A Multifaceted Exploration of Spatial Openness in Rental Housing: Big Data Analysis Across Tokyo's 23 Wards

賃貸住宅における空間開放性の多面的探究：東京23区のビッグデータ分析

OKI Lab 23M58033 リュ　ゲン (Liu Yuan)

1 Introduction

Spatial openness encompasses perceived spaciousness, visual connectivity, and flow within living environments. Traditional approaches using floor area and room count fail to capture nuanced spatial qualities from interior imagery and floor plans. This research proposes a methodology integrating interior images with semantic segmentation and visibility graph analysis to quantify spatial openness in rental housing. We analyze distributions across Tokyo's 23 wards and construction decades, investigating correlations with subjective impression scores. Our approach enables automated spatial quality analysis, opening possibilities for large-scale urban studies of residential design.

2 Openness Attributes Extraction

2.1 Dataset and Attributes

Our dataset comprises a few thousand rental properties from Tokyo's 23 wards (1960-present), sampled to ensure each decade has approximately equal representation and filtered for image quality and data completeness. The data selection process is shown as Figure 1.

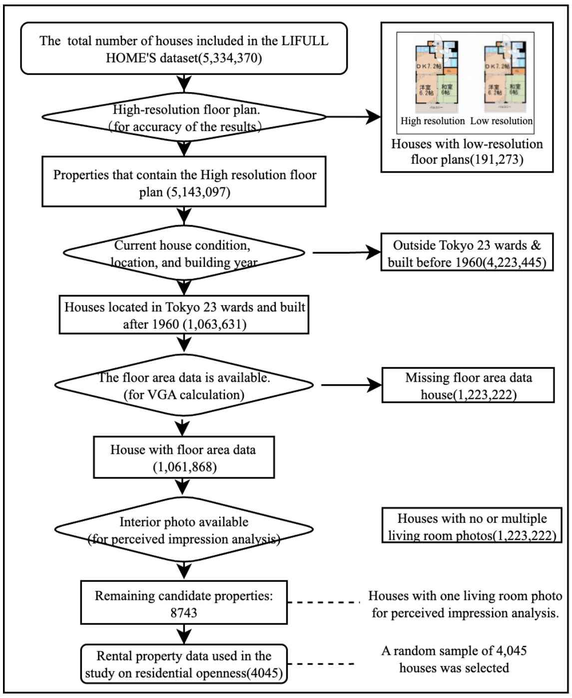

Figure 1: Data selection and filtering process.

As illustrated in Figure 2, this overview shows our methodology extracts multiple categories of spatial openness attributes through 3D interior images, 2D floorplan data, and property specifications, as well as our following data analysis.

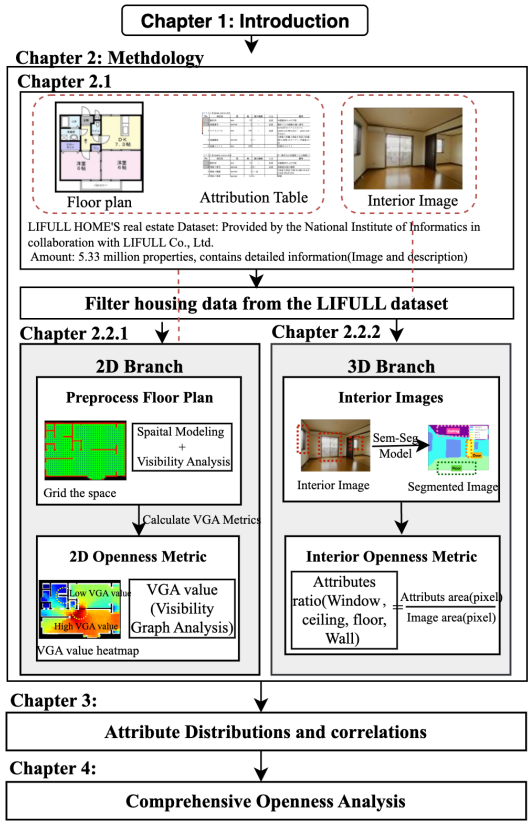

Figure 2: Feature extraction overview.

Impression scores (Q3) are extracted using Shimomura et al.'s pre-trained model. The following attributes are the focus of our analysis: **room size attributes**, **property specifications**, **interior elements**, and **vga-based statistics**. Room size attributes and property specifications are obtained directly from the database. Interior elements are extracted through semantic segmentation of interior images, while VGA-based statistics are computed from floorplan analysis to quantify spatial connectivity.

2.2 Interior Elements

We apply Mask2Former semantic segmentation to extract wall, ceiling, floor, and window ratios from interior images. A CNN extracts features for impression score generation via MLP, as shown in Figure 3.

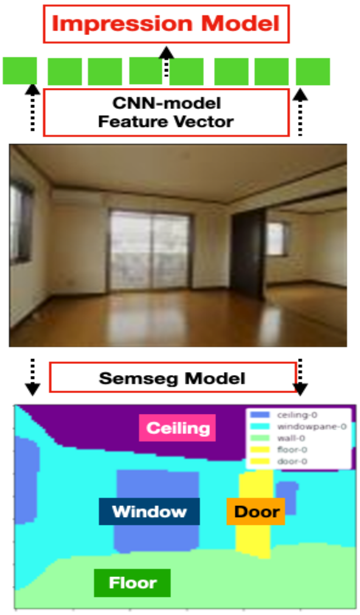

Figure 3: Interior semantic segmentation and impression scoring.

2.3 VGA-based Statistics

For floorplan data, we use semantic segmentation to identify walls, rooms, and open areas. We then create a grid over the open areas and calculate VGA (Visibility Graph Analysis) values:

$$S(i) = \sum_{j=1}^{N} V_{ij} \tag{1}$$

where $S(i)$ is the visibility score at node i, and $V_{ij} = 1$ if node j is visible from node i, otherwise $V_{ij} = 0$. This creates a VGA heatmap showing spatial connectivity throughout the floorplan. We extract summary statistics (mean, standard deviation, percentiles) from this heatmap as spatial openness features, as shown in Figure 4. After extracting the statistics, we can utilize them to analyze their distributions and correlations with other attributes to enhance the understanding of spatial openness. of the openness of the properties.

3 Attribute Distributions and Correlations

3.1 Temporal Analysis of Spatial Openness

From the tabular features of the properties, we have various information, with the most interesting being building age and geographic factors. We examined the distribution of VGA mean values across construction dates in decades, as shown in Figure 5.

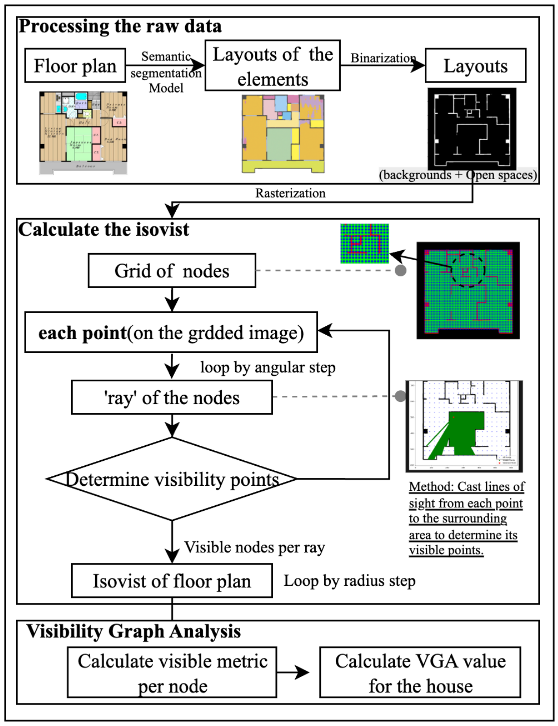

Figure 4: Workflow of creating VGA heatmap from floorplan data: raw floorplan, semantic segmentation output, physical gridding, and VGA heatmap.

The analysis shows VGA mean values contain large outliers but remain stable across decades, likely due to large-area properties. The impression scores show more balanced distributions, as they are evaluated from interior images that are less dependent of overall property layouts.

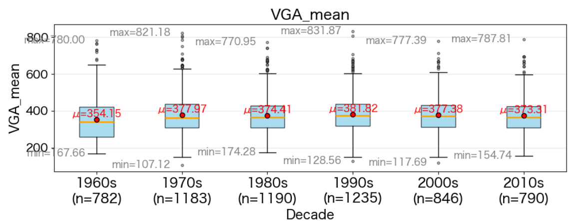

Figure 5: Distribution of VGA mean values across construction dates by decade

3.2 Geographic Distribution Analysis

We conducted a comprehensive analysis of the geographic distribution of VGA mean values and impression scores at the chome-level granularity using postcode alignment across Tokyo's 23 special wards, as shown in Figure 6. This fine-grained spatial analysis allows us to examine potential neighborhood-level variations in spatial openness characteristics. However, despite this detailed geographic examination, no clear clustering patterns or significant imbalances are found in the distributions across the different areas for either attribute, suggesting that spatial openness characteristics are relatively uniformly distributed across Tokyo's urban landscape.

3.3 Correlation Analysis Between Openness Attributes

We conducted a detailed correlation analysis examining the relationships between VGA statistics, impression scores, and interior element fractions to understand how different measures of spatial openness relate to each other. Figure 7 reveals only mild correlations between VGA metrics derived from floorplan analysis and impression scores obtained from interior imagery assessment, suggesting fundamental gaps between objective floorplan-based spatial analysis and subjective imagery-based perceptual assessment of spatial openness.

4 Comprehensive Openness Analysis

4.1 Analysis of Impression Scores and Interior Elements

We performed core-component dependency analysis combining impression scores with interior element ratios after applying rigorous outlier filtering using a 75th percentile cutoff to ensure data quality. The cluster analysis reveals distinct groupings based on floor and ceiling visibility characteristics, with moderate negative correlations observed with impression scores, indicating that higher visibility of structural elements tends to correspond with lower perceived spatial openness, as demonstrated in Figure 8 and quantified in Table 1.

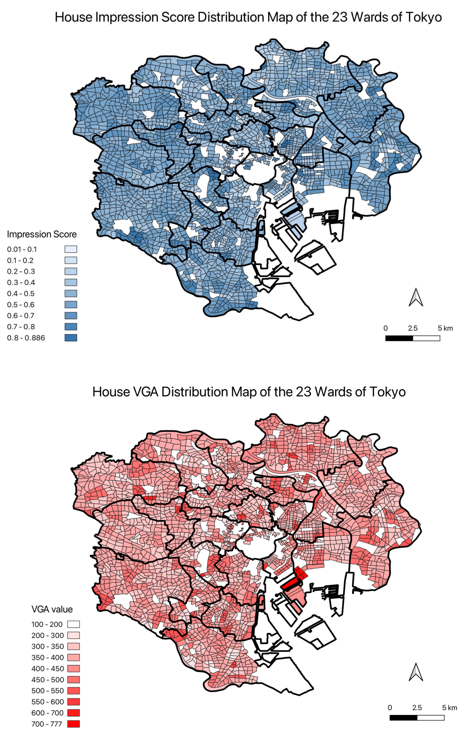

Figure 6: Geographic distribution analysis across Tokyo's 23 special wards: impression scores (top) and VGA mean values (bottom).

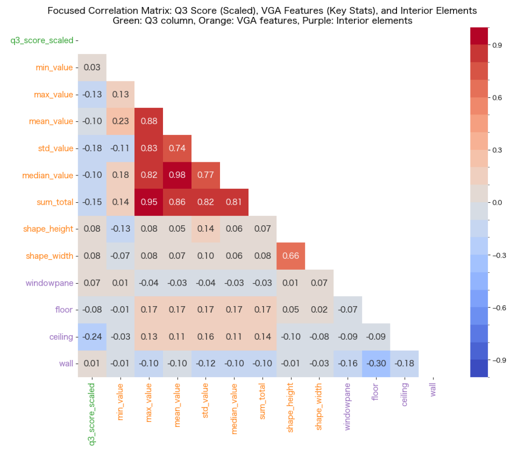

Figure 7: Correlation matrix between Q3 impression scores, VGA metrics (mean, std, min, max), and interior element fractions

Interior Element	Correlation	Sample Size (n)
Windowpane	0.0905	3012
Floor	-0.1577	4439
Ceiling	-0.3043	3559
Wall	0.0275	4430

Table 1: Correlations between interior element ratios and Q3 impression scores after 75% filtering

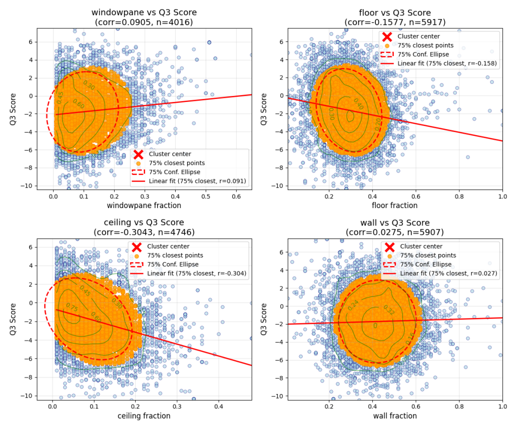

Figure 8: Cluster analysis of impression scores vs. interior element ratios

4.2 Regression Analysis Against Rent Cost

We performed regression analysis to predict rent cost using openness attributes. XGBoost achieved the best fit with acceptable R^2 scores. Analysis reveals that standard deviation and total sum of VGA values are the most important factors for rent prediction, thus there is a possibility that larger rooms with more spatial subdivisions tend to increase rental prices in a human understandable way. The distance towards the top 2 stations falls behind the above factors slightly. Additionally, our analysis reveals that user impression scores contribute to rent price trends moderately, suggesting a meaningful relationship between subjective spatial perception and market valuation. This indicates that user-perceived spatial quality aligns with economic value assessment in the rental market. As shown in Figure 9, training and testing results demonstrate reasonable performance. Residual analysis reveals weak bias in the fit, indicating the results are trustworthy.

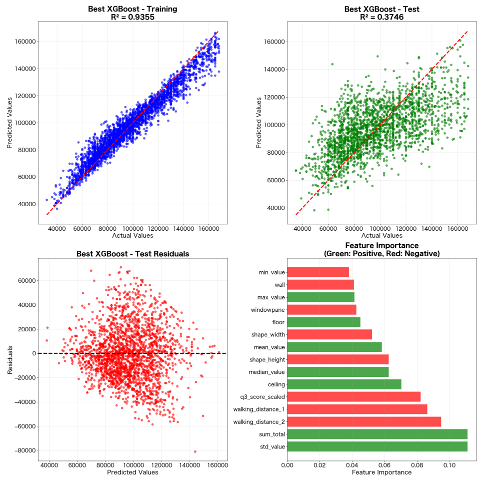

Figure 9: XGBoost regression results for rent prediction using openness attributes

4.3 Principal Component Analysis

To better understand the relationships between spatial features, we performed Principal Component Analysis (PCA) using VGA mean values, ceiling, wall, floor, and window visual fractions as input features. The analysis reveals the underlying structure of spatial characteristics in our dataset. The PCA plots are shown as Figure 10.

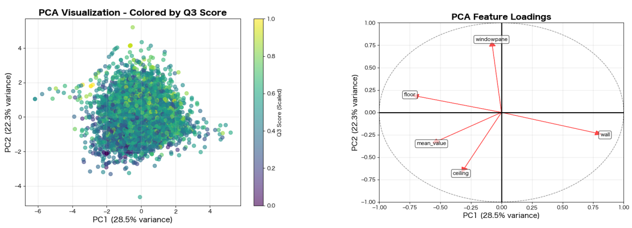

Figure 10: PCA distribution and top 2 components with raw feature axis

Table 2 presents the explained variance for each principal component. The first two components capture approximately 51% of the total variance.

Comp.	Eigenvalue	Var. %	Cum. %
PC1	1.42	28.46	28.46
PC2	1.12	22.33	50.80
PC3	1.07	21.36	72.16
PC4	0.88	17.57	89.73
PC5	0.51	10.27	100.00

Table 2: Principal component variance explanation

Table 3 shows the feature loadings for all principal components. PC1 is primarily dominated by wall fractions (0.65) opposing floor visibility (-0.57). PC2 captures the balance between natural lighting (window: 0.70) and ceiling prominence (-0.59).

Name	PC1	PC2	PC3	PC4	PC5
VGA. Avg	-0.44	-0.30	-0.15	0.82	-0.15
Window	-0.06	0.70	0.50	0.37	0.35
Floor	-0.57	0.17	-0.53	-0.23	0.56
Ceiling	-0.25	-0.59	0.62	-0.16	0.43
Wall	0.65	-0.22	-0.27	0.32	0.60

Table 3: Principal component feature loadings

5 Conclusions

This research presents a framework for analyzing spatial openness in rental housing through automated madori interpretation, visibility graph analysis, and interior semantic segmentation. Our methodology quantifies subjective spatial qualities using visual data from rental platforms. Key findings reveal weak correlations between VGA metrics and user impression scores, with excessive floor/ceiling visibility negatively impacting quality perceptions. PCA analysis reveals that the first two components capture 51% of variance, with PC1 dominated by wall-floor opposition and PC2 capturing natural lighting versus ceiling prominence. Regression analysis demonstrates that VGA standard deviation and room size are the most important factors for rent prediction. Spatial openness perception is impacted by the complex interactions between these factors. Our results identify their distinct contributions, revealing how visual effects and VGA metrics shape openness evaluation. This research provides practical tools for data-driven spatial openness analysis, which is a step towards openness-oriented housing design.