DATA-DRIVEN ANALYSIS OF SPATIAL PATTERNS THROUGH LARGE-SCALE DATASETS OF BUILDING FLOOR PLAN

HOYOUNG MAENG[1] and KYUNG HOON HYUN[2]
[1,2]*Department of Interior Architecture Design, Hanyang University*
[1,2]*{roach555|hoonhello}@hanyang.ac.kr*

Abstract. This paper introduces a unique quantitative analysis method and results that are differentiated from those in existing studies. We analyzed five types of information in floor plan images: the silhouette, number of rooms, room area, and direct and indirect room connectivity. Furthermore, the analysis used a large-scale apartment unit dataset consisting of 33,892 units. We present convincing and objective spatial pattern analysis results of Korean apartments by quantitatively analyzing a large-scale dataset. It is expected that the analysis results will clarify the characteristics of the residential environment of Korean apartments. The results suggest that changes in lifestyles lead to the modularization of bedrooms, increased numbers of private bathrooms and balconies with corridors as junctions, and the diversification of room layouts.

Keywords. Floor Plan Analysis; Design Quantification; Residential Layout; Spatial Pattern Analysis; Semantic Fingerprint.

1. Introduction

According to Statistics Korea (2020), apartments account for 51.1% of residential space in South Korea. Researchers have been actively investigating residential styles in Korea to understand apartment floor plan features (Bae et al. 2001; Choi 2003), to identify the effects of apartment units on lifestyles (Choi et al. 2004), and to infer residents‘ social and cultural structures from their spatial structures (Choi et al. 2004). Researchers have identified that Korean apartments’ floor plans are undergoing diversification over time, mainly in terms of the number of rooms or the functions of the rooms (Bae et al. 2001; Choi, 2003; Choi et al., 2004). However, previous studies were limited in terms of the number of datasets and were thus unable to show apartment trends over the longer term. Moreover, they organized their datasets manually. Thus, this study aims to propose a computer-aided data-driven analysis to provide a better understanding of the spatial patterns of apartments and to generate accurate and reliable results based on large-scale datasets. To do this, our study analyzes apartment units from the following three perspectives: 1) A morphological analysis of apartment floor plans (silhouette); 2) An analysis of the number of rooms, sizes of rooms and their types (room components); and 3) an analysis of floor plan layouts (room connectivity). Custom-made software was used to compile the dataset from an

PROJECTIONS, Proceedings of the 26th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2021, Volume 1, 301-310. © 2021 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

online real estate database (r114.com) consisting of floor plan images and 28 types of textual information. We conducted quantitative analyses of the silhouette, room components, and room connectivity data and examined the spatial structures of apartment floor plans over time.

2. Related Works

2.1. ANALYSIS OF UNIT FLOOR PLANS IN KOREAN APARTMENTS

Previous studies researched only partial spatial structures of Korean apartments. Bae et al. (2001) and Choi (2003) manually investigated 129 and 128 floor plans in the early 2000s in an effort to determine numbers of rooms and typical room layouts. Other researchers made attempts to quantify the connectivity of floor plans with the space syntax. Choi et al. (2004) collected 2,833 floor plans from 1966 to 2002 and quantified the minimum path from one room to another. Choi et al. (2014) compared the concept of spatial centrality with 60 floor plans. Given these studies, we can find three main aspects for a better analysis of apartment unit plans. The first is a larger dataset. We collected the most extensive data on 42,984 apartments and 5,890 floor plan images from 1970 to 2020 in Seoul. The second aspect is a perspective from a more thorough analysis that ensures an advanced understanding of the spatial structure of floor plans. The third aspect is automation of the recognition and data extraction process. Thus, we focused on these three developments to propose a novel method.

2.2. COMPUTER-AIDED QUANTITATIVE ANALYSIS OF SPATIAL STRUCTURES

Researchers that invented recognition systems for floor plan components were investigated. Ahmed et al. (2014) classified the features of floor plans using line thicknesses, recognized walls and openings, and extracted the semantic structure. However, his method is inadequate if applied to floor plans with various wall thicknesses. Liu et al. (2017) integrated a convolutional neural network and integer programming to distinguish floor plans' components, showing an accuracy rate of 94.7%. Given that the machine-learning dataset of Liu et al. (2017) mainly consists of square-oriented floor plans, there is a limit to interpreting curved or diagonal floor plans. The deep multi-task neural network developed by Zeng et al. (2019) uses an advanced method to recognize walls and types of rooms separately. This network learned diverse residential floor plans, specifically 1,047 with various wall thicknesses, including those with curved and diagonal wall lines, providing an improved floor plan recognition rate compared to that by Liu et al. (2017). In our study, to improve on the manual floor plans analysis methods in previous Korean apartment studies, a novel method was developed considering the aforementioned literature.

3. Methodology

This paragraph introduces five types of data extracted from the floor plan images. The first is the floor plan's silhouette, as it is possible to observe how the appearance of the apartment unit shape has changed. The second and third types are the number and area of the rooms. The room in the unit is a general index which can be used to discover changes in floor plan and is therefore the most fundamentally investigated aspect in the existing literature. The fourth and fifth types are direct and indirect connectivity based on the semantic fingerprint concept of Langenhan & Petzold (2010). We aimed to determine changes in the relationship between rooms and the layout of the floor plan over time. Various types of images were collected. Considering the investigation subjects, we excluded 3D renderings that were not appropriate floor plans and multi-floor units that were not in the standard Korean apartment format. Consequently, we studied 5,355 2D floor plan images and 33,892 instances of metadata to generate the five types of information. A custom-developed software package referring to the Grid-Based (GB) descriptor by Rodrigues et al. (2017), the floor plan recognition network of Zeng et al. (2019), OpenCV, and the NetworkX libraries were used for the information extraction process. Our system has the following differences with these earlier methods. Rodrigues et al. (2017) conducted a study to find a descriptor with high accuracy in cluster classification by comparing digitizing floor plans' silhouettes with labeled data. Zeng et al. (2019) focused on devising advanced machine learning networks capable of accurately recognizing the walls, doors, and types of rooms in floor plans. We aimed to extract various types of representable information from floor plans compared to those in previous studies. This approach is novel in that it allows more than 5000 floor plan silhouettes to be classified into clusters of an appropriate number and facilitates the quantification of the diversification, relationships, and connectivity aspects of rooms.

3.1. PREPROCESS

It was barely possible to extract information with a single image analysis system due to the various styles that existed. Therefore, to unify all of the images into the same format, unnecessary elements were manually trimmed. Simultaneously, 1,297 of 5,890 floor plans with different interior textures were modified into one drawing style with five color schemes for the living room, bedroom, toilet, balcony, and entrance (Modified Plan in Figure 1). After this alteration, we converted the images into pixel images in the 'Boundary Plans, format consisting of walls, doors, and windows, using a method developed by Zeng et al. (2019) (Figure 1). It processed input floor plans into three outputs - wall lines, connection parts (walls and windows), and room pixels - with different colors for each room type. Because the room types that were extracted by Zeng et al. (2019)'s method were significantly different relative to actual room types, we used only the wall lines and connection parts to generate Boundary Plans, in which black pixels represent wall lines and green pixels represent doors and windows. As illustrated in Figure 1, we unified the drawing style into a Modified Plan after removing external sections. We created Boundary Plans that express only the walls and openings from the Modified Plans in the preprocessing step, after which we performed the extraction process with our novel system.

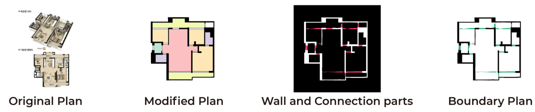


Figure 1. Preprocess steps.

3.2. SILHOUETTE EXTRACTION OF FLOOR PLANS AND SILHOUETTE CLUSTERING

Each Modified Plan image was changed into grayscale with white inside and a black background to obtain the corresponding silhouette. Because the transformed images contain large-scale pixel information of different sizes, the next step was to create a reduced binary image of an identical size. We utilized GB descriptor by Rodrigues et al. (2017) to do this (Figure 2). GB descriptor created a square with the floor plan located in the center and drew 28x28 equally spaced grids by dividing the square. If a grid's center is located inside the floor plan, the value is 1, and if the center is located outside the floor plan, the value is 0. It was necessary to classify silhouette data to check the morphological changes of apartment units over time. However, the floor plans' directions were not unified, meaning that even the same floor plans could be classified into different clusters. Using only GB descriptor without any processing, identical L-shaped floor plans simply rotated 90 degrees will be recognized as two different floor plans. Thus, it was necessary to acknowledge the two as the same L-shaped floor plan to find only the shape changes of the silhouette. Thus, floor plans undergo two processes so that we can standardize the directions. First, we rotate floor plans so that the long side is always in the vertical position. Second, we rotate the floor plans so that areas with more black pixels, i.e., empty segments, are located at the top. Afterward, the standardized GB described 28x28 matrices were transformed into 784 binary data in one-dimensional matrices. Subsequently, we classified clusters by means of the K-means method. The silhouette score and elbow method helped determine the optimized number of clusters.

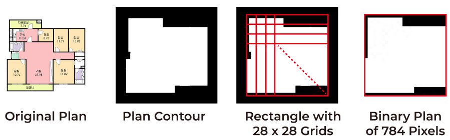


Figure 2. GB descriptor execution of the original floor plan.

3.3. ROOM COMPONENTS DETECTION

Boundary Plans are pixel images drawn into black wall lines and green-colored openings, and each room has individual contours based on the doors and walls. Our system selects the Modified Plan's color located in the centroid of each room contour of the Boundary Plan to assign room types to the contours in the Boundary Plan. After recognizing the room type, it colors each contour of the Boundary

Plan with the corresponding color. For example, if the center of a contour is orange in the Modified Plan, the contour will then be colored orange (Figure 3). However, the network devised by Zeng et al. (2019) could not draw the doors well compared to the walls, leading to errors. Occasionally, in poorly generated Boundary Plans, two or more rooms were recognized as one contour because the door was incomplete. In such cases, error detection was possible when two or more room centroids designate an identical contour. Our system searched for problematic door pixels close to an overlapped contour and extended these pixels in the direction of the longer side until the line hit the closest wall to become a proper door. The repair process was repeated until all of the contours marked only one room in each case. The system was completed, but some errors cases had multiple doors that were too close such that the system recognized two or more doors as one. We filtered these out according to their abnormal ratio for a door from the colored Boundary Plan and manually altered the doors given that our system could not identify overlapped doors separately. This left only valid walls, doors, and rooms in the colored Boundary Plans. Our system then stored the number and area of rooms depending on the number of colored contours and pixels.

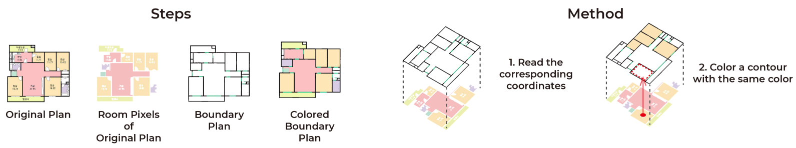


Figure 3. Colored Boundary plan generation steps and method.

3.4. ROOM CONNECTIVITY ANALYSIS

Room connectivity was analyzed using colored Boundary Plans. The system recognized the doors by color and selected two coordinates located at a certain distance in a direction parallel to the short side of the doors. If coordinates at both ends are located in different rooms, this indicates that the two rooms are connected through the door; i.e., direct connectivity is established. Meanwhile, there could be a colorless (empty) room between the rooms. The designated room types had their color assigned and, the colorless room in this case is either a closet or an extra room, such a walk-in closet with no assigned color. The former did not influence the connectivity because it was open to a single room. However, the latter was open to multiple rooms and acted as a corridor between rooms. We defined the corridor as a new room type for the latter cases. If a corridor candidate (colorless room) has more than two doors connected to other rooms, the system colored the candidate in cyan (Figure 4). Our system newly stored the number and area of corridors in the room components data. Because floor plans were drawn in different standards, in some cases it was necessary manually to modify direct connectivity after our system filtered out possible errors. Depending on the style, floor plans had a shower booth and bathroom connected through a door or without a door. In such a case, the system recognizes these elements as separate bathrooms. Considering this, we removed the shower booth doors after checking visually whether two or more bathrooms were actually connected. We were thus able to generate correct direct connectivity and colored Boundary Plans with added corridors. With regard to indirect connectivity, room pixels in the colored Boundary Plans were dilated to exceed the wall thickness, and the overlapping rooms were indirectly connected (Figure 4).

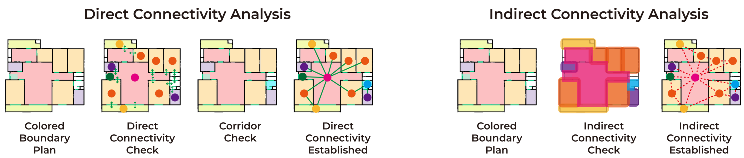


Figure 4. Direct and indirect connectivity analyses.

After the generation of the direct and indirect connectivity graphs, the degree centrality, the betweenness centrality, and the closeness centrality were calculated. The definitions of and equations for each centrality are as follows: degree centrality refers to how many rooms are connected to each room ($C_D = \sum E_w$; E_w: the set of edge connected to the node), closeness centrality refers to the closeness of all rooms in the floor plan to each other ($C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)}$; n: the number of nodes, d(v, u): the shortest path from v to u), and betweenness centrality refers to how many times the shortest paths pass through each room ($C_B(v) = \sum_{s,t \in V} \frac{(s,t|v)}{(s,t)}$; V: the set of nodes, σ (s, t): the number of shortest(s, t) -paths, σ (s, t|v): the number of shortest (s, t) -paths passing through some vertex v other than (s, t)).

4. Implementation Results

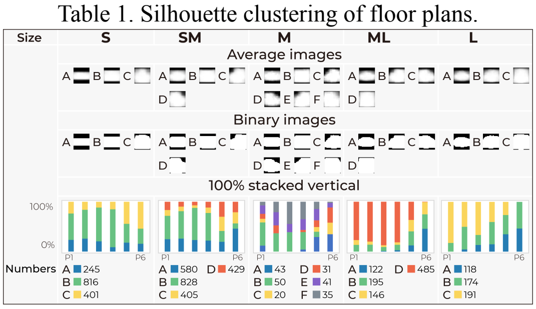
4.1. CLASSIFICATION OF FLOOR PLANS

We classified floor plans into five sizes by referring to Seoul's apartment-application deposit criteria. These are denoted here as Small (S) (-60m2), Small-medium (SM) (60-85m2), Medium (M) (85-102m2), Medium-large (ML) (102-135m2), and Large (L) (135m2 and larger). The numbers of floor plans by size were as follows: S: 1,462, SM: 2242, M: 220, ML: 948, and L: 483. The move-in year of the apartment units investigated in this study ranged from 1976 to 2020. Choi et al. (2004) noted that data reorganization by period based on events that affected the apartments is a very effective method when discussing socio-cultural phenomena via a spatial analysis. We reconstructed the years into several periods by referring to Choi et al. (2004). These are Period 1 (P1) (1976-1985), Period 2 (P2) (1986-1989), Period 3 (P3) (1990-1997), Period 4 (P4) (1998-2002), Period 5 (P5) (2003-2010), and Period 6 (P6) (2011-2020).

4.2. SILHOUETTE CLUSTERING

We grouped the silhouettes by size and investigated how the shapes of the units changed over time. The optimal numbers of clusterings for each size were S: 3; SM: 4; M: 6; ML: 4; and L: 3. Table 1 presents representative images of two versions of each cluster. The first image shows the average value obtained by adding the 28x28 binary silhouettes and dividing this value by the number of images in the cluster. The second representative image is a binary image of the first image. It can be inferred that the apartment unit's silhouette is standardized in several typical shapes, regardless of the size. Rectangle, rounded rectangle, L, mirrored L, and square shapes appeared in common, and the proportion of changes over time appeared to be similar. As a result, silhouettes varied the most in P5 and P6, which means that some diversification occurred in the 2000s and 2010s.

Table 1. Silhouette clustering of floor plans.



4.3. ANALYSIS OF THE NUMBER AND AREA OF ROOMS

Figure 5 illustrates the average of each room area's sum and the number of rooms excluding the living room according to the size over the period. The number of rooms increased to P5 overall, and all values decreased in P6. On average, the number of rooms increased from 7.3 for P1 to 9.2 for P6. We expected that the increase in the number of rooms would reduce the area of the rooms. However, the room area scarcely changed at most sizes, and L even increased up to P5. The fact that the room area did not decrease implied a decrease in the living room area. Next, we examined in more detail the room types. The numbers of bedrooms changed individually for each type; the number increased in S; there was little change in SM, M, and ML; and it decreased in L. The number of bathrooms for the L size increased from 2 to 2.5, and the number also increased in other sizes, settling at 2. Corridors showed slight increases in the number in general. The number of balconies increased up to P5, apart from the L size, and this number eventually decreased in P6. The areas of the bedrooms, bathrooms, and balconies were inversely proportional to their numbers, but both values decreased for the P6 balconies, representing less preference for balconies in the 2010s. Only the number of corridors and the corresponding area increased simultaneously, and the area expanded remarkably from P3, indicating growing demand for this element.

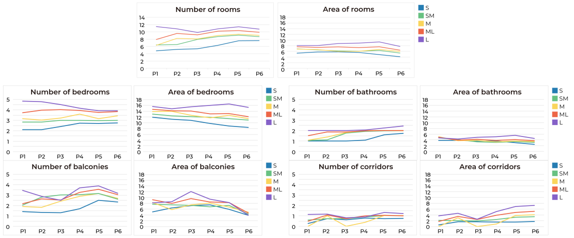


Figure 5. Number and area of rooms by period.

4.4. ANALYSIS OF DIRECT CONNECTIVITIES

In this section, we analyzed the direct connectivity of rooms. The average graphs for rooms shown in Figure 6 illustrate the average for the degree, closeness, and betweenness centrality of the rooms by period, excluding the living room in each case. With regard to the degree, even if there is a difference in P1, it converges to 1.5 at P6. Because closeness is a value that declines when the number of rooms or the depth of the space increase, S is largest and L is smallest. Betweenness increased the most for the SM size, and the other sizes similarly showed gradual increases. Next, we studied the centrality of each room in more detail (Figure 6). The degree and betweenness values for bedroom increased but closeness decreased, suggesting that bedrooms are becoming modularized and individualized over time due to the following reasons. First, while the bedroom's degree value increased, the numbers of balconies and bathrooms both increased (Figure 5) with decreases in their degrees. The first reason implies that added balconies and bathrooms over time are open to bedrooms only. Second, the betweenness value for bedrooms increased, indicating that bedrooms became an intermediate node between other rooms. These situations indicate increases in the numbers of private balconies and bathrooms. Except for a sharp decrease in P3, the corridor is the only room type for which the closeness value increased, as corridors serve as the bridge between bedrooms and private balconies or bathrooms. In rooms other than a corridor, the closeness value decreased due to the presence of modularized bedrooms, which means the minimum distance to other rooms grew. However, the betweenness value for corridors increased, as these elements served as the main junction in the modularized bedrooms.

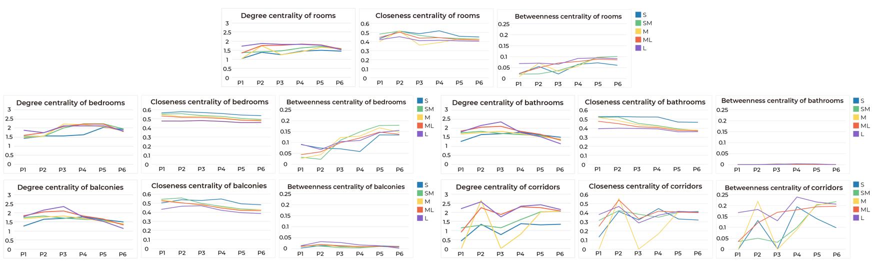


Figure 6. Degree centrality, closeness centrality and betweenness centrality of rooms by period

4.5. ANALYSIS OF ROOM LAYOUTS

In this section, we analyzed the indirect connectivity. Indirect connectivity is a component indicating the shape of the floor plan, similar to a silhouette (Section 4.2), as it connects nearby rooms. However, unlike silhouettes, which are represented by pixels, the layout information is identified by converting a floor plan into a network by examining the indirect connectivity between adjacent rooms. We utilized the graph edit distance (GED) to indicate how floor plans have become more diversified with Python's GMatch4Py library. GED was calculated by pairing two of the floor plans with the same size classification, and the average value was illustrated over time (Figure 7). As a result, we found that GED increased over time, implying that the diversification of the layout became more active over time. The larger sizes had larger GED, implying that a large area allows the diversification of the layout. It was able to describe the change of layouts in more specific values than the silhouette clusters in Table 1.

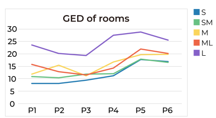


Figure 7. Graph edit distance (GED) values of rooms by period.

5. Discussion and Conclusion

P4 contains data from 1998 to 2002, a period of "The 1997 Asian Financial Crisis" in Korea. During the restoration period, P5, various apartments were designed to increase the construction volume with pump-priming by, for instance, financial support and deregulation. Furthermore, this socio-cultural background explains the peculiarity of the analyzed data in P5. We find a vigorous improvement in the number of cluster types, the number and area of rooms, modularization of bedrooms, and the diversity of layouts. Our method's limitation stems from the fact that floor plans available online vary in terms of the level of detail, the residential type, and the expression style. For example, whether the furniture is represented in the floor plan; the use of a non-constant wall thickness; and definitions of the range of room types, such as bedrooms, master bedrooms, bathrooms, shower booths, balconies, utility rooms, corridors, and walk-in closets depends on the architect or reader. Therefore, it was challenging to recognize floor plans completely without manual manipulation, and it was inevitable that we needed manual manipulation between the processes. A limitation in the results is that there is a significant difference in the number of floor plans for every period and size. Thus, the unevenness in the number of data instances led to the unintentional stressing of a specific situation. In P3 for the M size, there are relatively few samples, and corridors are absent in all 18 units. Accordingly, the dropped values in Figures 5 and 6 stood out. In future studies, data on more apartment units will be collected to secure data diversity.

This study quantitatively examined the historical changes and diversification of Korean apartment units from various perspectives through large datasets.

There are several important results. First, the silhouettes of the apartments were similar for all sizes, and the distribution by period changed similarly. Second, the number of rooms increased, and the area was maintained. Third, individualization and modularization of bedrooms were noted. Fourth, corridors are both evidence and a catalyst of the third finding above, and the preference for corridors increased over time. Fifth, the degree of layout diversity increased up to P5 and then declined. This study's crucial contribution is that employing large datasets for floor plan investigations quantitatively to analyze spatial patterns in morphological, numerical, and semantic criteria is important. Further research will explore more specific correlations between current values and conduct rigorous analyses with regions, prices, and nearby traffic conditions and convenience facilities.

Acknowledgement

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5A8034285).

References

"Real Estate 114" : 2019. Available from <https://www.r114.com/> (accessed 6th November 2019).

"Statistics Korea.: 2020, 2019 Population and Housing Census" : 2020. Available from <https ://kostat.go.kr/portal/korea/kor_nw/1/2/2/index.board?bmode=read&bSeq=&aSeq=384690&pageNo=1&rowNum=10&navCount=10&currPg=&searchInfo=&sTarget=title&sTxt= > (accessed 25th September 2020).

Ahmed, S., Weber, M., Liwicki, M., Langenhan, C., Dengel, A. and Petzold, F.: 2014, Automatic analysis and sketch-based retrieval of architectural floor plans, *Pattern Recognition Letters*, **35**, 91-100.

Bae, J.M., Jung, Y.S. and Yoon, J.S.: 2001, A Study on Characteristics of Apartment Housing Unit Plans, *Journal of the Korean Housing Association*, **12**, 1-12.

Choi, E.H.: 2003, A Study on the Characteristics of Unit Plan Composition according to Sizes in Public Housing : Focused on the Residential Apartments built after 1999 and occupied from 2001, *Journal of the Korean Institute of Interior Design*, **38**, 134-142.

Choi, J.P., Cho, H.K., Park, I.S. and Park, Y.S.: 2004, A Spatial Analysis of the Apartment Unit Plans from 1966 to 2002 in Seoul, *OURNAL OF THE ARCHITECTURAL INSTITUTE OF KOREA Planning & Design*, **20**, 153-162.

Choi, J., Kim, Y., Kang, J. and Choi, Y.: 2014, Comparative Analysis of the Spatial Structure of Apartment Unit Plans in Asia - Apartments in Korea, Vietnam, and Kazakhstan -, *Journal of Asian Architecture and Building Engineering*, **13**, 563-569.

Langenhan, C. and Petzold, F.: 2010, The fingerprint of architecture-sketch-based design methods for researching building layouts through the semantic fingerprinting of floor plans, *International electronic scientific-educational journal: Architecture and Modern Information Technologies*, **4**, 1-8.

Liu, C. and Kohli, P.: 2017, Raster-to-vector: Revisiting floorplan transformation, *Proceedings of the IEEE International Conference on Computer Vision (ICCV) 2017*.

Rodrigues, E., Rodrigues, D., Sampayo, M., Gaspar, A., Gomes, A. and Antunes, C.: 2017, Clustering of architectural floor plans: A comparison of shape representations, *Automation in Construction*, **80**, 48-65.

Zeng, Z., Li, X., Yu, Y. and Fu, C.: 2019, Deep Floor Plan Recognition Using a Multi-Task Network With Room-Boundary-Guided Attention, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) 2019*.