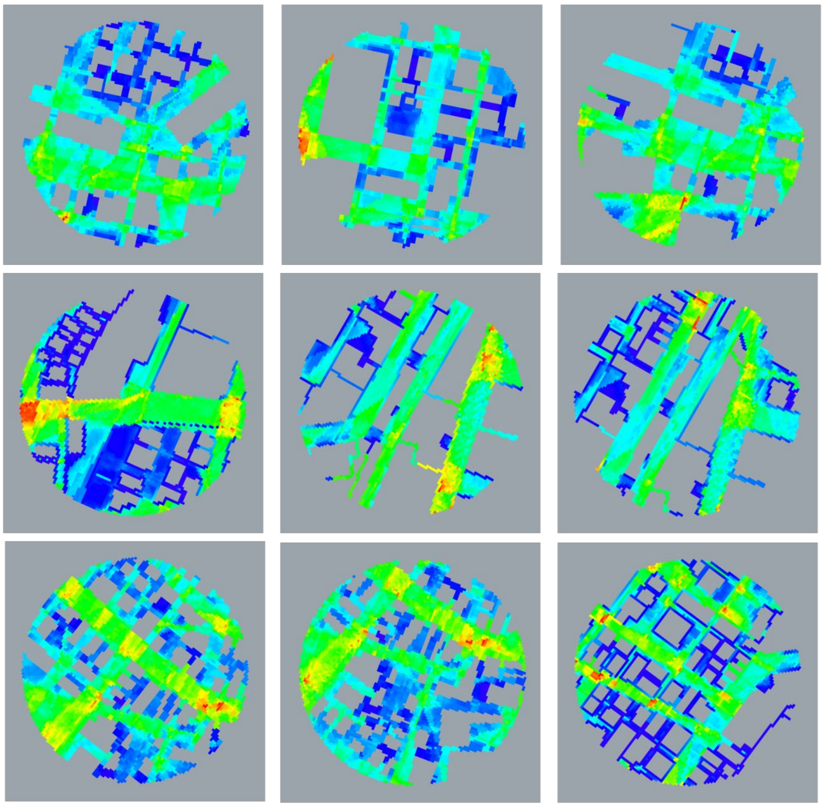

Figure A-3. Visual occlusivity for the case studies

View publication stats