Portfolio of Design and Research Work

Dr, Taro Narahara
Associate Professor

New Jersey Institute of Technology
University Heights
Newark, NJ 07102-1982
Email: taronarahara@gmail.com

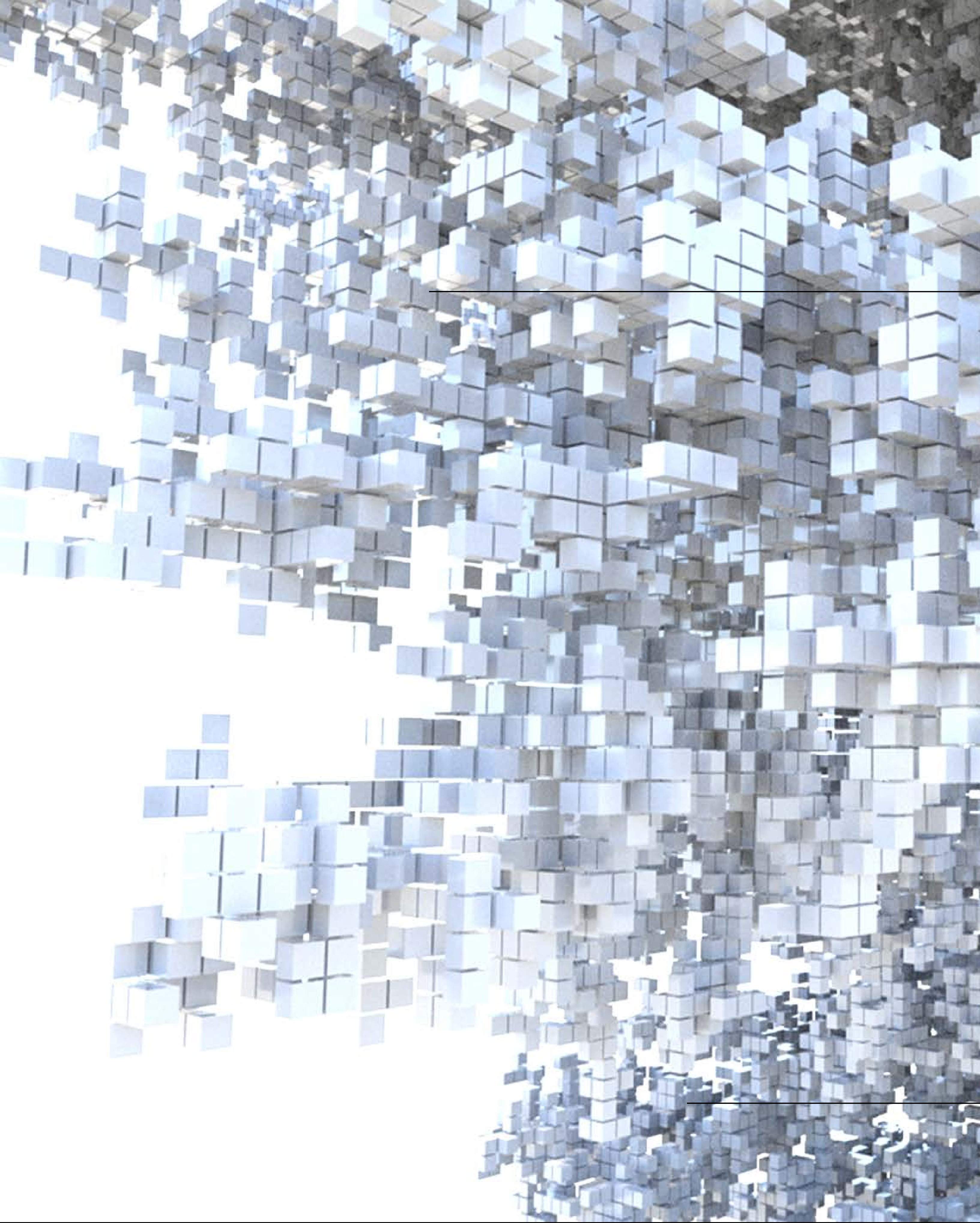


Table of Contents:

Research Work Samples

Professional Experience

Teaching Projects and Achievements

PROJECT VIDEO

Subjective Functionality and Comfort Prediction for Apartment Floor Plans and Its Application to Intuitive Online Property Search

Taro Narahara, *Member, IEEE*, and Toshihiko Yamasaki, *Member, IEEE*

***Abstract*—This paper presents a new user experience for online apartment search using functionality and comfort as query items. Specifically, it has three technical contributions. First, we present a new dataset on the perceived functionality and comfort scores of residential floor plans using nine question statements about the level of comfort, openness, privacy, etc. Second, we propose an algorithm to predict the scores from the floor plan images. Lastly, we implement a new apartment search system and conduct a large-scale usability study using crowdsourcing. The experimental results show that our apartment search system can provide a better user experience. To the best of our knowledge, this is the first work to propose a highly accurate machine learning model for predicting the subjective functionality and comfort of apartments.**

***Index Terms*—Attractiveness prediction, crowdsourcing, graph analysis, real estate floor plans.**

I. Introduction

IN RECENT years, the real estate industry has been showing increasing interest in applying machine learning-assisted solutions such as price prediction [1], [2], [3], [4] and apartment-searching [5] tools. Some online platforms can help users search for properties by specifying metadata, such as the type of apartment and room size. However, many users inspect floor plans based on more intuitive sensory impressions, such as *living comfort*, *openness*, and *privacy*. This makes it difficult to estimate the perceptive values of apartments through any currently available retrieval system, as there are no quantifiable data that represent such subjective characteristics of apartments. Moreover, apartment properties listed with the same size and type in their metadata (e.g., two-bedroom apartments) could feature different room arrangements, which will have a significant impact on their functionality and overall livability. Therefore, further understanding the relationships between floor plan images and structured data, including the connectivity of rooms and metadata, could improve the user experience on real estate search platforms.

Information on floor plans has been widely adopted by users to evaluate the values of properties over the years, as can be seen in many real estate portal sites today. A floor plan image of an apartment includes various room types, room sizes, and connections and spatial layouts of the rooms. According to a customer survey conducted in Japan[1], a floor plan was found to be among the top five priorities for customers during their apartment search. Moreover, it was found that customers are very reluctant to compromise on their preferred floor plans and are often willing to invest more for their pursuits. Essential elements that largely influence functional, environmental, and some perceptive characteristics of apartments, such as locations of walls, columns, windows, and wet areas, are already set in the floor plans, and cannot be changed no matter the finish materials or furniture used. Although we can estimate the subjective quality of apartments, such as *living comfort*, simply by looking at the floor plan images, no related work nor dataset has been reported for such a task.

In this study, therefore, we first constructed a new dataset that contains 1,000 floor plan images (hereafter, "dataset A"). Each image has a subjective score from nine perspectives relating to perceived quality and functionality of the apartments. Examples of the generated dataset are shown in Figure 1. Based on this dataset, we developed a multimodal neural network-based framework to predict subjective apartment scores via their floor plan images, graph representations, and various metadata. The experimental results showed that we can predict the scores with a correlation coefficient of 0.701 on average. This is relatively high, considering that they are all subjective scores. Our study shows that the baseline model, which uses features based on images alone, has a much lower average correlation coefficient of 0.491, even using ResNet50, the state-of-the-art network for image recognition tasks.

Manuscript received 15 March 2022; revised 3 July 2022 and 7 August 2022; accepted 25 September 2022. Date of publication 27 October 2022; date of current version 1 November 2023. This work was supported by the CREI Research Project of The University of Tokyo. The Associate Editor coordinating the review of this manuscript and approving it for publication was Prof. Abderrahim Benslimane. *(Corresponding author: Taro Narahara.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB) of New Jersey Institute of Technology (F466-20) and The University of Tokyo (UT-IST-RE190927-1).

Taro Narahara is with the Hillier College of Architecture and Design (HCAD), New Jersey Institute of Technology (NJIT), Newark, NJ 07102 USA (e-mail: narahara@njit.edu).

Toshihiko Yamasaki is with the Department of Information and Communication Engineering, Graduate School of Information Science and Technology, The University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan (e-mail: yamasaki@cvm.t.u-tokyo.ac.jp).

This article has supplementary downloadable material available at https://doi.org/10.1109/TMM.2022.3214072, provided by the authors.

Digital Object Identifier 10.1109/TMM.2022.3214072

[1][Online]. Available:https://suumo.jp/article/oyakudachi/oyaku/chintai/fr_data/hikkoshi-sumikae2017/, accessed 09/16/2020

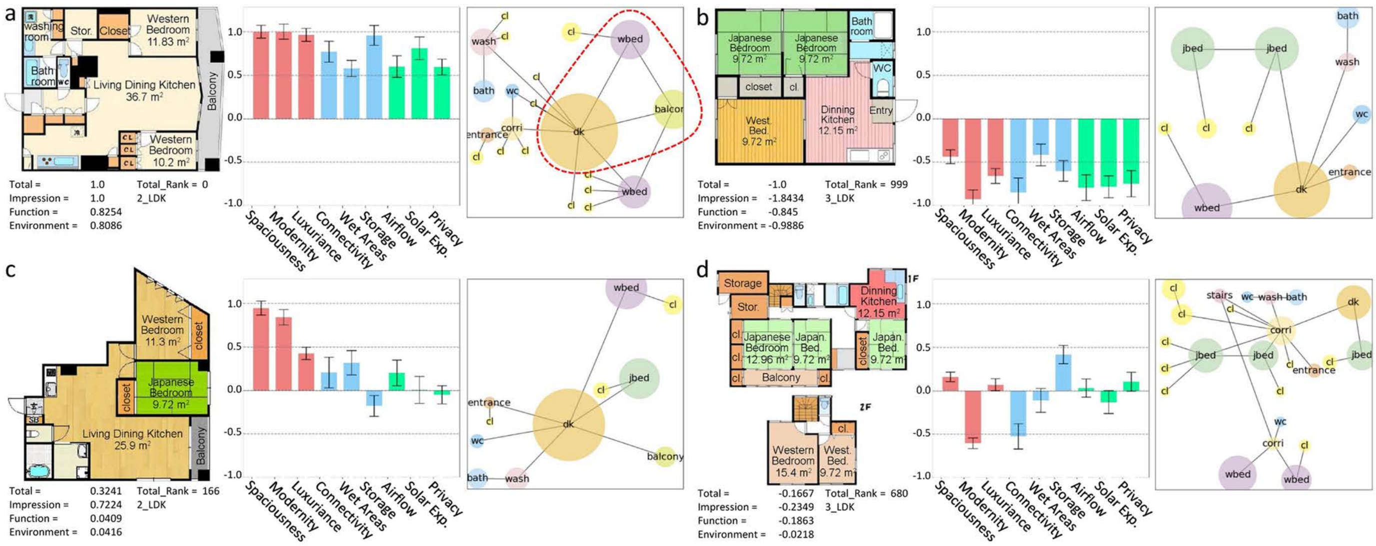


Fig. 1. Examples of our functionality evaluation dataset using real estate floor plan images. Each example shows a floor plan image, a bar graph for nine evaluation measures, and a graph from left to right. (a) Highest scoring floor plan; (b) lowest scoring floor plan; (c) floor plan that scored high on modernity; (d) floor plan that scored low on modernity.

The contributions of this paper can be summarized as follows.

- We created and analyzed a large-scale dataset of subjective evaluations of both perceived quality and functionality of real estate floor plan images using crowdsourcing. (i.e., 3,128 participants rated 1,000 floor plans based on nine subjective criteria.)
- Our proposed prediction model, which extracts features from floor plan images and their graph structures, proved to be effective and highly accurate with average correlation coefficient of 0.701. We also developed a workflow to extract semantically segmented floor plan images, graphs, and graph-related features for our multimodal deep neural network model.
- Upon using a new set of floor plan images (with predicted scores by our model as a dataset), our proposed apartment search tool was found to provide a significantly better user experience than the baseline tool without the proposed feature.

This paper is organized as follows. In Section II, we report the results of our literature survey. Section III explains dataset creation. Section IV describes the methodology utilized in the study. Section V explains the experiments and their results. Section VI describes the usability study of our proposed apartment search tool. Section VII discusses the limitations of our approach, with concluding remarks included in Section VIII.

II. Related Works

A. Real Estate Tasks Using Property Images

Several researchers have worked on tasks related to real estate using property images. In [2], the authors tried to improve the accuracy of real estate price prediction by predicting the luxury levels of the rooms using the appearance and interior images of properties. Law et al. [6] showed that street view and satellite images are also helpful when predicting house prices. In [7], the researchers attempted to predict the construction age of the property by combining the predictions for each of its salient image patch, resulting in greater accuracy than human prediction.

Moreover, deep learning has been applied to real estate property images, and some studies have analyzed real estate images themselves. In [8], real estate images were classified into different types (e.g., bedroom, kitchen, living, and garden) by employing contrast-limited adaptive histogram equalization (CLAHE) and applying long short-term memory (LSTM) in both vertical and horizontal directions. Wang et al. [9] predicted which among two images of the same property is more attractive using a pairwise comparison network.

These works demonstrated the use of images in specific tasks that are related to the appearance of the property. In our task, to comprehensively represent the user experience in term of the perceived quality and functionality, we assumed that floor plan information with related metadata can provide more structured and detailed information on the property.

B. Real Estate Tasks Using Floor Plan

Several related works have been conducted on floor plan image analysis. Before the development of deep learning, some researchers manually and graphically analyzed floor plans using adjacency graphs (with rooms, corridors, and other features as labeled nodes) and used them to classify apartments into different types [10]. Takizawa et al. [11] analyzed the relationships among adjacency graph structures of apartment floor plans and their rental fees. They extracted subgraphs from the adjacency graphs and effectively estimated the rent from the presence and absence of common subgraphs. However, the cost of creating adjacency graphs by hand was very high.

Floor plan images have also proved to be useful for rent price prediction [12], [13], [14]. In [12], [13], it was demonstrated that conventional bag-of-features (BoFs) [15] has the potential to achieve lower-error prediction with smaller variance, although the contribution of BoFs was smaller than that of other apartment attribute information. In [14], image features were applied to hedonic price models [16] to predict the apartment rent price after controlling for locational and structural characteristics of an apartment.

Recently, machine learning has been applied to the analysis of floor plan images. Yamasaki et al. [5] used deep neural networks (DNNs) to conduct semantic segmentation of floor plan images. They further developed a method to systematically generate adjacency graphs of floor plans from the semantically segmented images. Takada et al. [17] utilized multi-task learning for floor plan images and retrieved similar floor plans to the query. The floor plan recognition was then applied to property recommendation [18], [19] and retrieval [20]. Additionally, a toolbox for converting floor plan images to a vector format was developed in [21]. Furthermore, generating floor plans using graphs [22], [23], panoramic images [24], or 3D scans [25], [26], [27], [28], [29], [30] is emerging. Generating furniture layouts using graphs was also discussed in [31].

C. Prediction of Subjective Scores of Multimedia Content

There are extensive surveys in the literature introducing quality assessment studies and methods using images [32] and videos [33], [34].

Assessing the perceived low-level quality of images [35], [36], [37], [38] and videos [39], [40], [41], [42] has been an important topic in multimedia. These works tried to predict the perceived quality when the quality of the multimedia content is somewhat degraded by low-level factors such as compression and noise.

Predicting higher-level subjective scores has also been an active research area. For instance, analysis of image aesthetics relates more to color usage, composition, etc., and not to low-level noise in the content [43], [44], [45], [46], [47], [48]. Sentiment and emotion classification and affective analysis constitute another direction of research for analyzing subjective evaluation of multimedia content. In this regard, many related works can be found in the literature for texts [49], [50], [51], images [52], [53], [54], [55], [56], [57], [58], speech [59], music [60], videos [53], [61], [62], [63], [64], and their combinations [65], [66], [67].

Skills and creativeness have also been research targets, including skill assessment [68], creativity [69], click through rate prediction for online advertisements [70], [71], presentation slide assessment [72], [73], and so on.

To the best of our knowledge, this is the first work on subjective evaluation of apartment floor plans. In the paper, we show that predicting the subjective functionality and comfort is possible, and we also demonstrate potential applications of such predictions.

III. Dataset Creation

A. Subjective Scores

In this study, we used crowdsourcing to create a new dataset based on subjective evaluation of real estate floor plan images through a set of statements that question their levels of comfort, openness, privacy, and other characteristics.

In total, 3,128 workers participated in this evaluation. We recruited 400 participants separately from ten groups: two genders over five age ranges (20-29, 30-39, 40-49, 50-59, and 60+) using a crowdsourcing service.

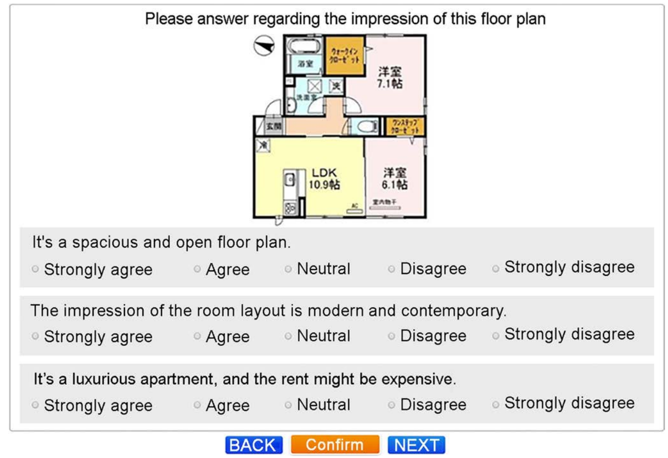


Fig. 2. A screenshot of the crowdsourcing page.

We used floor plan images of Japanese rental apartments from the “Home’s dataset” released by LIFULL Co., Ltd. with the cooperation of National Institute of Informatics,[2] which has been widely used as a general floor plan image dataset in the international research community [22], [74], [75], [76]. We randomly selected 1,000 floor plan images that included apartments with one, two, three, and four or more bedrooms in balanced proportions and prepared the following nine question statements for each image (questions Q1 to Q3 are about impressions, Q4 to Q6 are about functionality, and Q7 to Q9 are about environmental criteria):

- Q1 (Spaciousness): It is a spacious and open floor plan.
- Q2 (Modernity): The impression of the room layout is modern and contemporary.
- Q3 (Luxuriance): It is a luxurious apartment, and the rent might be expensive.
- Q4 (Connectivity): Connectivity, adjacency, and layout of rooms and circulation are efficient and look comfortable.
- Q5 (Wet Areas): The traffic paths for the kitchen, bathroom, and restroom are good.
- Q6 (Storage): Locations and sizes of storage are good.
- Q7 (Airflow): Airflow inside the floor plan is good.
- Q8 (Solar Exp.): Solar exposure of the floor plan is good.
- Q9 (Privacy): The arrangement and adjacency of rooms fully consider the privacy of each family member.

In Q9, participants were provided a family size to evaluate the privacy of the image. Each floor plan was evaluated based on a five-grade score on a scale of 1 (strongly disagree) to 5 (strongly agree) by participants. Only the floor plan images were shown to participants for the evaluation (Figure 2). Each participant was asked to evaluate 25 floor plan images. A task completion control was implemented, and the participants were required to answer all the assigned questions in order to complete the task; otherwise, their responses were not included in the study. After removing those from whom we did not receive all responses and also those who chose the same rating for all (i.e.,

[2]National Institute of Informatics (NII), https://www.nii.ac.jp/dsc/idr/lifull/homes.html (accessed: 05.05.2020)

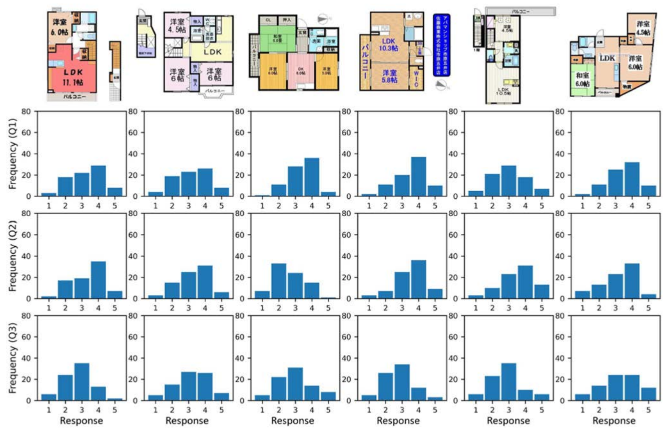


Fig. 3. Histograms of the responses (5-point Likert scale) for several examples of randomly selected floor plan images.

"straight-lining"), we obtained 3,128 valid participants' results out of the 4,000 participants that were recruited. The remaining 871 include those who either dropped out or were not included for the reasons stated above. Thus, each floor plan was evaluated by 78 to 80 individuals (i.e., 2 genders $\times$ 5 age ranges $\times$ 400 participants $\times$ 25 images / 1,000 total images = 100 participants per image; with approximately 20% drop outs, this coincides with 3128/4000 = 0.782). We verified that the histograms of responses by participants were approximately normally distributed from random selections of over 200 floor plans and that no isolated peaks with unusual values were found. Figure 3 shows some examples of histograms from our randomly selected floor plans. Each participant would have different mean and standard deviation in their evaluation scores, and therefore the scores were standardized before taking the average among the participants. We used the value obtained by subtracting the mean value from each raw score and dividing it by the standard deviation. The resulting scores were normalized between -1 and 1 for each question. The effect of this standardization is discussed in Section V-B. Figure 1 shows selected examples of floor plans with the corresponding results of the nine scores of subjective measure. The error bar represents the standard deviation. Next, for each floor plan, we averaged the scores from the nine questions. Since we averaged nine scores that originally came from different populations of scores with different distributions, we standardized the mean values from all floor plans and normalized them between -1 and 1; we called this value the "Total Score" of the floor plan. The Total Score represents the overall performance of each floor plan across all nine questions.

B. Segmentation and Graph Extraction

A floor plan can also be viewed as a graph with nodes representing rooms and with edges representing connections between them [5], [20], [22], [23]. Therefore, we extracted graphs from floor plan images and used them as features to predict the nine subjective scores defined above. This is a reasonable assumption because floor plan images are inspected for the connections among rooms and their adjacency.

In order to extract corresponding graphs automatically from floor plan images, we used the following two steps (Figure 4 top left). First, we prepared 3,800 new floor plan images (no overlap with dataset A), which we call dataset B, with their manually annotated segmented images using an online annotation tool (please see Figure 4). They were consistently color-coded and semantically segmented into the following 15 classes of elements: wall, western bedroom (wbed), Japanese bedroom (jbed), dining kitchen (dk), restroom (wc), bathroom (bath), washroom (wash), balcony (balc), entrance (ent), corridor (corri), stairs, closet (cl), door, window, and unknown elements that do not belong to any category (abbreviations in parentheses are used in figures representing graph nodes). Then, we prepared a segmentation prediction network based on the method introduced in [5] using an improved network architecture, DeepLab v3+ [77], [78], instead of fully convolutional networks. Using 3,040 images for training, 380 for validation, and 380 for testing, we trained the network using the segmented images as ground truth data. We automatically obtained 1,000 segmented images with subjective scores for our dataset by feeding dataset A into this pre-trained network using DeepLab v3+.

Second, we used the rule-based method to extract 1,000 graphs from the inferred segmented images using images in dataset A following the same procedure in [20]. Figure 5 shows examples of floor plan images, inferred segmented images, extracted graphs, and ground truth (GT) segmented images, which were manually annotated using an online annotation tool. We used 11 elements for nodes of the graphs, excluding the wall, door, window, and unknown elements from the above 15 elements. We added edges between two rooms only if they are directly accessible through a physical opening or door. Nodes were created by extracting regions representing rooms with a certain area in the inferred segmented images. The resulting dataset was used to determine whether the differences in graph structures influence the impression and functionality of apartments from subjective evaluations.

In Figure 5, it is noticeable that the inferred segmented images using the segmentation prediction network are noisy and slightly degraded compared to the GT segmented images. As a result, the extracted graphs from the inferred images are not as accurate as the grand truth graphs extracted from the GT segmented images. In Section V-B, we discuss how the imperfections in the use of inferred images and extracted graphs affects the prediction accuracy of our proposed model compared to the best-case using GT segmented images and GT graphs as input features as the upper bound performance. In fact, the average performance drop in Pearson correlation coefficient (PCC) is only 0.046 between the two cases, which is statistically insignificant (see Section V-B). It is thus acceptable to use the inferred segmented images instead of GT segmented images, which allows us to automatically generate all required features only from floor plan images without preparing manually annotated floor plan images for new data inputs.

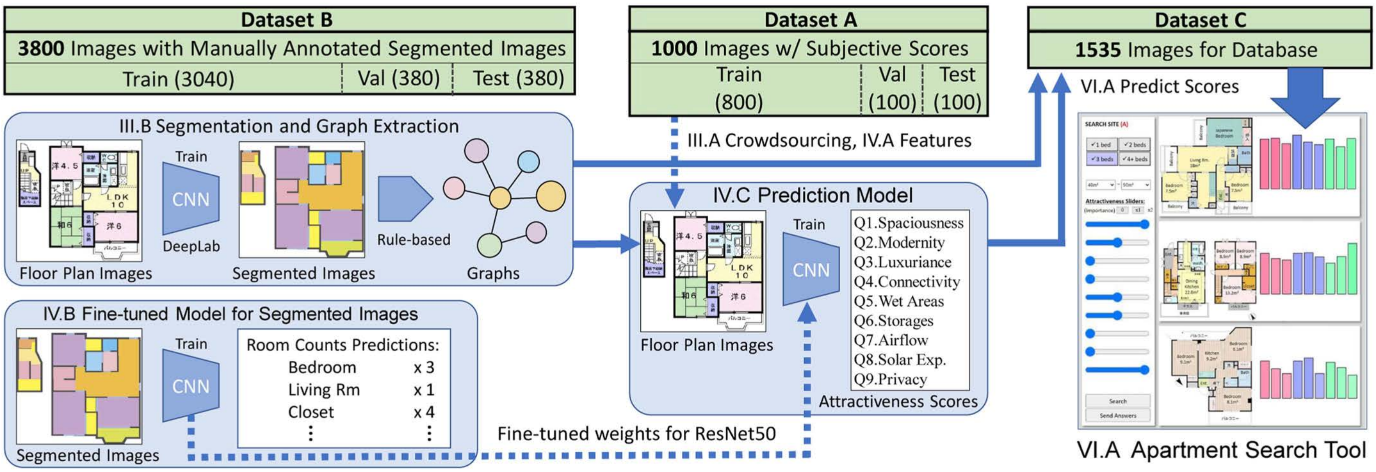


Fig. 4. Overview of our framework.

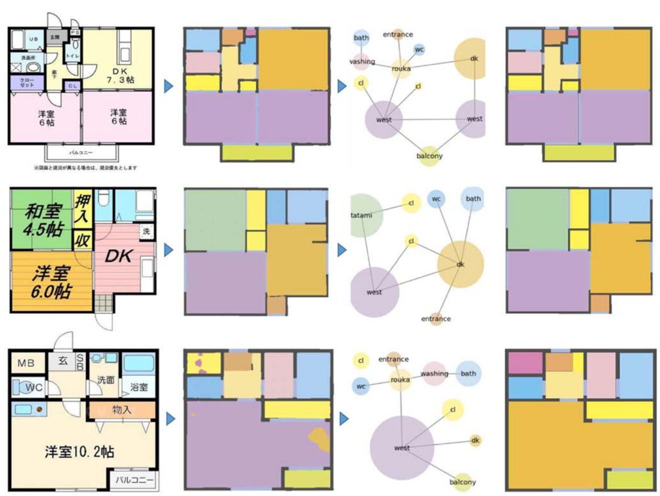


Fig. 5. Examples of input images (first column on the left), inferred semantically segmented images (second column), extracted graphs (third column), and grand truth segmented images (forth column).

C. *Overview of Our Datasets*

Figure 4 shows an overview of our framework. To predict subjective scores only from the real-estate floor plan images, we prepared three sets of floor plan images without any overlaps to avoid bias for network models' training.

The first set of 1,000 images (dataset A) was used for the dataset with subjective scores and to train our prediction model in Section IV.

The second set with 3,800 images (dataset B) was used to obtain the network that outputs segmented images from the floor plan images as input, which as explained in Section III-B. We also used dataset B to train an ImageNet-pretrained feature extractor network using color-coded semantically segmented images, as shown in bottom left of Figure 4. We additionally developed this network since our proposed network in Figure 8 explained in Section IV-C shows improved performance using the new weights fine-tuned by this network for ResNet50 [79] instead of using the weights only pre-trained on ImageNet on segmented images. The network was based on ResNet50 and used the segmented image as input. It was pre-trained first using ImageNet and then fine-tuned for the multi-task classification task to predict the number of rooms of the 13 room types, excluding the wall and unknown elements, from the 15 types defined in Section III-B.

Finally, we predicted subjective scores using our model from a separate set of 1,535 floor plan images (hereafter, "dataset C"; please see Figure 4), and it was also used for our proposed apartment search tool in Section VI.

IV. Proposed Methods

In this section, we introduce our prediction model. We propose to use two types of inputs for the model: floor plan images and structured data features. For the images, we used both floor plan images and semantically segmented images from Section III. In Section IV-A, we explain our method to extract four features from the structured data based on graphs and metadata of floor plans. The image features are described in Section IV-B. In Section IV-C, we introduce our prediction model in detail.

A. *Features*

1) Emerging Subgraphs: To extract frequently appearing common subgraphs that are very important and contribute to either higher or lower evaluation scores for each question regarding the floor plans, we first performed frequent subgraph mining on a total of 1,000 graphs in dataset A using graph-based substructure pattern mining (gSpan) [80]. For the condition to evaluate subgraph isomorphism, we considered node attributes that represent the room types for graph matching. The edge attributes that represent door and window types were not considered. As a result, 162,470 subgraphs were extracted by setting the minimum support threshold to 5 (i.e., the condition for extracting the common subgraph corresponding to at least five out of 1,000). The following three steps were further performed to extract subgraphs that are more relevant to each subjective score.

Step 1: For each question from Q1 to Q9 and the "Total Score" from Section III-A (ten items in total), common subgraphs that were included in the floor plan graphs with evaluation scores of the top and bottom 10% were extracted and separated into a total of 20 classes. We further narrowed down the selection of subgraphs by setting the following three thresholding strategies.

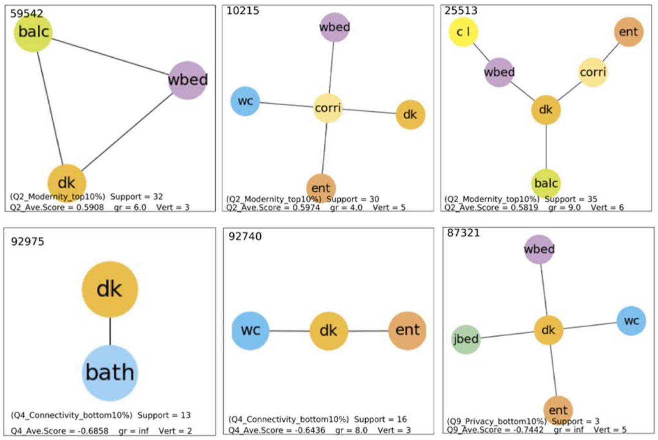


Fig. 6. Examples of emerging subgraphs frequently found in floor plans having a high Q2 (modernity) score (top row). The top left also appears in Figure 1 a (circled in red). Examples found in floor plans having a low Q4 (connectivity) (bottom left & middle) and a Q9 (privacy) score (bottom right).

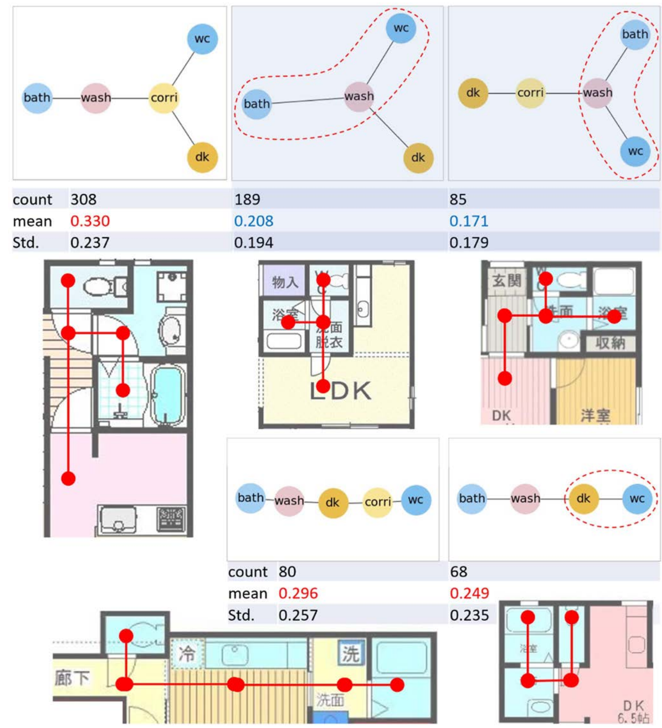


Fig. 7. The main five types of subgraphs for wet areas that make up 76% of all floor plans. Two subgraphs that include the linkage (wc)-(wash)-(bath), circled in red, have lower mean scores for Q5 (wet areas) than the others.

First, the minimum support threshold was set to 10. Second, the average score of the apartments that contain the target subgraph should be 0.25 or larger for those in the top 10% classes and -0.25 or lower for those in the bottom 10% classes. Third, we also used the growth rate (GR), which is widely used for discovering discriminative patterns in emerging pattern mining [81]. The ratio between GR for the top 10% and that for the bottom 10% should be larger than 4 or smaller than $1/4$. These thresholds were set by our empirical study.

Step 2: To further extract relevant common subgraphs in the above 20 classes, we considered identification numbers of extracted subgraphs as "words" in 20 different "documents." Then, we performed term frequency - inverse document frequency (TF-IDF), which evaluates how relevant a word (subgraph) is to a document (class) in a collection of documents (classes that represent each question). Based on the obtained TF-IDF weights, we sorted important subgraphs that are relevant to the top or bottom 10% of floor plans evaluated based on a specific question in each class. The use of TF-IDF helped eliminate frequently appearing subgraphs found in multiple classes that are not yet relevant to any specific class.

For any subgraph that included other subgraphs (i.e., inclusion dependency) within the same class, subgraphs with a minimum number of nodes were always kept. If the larger-size subgraph in the top 10% class has a larger mean evaluation score than those of the minimum-size subgraphs defined above, it is also kept (and vice-versa for the bottom 10% classes).

Step 3: We sorted all the extracted subgraphs from the previous steps based on the mean evaluation score in each class and obtained the top 20 subgraphs for each of the 20 classes. By eliminating duplicates that appeared in multiple classes from a total of 400 extracted emerging subgraphs, we obtained 230 unique emerging subgraphs. For each floor plan in the dataset, a 230-dimensional feature vector that contained 0 or 1 based on the absence or presence of 230 emerging subgraphs, respectively, were extracted.

Figure 6 shows examples of the emerging subgraphs that were extracted using the above three steps. The subgraph on the top left is a triangle graph connecting (wbed)-(dk)-(balc) as a subgraph that appears in the top 10% of the Q2 (modernity) class. For example, the floor plan in Figure 1 a includes this subgraph and has a very high evaluation score for Q2. It represents the wide balcony space that is open to both the bedroom and kitchen and is often considered a contemporary layout among Japanese apartments (advertised as a "wide-span balcony"). The subgraph with two nodes, (dk)-(bath), was in the class representing the bottom 10% of the Q4 (adjacency and connectivity) class (bottom left in Figure 6). This has a bathroom door that immediately opens to a living room, allowing anyone to enter a communal space directly after taking a bath, which is not considered desirable. Adding a washing room between the two nodes is a common practice in Japan, as it provides a room to change clothes before entering the bathroom. This subgraph was also included in the class for the bottom 10% for the Q9 (privacy) class. Our method was able to extract emerging subgraphs representing such characteristics of floor plans in 20 classes.

2) Subgraphs for Wet Areas: Wet areas are particularly important because they are more personal spaces. Therefore, special attention was paid to them. For each floor plan in dataset A, a connected subgraph containing rooms related to the usage of water, including dining kitchen (dk), washroom (wash), bathroom (bath), and restroom (wc), was extracted. Any nodes bridging the above four nodes such as a corridor (corri) that were necessary to form a single connected subgraph that contained all four of the above nodes were also included depending on the floor plan layouts. If there were more than two such bridging nodes and only one node was sufficient to form a connected subgraph, the node that served a communal use or as circulation, such as corridor, was selected over a node use for a private purpose, such as a bedroom. In total, 162 unique subgraph types for water-related rooms were extracted. Out of all the floor plans, 76% belonged to the five types. 120 types of subgraph appeared only once and therefore they were discarded. As a result, 42 subgraph types were used in this study.

Two out of the five remaining types of subgraphs included three nodes that were directly linked: (wc)-(wash)-(bath). However, this is not a desirable layout, as its circulation paths for (wc) and (bath) crisscross at the washing room (Figure 7). It forces one to enter (wash) while someone else is still present after using either room. The mean evaluation scores for Q5 (wet areas) for those two types were lower (0.171 and 0.208) than the mean scores from the other three types (0.330, 0.296, and 0.249). As a result, we extracted a 42-dimensional one-hot feature vector based on the presence of the 42 types of subgraphs for wet areas.

3) Feature Based on MCS Graph Similarity: The similarity between graphs of all 800 floor plans in the training set of dataset A was calculated based on the method described in [17], [82], [83] using the maximum common subgraph (MCS) as a graph similarity measure. The similarity was 1 when the two graphs perfectly matched, and 0 when there were no common parts. Any input graph could be expressed by an 800-dimensional vector that represents distances from each of the 800 graphs in the training set of the dataset A. We used this vector based on the MCS similarity to extract a feature that represented an entire (global) characteristic of each apartment, as opposed to a local sub-structure from a subgraph.

4) Feature Based on Metadata: In addition to the above three features extracted from graph structures, we listed areas and numbers of room types using the metadata for each floor plan. This resulted in 30-dimensional feature vectors that represented the areas and numbers of room types.

From the above, we obtained four feature vectors that represented structured data based on emerging subgraphs, subgraphs for wet areas, MCS graph similarity, and metadata. All feature vectors were standardized before using them for machine learning (see the next section) such that their distributions had a mean value of 0 and a standard deviation of 1.

B. Image Features

We also fed the network two types of images: floor plan images and consistently color-coded semantically segmented images (prepared for the automated graph generation in Section III). The inputs to the network were two sets of RGB images of resolution 224×224.

First, for the floor plan images, using ResNet50 [79] pre-trained on ImageNet [84], a 2,048-dimensional vector of the pool5 layer was extracted from deep features of the images. Then, a new fully connected (FC) layer was added in place of the original one.

Second, for the segmented images, we used the same network architecture, but the network was pre-trained first using ImageNet and then fine-tuned for the multi-task classification task to predict the number of rooms of the 13 room types, excluding walls and unknown elements, from the 15 types defined

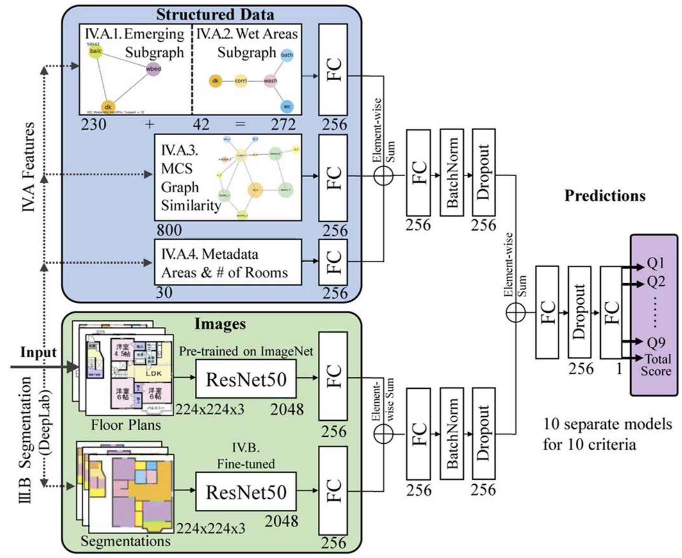


Fig. 8. Proposed network architecture.

in Section III-B. Similarly, a 2,048-dimensional vector of the pool5 layer was extracted.

Two features extracted from the two sets of images described above were also added in the same manner as above, followed by the FC, batch normalization (BN) [85], and dropout [86] layers. Finally, these two added features from the structured data and images were added again, followed by the FC and dropout layers. The final FC layer generated a value predicting the score (see Figure 8 for details). For the FC layers, we used Leaky ReLu [89] as an activation function.

C. Prediction Model

Figure 8 shows our proposed network architecture to predict the evaluation scores from Q1 to Q9 and the "Total Score" (ten separate models in total).

Two features based on emerging subgraphs (IV-A1) and subgraphs for wet areas (IV-A2) were concatenated into one 272-dimensional vector, which was reduced to 256 dimensions using a FC layer. The 800-dimensional feature vector from the MCS graph similarity (IV-A3) and the 30-dimensional feature based on metadata, such as areas and numbers of room types (IV-A4), were also reduced to 256 dimensions using FC layers. Then, the above three input features were added, followed by the FC, BN, and dropout layers. The latter two layers were added to improve generalization and to reduce overfitting.

V. Experiments

We used the dataset created in Section III and divided it into 800 floor plans for training, 100 for validation, and 100 for testing. We trained ten separate models for predicting ten scores using the network in Section IV-C because it is better than multi-task learning using a single model in our preliminary experiment. Regression models were created for each using the mean squared error (MSE) as a loss function, and we trained these networks using 800 training images. We applied the Momentum Stochastic Gradient Descent (SGD) algorithm to train models with a batch size of 20, a learning rate of 10^{-3}, a decay rate of 2.86×10^{-5}, and a momentum of 0.9 for the 35 epochs. The system was implemented using Keras.[3] We applied the pearson correlation coefficient (PCC) to measure the correlation between the predicted values and the evaluation scores from the dataset as ground truths for 100 floor plans in the test data. We also used the root-mean-squared error (RMSE) to evaluate the performance. We repeated the above process five times using a different randomly selected set of 800 training, 100 validation, and 100 test floor plans each time, and used the mean values of all five results as the final PCCs for the proposed as well as the following comparative baseline models, which were prepared to compare and verify the prediction results.

A. Baseline Methods

As baseline methods, we prepared the following seven models, of which five models use DNNs by subtracting one of five input features from the proposed network in Figure 8:

- Images Only: The model with only two input features based on floor plans and color-coded semantically segmented images.
- w/o SubG: The model without the input features based on subgraphs for emerging (IV-A1) and for wet areas (IV-A2).
- w/o MCS: The model without the input features based on MCS graph similarity (IV-A3).
- w/o Meta: The model without the input features based on metadata such as area and number of room types (IV-A4).
- w/o Img: The model without the features for floor plan images.
- w/o Segm: The model without the input features based on consistently color-coded semantically segmented images.
- Upper Bound: The model's network architecture is identical to the proposed model. However, the dataset based on GT segmented images and GT graphs was used for training and testing of the model as the best case, instead of using inferred segmented images and extracted graphs used in our proposed model explained in Section III-B)

B. Results

1) Segmentation Performance: First, we discuss the accuracy of the semantic segmentation using dataset B introduced in III-B. The number of train/validation/test data were set to 3040/380/380, respectively, and we trained the segmentation prediction network model to learn the correspondence between the floor plan images and the GT label masks using the training and validation data. After the training, the test data was used for evaluation with a metric mean intersection over union (IoU) [87] defined by (1), where n_c is the number of classes, t_i is the total number of pixels belonging to class i, and n_j is the total number of predicted as class j belonging to class i.

$$\text{mean IoU} = \frac{1}{n_c} \sum_{i=1}^{n_c} \frac{n_{ii}}{t_i + \sum_{i'=1}^{n_c} n_{i'i} - n_{ii}}. \qquad (1)$$

[3][Online]. Available:https://keras.io

The mean IoU for the test data was 82.1%, and the average mean accuracy was 89.0% for the test set. Although the segmentation accuracy is not perfect, it is acceptable considering the prediction accuracy of functionality and comfort prediction as discussed below.

2) Performance Comparison: To evaluate the performance of each model, we used two metrics, namely PCC and RMSE. Table I shows the PCC results of both the proposed and baseline models using the test set of dataset A, and Table II shows the RMSE values. Note that a higher PCC and a lower RMSE mean better prediction performance. Our proposed model, which uses all features extracted from the structured data, floor plan images, and segmented images, outperformed all the comparative baseline models in terms of the mean value of PCCs from all ten criteria (0.701) and recorded the highest PCC for nine out of ten criteria (Table I). While we used ResNet50, which is known for its high performance in a wide range of image-recognition tasks, the naive approach only using two sets of images, The Images Only model, has a lower mean value of PCCS, 0.491. Therefore, it can be inferred that the network compensates for the weakness of each feature by combining them. The PCC values for our proposed model were over 0.7 for five criteria (Q1 = 0.721, Q2 = 0.793, Q3 = 0.776, Q6 = 0.751, and Total Score = 0.794), and over 0.8 for the "privacy" criteria (Q9 = 0.816). Table II also shows that our proposed model outperforms baseline models with the lowest mean value of RMSEs from all ten criteria.

We carried out the statistical significance tests to determine whether PCCs and RMSEs between the proposed and baseline models are statistically significant or not. Following the recommended procedure in Section 7.6.1 of ITU-T Rec. P.1401 [88], the statistical significance tests for PCC use statistics derived from Fisher's z-transformed correlation coefficients in each comparison, compared with the 95% two-tailed Student's t-test critical value. In Table I, we have used an asterisk to denote the case when 0.05 significance level of difference compared to the proposed model's PCC is found. Table III shows p-values from Images Only and Upper Bound models. Our multimodal network model outperformed the Images Only model with p-values$>$0.05 in eight out of ten criteria (all criteria except for Q5 and Q8) (Table III).

The statistical significance tests for the differences in RMSEs were also performed, following the recommended procedure in Section 7.6.4 of ITU-T Rec. P.1401 [88]. In Table II, we have used an asterisk to the RMSE values to denote the case when significant differences between the proposed model's RMSEs are found. The RMSEs from our proposed model and the Images Only model were significantly different in eight out of ten criteria (all criteria except for Q5 and Total Score), indicating that the improvements using our multimodal network are statistically significant. Only Q5 (Wet Areas) does not show significant improvement using our model suggested by the differences in both PCC and RMSE values. As stated in Section IV-A2, subgraphs for wet areas have fewer pattern variations than other features, which may make prediction of the Q5 (Wet Areas) score more difficult. The test also suggests that our model is statistically significantly better than the model without the features for floor

TABLE I
PREDICTION ACCURACY COMPARISON. VALUES ARE IN TERMS OF PEARSON CORRELATION COEFFICIENTS (PCC). BLACK CELLS REPRESENT PCC>0.7 AND GRAY CELLS INDICATE PCC<0.7. STATISTICAL SIGNIFICANCE BETWEEN PROPOSED AND BASELINE MODELS WERE DETERMINED BY A T-TEST

Model	Images Only	w/o SubG	w/o Meta	w/o MCS	w/o Img	w/o Segm	Proposed	Upper Bound
Q1.Spaciousness	0.490*	0.669	0.707	0.730	0.637	0.648	0.721	0.794
Q2.Modernity	0.600*	0.782	0.772	0.787	0.771	0.765	0.793	0.838
Q3.Luxuriance	0.549*	0.762	0.751	0.753	0.749	0.747	0.776	0.805
Q4.Connectivity	0.432*	0.542	0.592	0.620	0.546	0.571	0.637	0.698
Q5.Wet Areas	0.392	0.477	0.422	0.452	0.439	0.499	0.525	0.585
Q6.Storage	0.518*	0.715	0.728	0.751	0.713	0.733	0.751	0.778
Q7.Airflow	0.363*	0.494	0.573	0.514	0.446	0.592	0.607	0.657
Q8.Solar Exp.	0.402	0.520	0.531	0.571	0.525	0.564	0.591	0.680
Q9.Privacy	0.567*	0.764	0.786	0.780	0.791	0.748	0.816	0.822
Total Score	0.553*	0.727	0.763	0.740	0.712	0.755	0.794	0.809
Average	0.491	0.645	0.662	0.670	0.633	0.662	0.701	0.747

(* Denotes p-Value<0.05).

TABLE II
ROOT-MEAN-SQUARED ERROR (RMSE) COMPARISON

Model	Images Only	w/o SubG	w/o Meta	w/o MCS	w/o Img	w/o Segm	Proposed	Upper Bound
Q1.Spaciousness	0.338*	0.266	0.254	0.242	0.285*	0.259	0.236	0.218
Q2.Modernity	0.357*	0.257	0.266	0.255	0.272	0.267	0.253	0.202
Q3.Luxuriance	0.281*	0.213	0.217	0.205	0.235*	0.209	0.198	0.192
Q4.Connectivity	0.342*	0.252	0.235	0.227	0.307	0.259	0.245	0.251
Q5.Wet Areas	0.253	0.215	0.225	0.209	0.276*	0.203	0.216	0.238
Q6.Storage	0.344*	0.282	0.274	0.259	0.265	0.282	0.249	0.228
Q7.Airflow	0.332*	0.272	0.255	0.257	0.318*	0.265	0.233	0.267
Q8.Solar Exp.	0.303*	0.255*	0.253*	0.247	0.255*	0.239	0.211	0.227
Q9.Privacy	0.374*	0.242	0.242	0.248	0.235	0.265	0.239	0.218
Total Score	0.257	0.253	0.234	0.244	0.253	0.240	0.218	0.243
Average	0.318	0.251	0.246	0.239	0.270	0.249	0.230	0.228

(* Denotes 0.05 Significance Level of Difference Compared to the Proposed Model's RMSE).

TABLE III
SIGNIFICANCE OF THE DIFFERENCE IN PCCS COMPARED TO THE PROPOSED MODEL'S PCCS. VALUES ARE IN TERMS OF P-VALUES

Model	Images Only	Upper Bound
Q1.Spaciousness	0.009*	0.230
Q2.Modernity	0.007*	0.350
Q3.Luxuriance	0.004*	0.589
Q4.Connectivity	0.042*	0.446
Q5.Wet Areas	0.240	0.544
Q6.Storage	0.005*	0.647
Q7.Airflow	0.024*	0.560
Q8.Solar Exp.	0.079	0.294
Q9.Privacy	0.001*	0.895
Total Score	0.001*	0.768

(* Denotes p-Value<0.05).

plan images, w/o Img, in five out of ten criteria, including Q1, Q3, Q5, Q7, and Q8. The results indicate that the use of both features based on images and structured data extracted from graphs contributes to the improvement in prediction accuracy.

The differences in PCC and RMSE values between the proposed and Upper Bound models are not significant, indicating that the performance drops caused by the use of graphs extracted from inferred segmented images, instead of the GT graphs described in Section III-B, are not significant (average difference in PCCs from nine questions is 0.046). Our model can automatically generate all input features from floor plan images alone as inputs using the processes in Section III-B to extract segmented images and graphs, allowing us to develop the search tool described in Section VI without the time-consuming and labor-intensive manual annotation process for segmented images. We also tested the prediction accuracy of our proposed network architecture trained and tested using a dataset based on raw values rated by the participants as evaluation scores, instead of using standardized evaluation scores in our dataset explained in Section III-A. We found no statistically significant difference in PCC and RMSE values between the two datasets (i.e., average PCC was 0.699 using raw score values and 0.701 using ours).

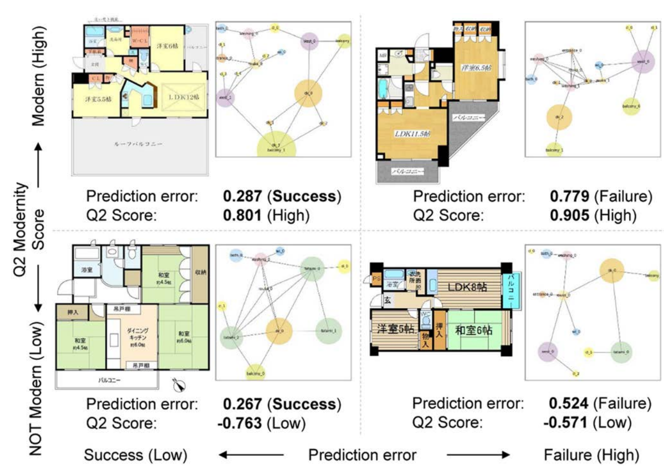


Fig. 9. Success and failure cases from our model's prediction results. While both floor plans in the top row have high ground-truth scores for Q2 Modernity in our dataset, the prediction error for the top-left was lower than the prediction error for the top-right, indicating a more successful prediction result.

3) Success and Failure: Figure 9 depicts some success and failure examples from our prediction results for Q2 Modernity scores. Here, "success" means that the prediction error between the predicted score and the ground-truth score from the dataset is

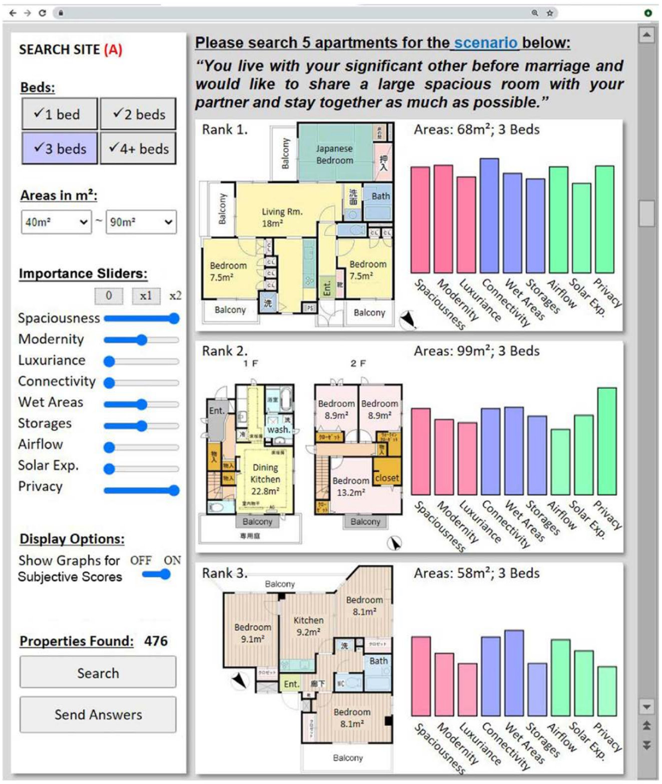


Fig. 10. The proposed apartment search tool.

lower. It is inferred that if similar floor layouts are included in the proposed network's training data, prediction errors would likely be reduced. If the floor plans are rare or uncommon examples that do not exist in the training set, prediction errors could increase.

In conclusion, although the measurement of the functionality of floor plans is a very subjective problem, our proposed prediction models are able to achieve a very strong correlation with human evaluation.

VI. Application to Online Property Search

A. Implementation

We introduce a new interface for an apartment search tool that implements functionality and comfort as query items. In addition to a common search interface based on user selection of the number of bedrooms and a range of areas, our tool offers *importance sliders* with adjustable weights for importance in three levels: 0 (none); $\times 1$ (important); and $\times 2$ (very important), for nine subjective criteria (see Figure 10). After a user presses the search button, the tool calculates the scores for all floor plans based on the weighted sum of the predicted subjective values using the information from the sliders, and displays the scores in ranked order. This feature allows a user to search apartments based on a controlled weighted priority for qualitative criteria (i.e., extremely spacious apartments with sufficient privacy and storage spaces).

For our proposed search tool's floor plan database, we prepared a new set of 1,535 floor plan images that include apartments with one, two, three, and four or more bedrooms in balanced proportions (dataset C). These images were completely unbiased and were not previously used in this study. Using the pre-trained network using DeepLab v3+ and the rule-based method in Section III-B, we obtained the 1,535 corresponding segmented images and graphs. Next, we executed the procedures outlined in Sections IV-A and IV-B to extract the five features from these images, which were then used to obtain the predicted attractiveness scores for the nine criteria for each of the 1,535 floor plans using the best model (i.e., the one with the highest PCC) from the proposed network introduced in Section IV-C.

B. Procedure

We evaluated our method through a large-scale user study. Among the 200 participants recruited through a crowdsourcing service, we removed three who chose the same rating for all questions (i.e., "straight-lining") and another six who did not complete the survey to obtain a final count of 191 participants with a wide variety of attributes. Out of the 191 participants, 81 identified themselves as male and 109 as female. The numbers of participants in their 20 s, 30 s, 40 s, 50 s, and over 60 were 30, 78, 53, 20, and 4, respectively. Of the total, 38 lived alone at the time of the experiment, while 81 participants were married. While 115 lived in urban areas, 60 lived in suburban areas, and 11 in rural areas. Participants were required to answer all questions to complete the task. Incomplete responses were not included in the study. We verified that all histograms of responses for 5-point Likert scale questions were approximately normally distributed, and that no isolated peaks with unusual values were found.

To make the study task realistic and provide internal motivation, we prepared hypothetical scenarios with different demands from five unique families and asked participants to search five apartments for each scenario. We set the scenarios based on five completely different family structures to avoid demographic background bias of the participants.

- "You are a married couple, and your two children need their own rooms soon. You want a functional floor plan layout and don't want to pay extra for unnecessarily large spaces."
- "You are a family of five, living with one child and your spouse's parents. A well-functioning home with large storage spaces, kitchen, and wet areas are your top priority since you have a big family."
- "Due to the COVID-19 pandemic, you have been working from home and would like to have your own study room. You and your spouse want to have separate rooms to respect each other's privacy."
- "You live with your significant other before marriage and would like to share a large spacious room with your partner and stay together as much as possible."
- "You and your partner are a young couple and hope to have at least one child in the future. You prefer to have a large balcony for your family to spend the weekend together."

We assigned these tasks using both our proposed tool and another baseline tool. The baseline tool represented commonly available real estate portal sites without our proposed features,

TABLE IV
THE DIRECT COMPARISON QUESTIONS WERE ASKED ON A 5-POINT LIKERT SCALE. A HIGHER SCORE INDICATED A PREFERENCE FOR OUR PROPOSED TOOL, WHILE A LOWER SCORE INDICATED A PREFERENCE FOR THE BASELINE TOOL. A SCORE OF 3 INDICATED NO PREFERENCE

Question	Mean	CI	p-value
Which tool helped you find more desirable floor plans?	4.03*	[3.90, 4.17]	4.9×10^{-34} < 0.001
Which tool did you enjoy better while searching?	3.97*	[3.83, 4.11]	1.6×10^{-30} < 0.001
Which tool was faster for you to search floor plans?	3.84*	[3.67, 4.01]	1.1×10^{-18} < 0.001
Which tool was easier to search apartments?	3.76*	[3.60, 3.93]	2.5×10^{-16} < 0.001
Which tool was more intuitive for you to search floor plans?	3.43*	[3.25, 3.62]	7.0×10^{-6} < 0.001

* Significantly different based on 95% confidence interval.

featuring a search interface for a user to select the number of bedrooms and a range of areas. We studied major real estate portal sites [4567], and found the above two items as common search features. The differences between the proposed and baseline tool are that the baseline tool has an identical interface except that it does not have functionality- and comfort-related options. As we wanted participants to focus on analyzing information only readable from floor plans to enable a fair comparison with our proposed tool, we excluded other common search features based on the location and cost of properties. The study employed a within-participant design in which participants used both tools (counterbalanced across participants) and provided feedback on them.

Participants were asked to search five apartments that met needs of each of the five scenarios using one tool, and then switch to the other tool for the same search (with the order of tool counterbalanced across participants). After completing both search tasks, participants completed a post-experiment survey. The survey asked participants to directly compare their experience with the proposed tool to the baseline tool in five Likert-scale questions (shown in Table IV). We also asked them to rate 50 selected floor plans (i.e., five plans $\times$ five scenarios $\times$ two tools) in a five-grade score on a scale of 1 (very unsatisfied) to 5 (very satisfied) (i.e., each participant gave a score of 1 to 5 for each retrieved floor plan image.).

C. Results

Overall, it can be observed that the participants showed a significant preference for our proposed tool in response to all five direct comparison questions in Table IV. For each question, 95% confident intervals of mean scores are indicated. In addition, a one-sample one-tailed t-test ($p < 0.05$) was used to evaluate whether the mean responses are greater than 3 (i.e., a score of 3 indicates no preference) at the significance level. Participants appreciated the proposed tool as well. Compared to the baseline tool, it helped them find significantly more desirable floor plans

[4][Online]. Available:https://lifull.com
[5][Online]. Available:https://www.livable.co.jp
[6][Online]. Available:https://suumo.jp
[7][Online]. Available:https://www.redfin.com

TABLE V
THE 5-GRADE RATINGS OF ALL FLOOR PLANS SELECTED BY PARTICIPANTS USING TWO METHODS

	Baseline	Proposed	p-value
Mean scores of the 5-grade ratings of all floor plans selected	3.75	4.01	1.67×10^{-40} < 0.001

An independent two-sample one-tailed t-test was used on two samples of scores of floor plans using our tool and the baseline tool.

(M = 4.03, 95% CI[3.90, 4.17], p = $4.9 \times 10^{-34} < 0.001$), and gave them a significantly more enjoyable experience (M = 3.97, 95% CI[3.83, 4.11], p = $1.92 \times 10^{-30} < 0.001$).

Even though the participants spent a longer average time to complete a search task using the proposed tool (M = 158.0 s, median = 131 s, SD = 105.7 s) than the baseline tool (M = 129.0 s, median = 105 s, SD = 97.4 s), the result from the direct comparison question shows that participants felt that they found the desired floor plans faster using the proposed tool (M = 3.84, 95% CI[3.67, 4.01], p = $1.1 \times 10^{-18} < 0.001$). The number of clicks on buttons per search increased using the proposed tool (M = 18.3, median = 17, SD = 7.6) compared to the baseline tool (M = 13.2, median = 12, SD = 5.9). These results show that our proposed tool required more time and mouse clicks for users due to additional features that are not included in the baseline tool (e.g., importance sliders). However, these additional features did not lead them to believe that the proposed system is harder to use. Participants felt that the proposed tool made the task significantly easier (M = 3.76, 95% CI[3.60, 3.93], p = $2.5 \times 10^{-16} < 0.001$) and more intuitive (M = 3.43, 95% CI[3.25, 3.62], p=$7.0 \times 10^{-6} < 0.001$). The longer search time and additional operations did not make them feel burdened with tasks, and they preferred the proposed tool over the baseline one.

In Table V, the 5-grade ratings of all floor plans selected by participants also indicated that they were significantly more satisfied with their selections using the proposed tool than the baseline one (M(proposed) = 4.01, M(baseline) = 3.75, p=1.67×10^{-40} <0.001). To compute the p-value, we used an independent two-sample one-tailed t-test, and the mean score for selected floor plans by participants using our proposed tool was found to be significantly higher than the mean score for selected floor plans using the baseline tool.

VII. LIMITATIONS

One limitation of our work is that predicted functionality and comfort scores do not come with explanations. People may have different opinions and viewpoints on such subjective scores, and visualizing how the system evaluates floor plan images would be preferred.

Our proposed method has been applied only to apartment floor plans in Japan. As can be seen in the figures, the drawing styles employed in the floor plan images are very diverse; however, the style employed in other counties may be more distinct. Therefore, the prediction model will need to be trained for each country. We expect that our segmentation and functionality and comfort prediction models can be used as pre-trained models for fine-tuning, but this is left as a future work.

VIII. Conclusion

We created and analyzed a large-scale dataset based on subjective evaluation of real estate floor plan images using crowdsourcing. Our proposed methods for extracting features from graph structures and images of floor plans proved to be effective, as we obtained functionality and comfort prediction models with relatively high accuracy (PCC = 0.701) for very subjective scores, which is essentially different from conventional image recognition tasks where the answer is apparent to all the evaluators. This study is the first work to propose a highly accurate prediction model for dwelling functionality and comfort using machine learning. We applied the results of the prediction model to our new apartment search tool using functionality and comfort as query items, and our user study showed that our tool could provide a better user experience.

References

[1] V. James, S. Wu, A. Gelfand, and C. Sirmans, "Apartment rent prediction using spatial modeling," *J. Real Estate Res.*, vol. 27, no. 1, pp. 105–136, 2005.

[2] O. Poursaeed, T. Matera, and S. Belongie, "Vision-based real estate price estimation," *Mach. Vis. Appl.*, vol. 29, no. 4, pp. 667–676, 2018.

[3] M. Heidari, S. Zad, and S. Rafatirad, "Ensemble of supervised and unsupervised learning models to predict a profitable business decision," in *Proc. IEEE Int. IOT, Electron. Mechatronics Conf.*, 2021, pp. 1–6.

[4] H. Seya and D. Shiroi, "A comparison of residential apartment rent price predictions using a large data set: Kriging versus deep neural network," *Geograph. Anal.*, vol. 54, no. 2, pp. 239–260, 2022.

[5] T. Yamasaki, J. Zhang, and Y. Takada, "Apartment structure estimation using fully convolutional networks and graph model," in *Proc. ACM Workshop Multimedia Real Estate Tech.*, 2018, pp. 1–6.

[6] S. Law, B. Paige, and C. Russell, "Take a look around: Using street view and satellite images to estimate house prices," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 5, Sep. 2019, Art. no. 54.

[7] M. Zeppelzauer, M. Despotovic, M. Sakeena, D. Koch, and M. Döller, "Automatic prediction of building age from photographs," in *Proc. ACM Int. Conf. Multimedia Retrieval*, 2018, pp. 126–134.

[8] J. H. Bappy, J. R. Barr, N. Srinivasan, and A. K. Roy-Chowdhury, "Real estate image classification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2017, pp. 373–381.

[9] X. Wang, Y. Takada, Y. Kado, and T. Yamasaki, "Predicting the attractiveness of real-estate images by pairwise comparison using deep learning," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2019, pp. 84–89.

[10] T. Hanazato, Y. Hirano, and M. Sasaki, "Syntactic analysis of large-size condominium units supplied in the Tokyo metropolitan area," *J. Architecture Plan.*, no. 591, pp. 9–16, 2005.

[11] A. Takizawa, K. Yoshida, and N. Katoh, "Applying graph mining to rent analysis considering room layouts," *J. Environ. Eng. (Trans. AIJ)*, vol. 73, no. 623, pp. 139–146, 2008.

[12] R. Hattori, K. Okamoto, and A. Shibata, "Visualizing the importance of floor-plan image features in rent-prediction models," in *Proc. Joint 11th Int. Conf. Soft Comput. Intell. Syst. 21st Int. Symp. Adv. Intell. Syst.*, 2020, pp. 1–3.

[13] R. Hattori, K. Okamoto, and A. Shibata, "Impact analysis of floor-plan images for rent-prediction model," *J. Jpn. Soc. Fuzzy Theory Intell. Informat.*, vol. 33, no. 2, pp. 640–650, 2021.

[14] K. Solovev and N. Pröllochs, "Integrating floor plans into hedonic models for rent price appraisal," in *Proc. Web Conf.*, 2021, pp. 2838–2847.

[15] G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bags of keypoints," *Proc. Workshop Stat. Learn. Comput. Vis.*, 2004, pp. 1–22.

[16] S. Sirmans, D. Macpherson, and E. Zietz, "The composition of hedonic pricing models," *J. Real Estate Literature*, vol. 13, no. 1, pp. 1–44, 2005.

[17] Y. Takada, N. Inoue, T. Yamasaki, and K. Aizawa, "Similar floor plan retrieval featuring multi-task learning of layout type classification and room presence prediction," in *Proc. IEEE Int. Conf. Consum. Electron.*, 2018, pp. 1–6.

[18] K. Naoki, Y. Toshihiko, A. Kiyoharu, and O. Takemi, "Users' preference prediction of real estates featuring floor plan analysis using floornet," in *Proc. ACM Workshop Multimedia Real Estate Tech.*, 2018, pp. 7–11.

[19] N. Kato, T. Yamasaki, K. Aizawa, and T. Ohama, "Users' preference prediction of real estate properties based on floor plan analysis," *IEICE Trans. Inf. Syst.*, vol. E103-D, no. 2, pp. 398–405, 2020.

[20] M. Yamada, X. Wang, and T. Yamasaki, "Graph structure extraction from floor plan images and its application to similar property retrieval," in *Proc. IEEE Int. Conf. Consum. Electron.*, 2021, pp. 1–5.

[21] V. Trinh and R. Manduchi, "Semantic interior mapology: A toolbox for indoor scene description from architectural floor plans," in *Proc. 24th Int. Conf. 3D Web Technol.*, 2019, doi: 10.1145/3329714.3338141.

[22] N. Nauata, K.-H. Chang, C.-Y. Cheng, G. Mori, and Y. Furukawa, "House-GAN: Relational generative adversarial networks for graph-constrained house layout generation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 162–177.

[23] R. Hu et al., "Graph2plan: Learning floorplan generation from layout graphs," *ACM Trans. Graph.*, vol. 39, no. 4, pp. 118–121, 2020.

[24] S.-T. Yang et al., "Dula-Net: A dual-projection network for estimating room layouts from a single RGB panorama," in *Proc. IEEE/CVF Comput. Vis. Pattern Recognit.*, 2019, pp. 3363–3372.

[25] C. Lin, C. Li, and W. Wang, "Floorplan-jigsaw: Jointly estimating scene layout and aligning partial scans," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 5674–5683.

[26] C. Liu, J. Wu, and Y. Furukawa, "FloorNet: A unified framework for floorplan reconstruction from 3D scans," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 203–219.

[27] C. Lin, C. Li, and W. Wang, "Floorplan-Jigsaw: Jointly estimating scene layout and aligning partial scans," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 5674–5683.

[28] J. Chen, C. Liu, J. Wu, and Y. Furukawa, "Floor-SP: Inverse CAD for floorplans by sequential room-wise shortest path," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 2661–2670.

[29] Y. Cui et al., "Automatic 3-D reconstruction of indoor environment with mobile laser scanning point clouds," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 3117–3130, Aug. 2019.

[30] A. Phalak, V. Badrinarayanan, and A. Rabinovich, "Scan2plan: Efficient floorplan generation from 3D scans of indoor scenes," 2020, *arXiv:2003.07356*.

[31] K. Wang et al., "Planit: Planning and instantiating indoor scenes with relation graph and spatial prior networks," *ACM Trans. Graph.*, vol. 38, no. 4, pp. 1–15, 2019.

[32] G. Zhai and X. Min, "Perceptual image quality assessment: A survey," *Sci. China Inf. Sci.*, vol. 63, no. 11, pp. 1–52, 2020.

[33] X. Min et al., "Screen content quality assessment: Overview, benchmark, and beyond," *ACM Comput. Surv.*, vol. 54, no. 9, pp. 1–36, 2021.

[34] W. Zhou, X. Min, H. Li, and Q. Jiang, "A brief survey on adaptive video streaming quality assessment," *J. Vis. Commun. Image Representation*, 2022, Art. no. 103526.

[35] Q. Li, W. Lin, J. Xu, and Y. Fang, "Blind image quality assessment using statistical structural and luminance features," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2457–2469, Dec. 2016.

[36] Q. Wu et al., "Blind image quality assessment based on rank-order regularized regression," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2490–2504, Nov. 2017.

[37] Z. Fan, T. Jiang, and T. Huang, "Active sampling exploiting reliable informativeness for subjective image quality assessment based on pairwise comparison," *IEEE Trans. Multimedia*, vol. 19, no. 12, pp. 2720–2735, Dec. 2017.

[38] H. Hofbauer, F. Autrusseau, and A. Uhl, "Low quality and recognition of image content," *IEEE Trans. Multimedia*, vol. 24, pp. 3595–3610, 2022.

[39] Y. Li et al., "User-generated video quality assessment: A subjective and objective study," *IEEE Trans. Multimedia*, early access, Oct. 29, 2021, doi: 10.1109/TMM.2021.3122347.

[40] Y. Liu et al., "Video quality assessment with serial dependence modeling," *IEEE Trans. Multimedia*, vol. 24, pp. 3754–3768, 2022, doi: 10.1109/TMM.2021.3122347.

[41] J. Gutierrez et al., "Subjective evaluation of visual quality and simulator sickness of short 360 videos: ITU-T REC," *IEEE Trans. Multimedia*, vol. 24, pp. 3087–3100, 2021.

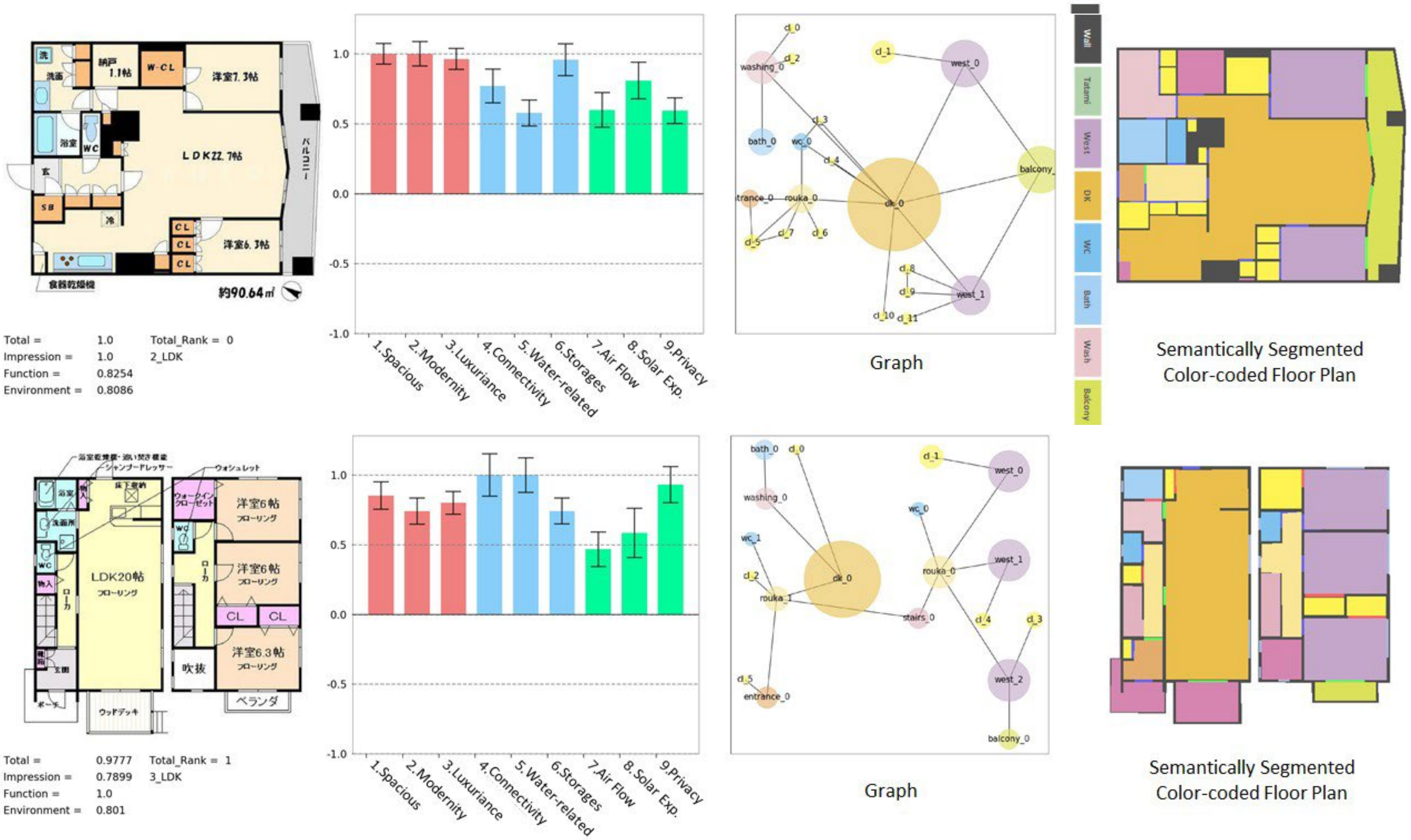
[42] P. Lebreton and K. Yamagishi, "Predicting user quitting ratio in adaptive bitrate video streaming," *IEEE Trans. Multimedia*, vol. 23, pp. 4526–4540, 2021.

[43] Y. Deng, C. C. Loy, and X. Tang, "Image aesthetic assessment: An experimental survey," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 80–106, Jul. 2017.
[44] G. Guo, H. Wang, C. Shen, Y. Yan, and H.-Y. M. Liao, "Automatic image cropping for visual aesthetic enhancement using deep neural networks and cascaded regression," *IEEE Trans. Multimedia*, vol. 20, no. 8, pp. 2073–2085, Aug. 2018.
[45] V. Hosu, B. Goldlücke, and D. Saupe, "Effective aesthetics prediction with multi-level spatially pooled features," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 9367–9375.
[46] P. Lv et al., "User-guided personalized image aesthetic assessment based on deep reinforcement learning," *IEEE Trans. Multimedia*, early access, Nov. 25, 2021, doi: 10.1109/TMM.2021.3130752.
[47] H. Zhu, Y. Zhou, L. Li, Y. Li, and Y. Guo, "Learning personalized image aesthetics from subjective and objective attributes," *IEEE Trans. Multimedia*, early access, Nov. 02, 2021, doi: 10.1109/TMM.2021.3123468.
[48] Y. Bai, Z. Zhu, G. Jiang, and H. Sun, "Blind quality assessment of screen content images via macro-micro modeling of tensor domain dictionary," *IEEE Trans. Multimedia*, vol. 23, pp. 4259–4271, 2021.
[49] X. Lei, X. Qian, and G. Zhao, "Rating prediction based on social sentiment from textual reviews," *IEEE Trans. Multimedia*, vol. 18, no. 9, pp. 1910–1921, Sep. 2016.
[50] F. Chen, R. Ji, J. Su, D. Cao, and Y. Gao, "Predicting microblog sentiments via weakly supervised multimodal deep learning," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 997–1007, Apr. 2018.
[51] R. Ji, F. Chen, L. Cao, and Y. Gao, "Cross-modality microblog sentiment prediction via bi-layer multimodal hypergraph learning," *IEEE Trans. Multimedia*, vol. 21, no. 4, pp. 1062–1075, Apr. 2019.
[52] X. Lu, Z. Lin, H. Jin, J. Yang, and J. Z. Wang, "Rapid: Rating pictorial aesthetics using deep learning," in *Proc. Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 457–466.
[53] T. Li et al., "Data-driven affective filtering for images and videos," *IEEE Trans. Cybern.*, vol. 45, no. 10, pp. 2336–2349, Oct. 2015.
[54] J. Yang et al., "Visual sentiment prediction based on automatic discovery of affective regions," *IEEE Trans. Multimedia*, vol. 20, no. 9, pp. 2513–2525, Sep. 2018..
[55] X. Yao et al., "Adaptive deep metric learning for affective image retrieval and classification," *IEEE Trans. Multimedia*, vol. 23, pp. 1640–1653, 2021.
[56] S. Ruan et al., "Color enhanced cross correlation net for image sentiment analysis," *IEEE Trans. Multimedia*, early access, Oct. 11, 2021, doi: 10.1109/TMM.2021.3118208.
[57] Y. Su, W. Zhao, P. Jing, and L. Nie, "Exploiting low-rank latent Gaussian graphical model estimation for visual sentiment distribution," *IEEE Trans. Multimedia*, early access, Jan. 06, 2022, doi: 10.1109/TMM.2022.3140892.
[58] H. Zhang and M. Xu, "Multiscale emotion representation learning for affective image recognition," *IEEE Trans. Multimedia*, early access, Jan. 25, 2022, doi: 10.1109/TMM.2022.3144804.
[59] S. Zhang, S. Zhang, T. Huang, and W. Gao, "Speech emotion recognition using deep convolutional neural network and discriminant temporal pyramid matching," *IEEE Trans. Multimedia*, vol. 20, no. 6, pp. 1576–1590, Jun. 2018.
[60] T. Li and M. Ogihara, "Toward intelligent music information retrieval," *IEEE Trans. Multimedia*, vol. 8, no. 3, pp. 564–574, Jun. 2006.
[61] G. Irie, T. Satou, A. Kojima, T. Yamasaki, and K. Aizawa, "Affective audio-visual words and latent topic driving model for realizing movie affective scene classification," *IEEE Trans. Multimedia*, vol. 12, no. 6, pp. 523–535, Oct. 2010.
[62] R. Teixeira, T. Yamasaki, and K. Aizawa, "Affective determination of video clips by low level audiovisual features -A dimensional and categorial approach," *Multimedia Tools Appl.*, vol. 61, no. 1, pp. 21–49, 2011.
[63] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Trans. Affect. Comput.*, vol. 3, no. 2, pp. 211–223, Apr.–Jun. 2012.
[64] T. Liu et al., "Sentiment recognition for short annotated GIFs using visual-textual fusion," *IEEE Trans. Multimedia*, vol. 22, no. 4, pp. 1098–1110, Apr. 2020.
[65] Q. Fang, C. Xu, J. Sang, M. S. Hossain, and G. Muhammad, "Word-of-mouth understanding: Entity-centric multimodal aspect-opinion mining in social media," *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2281–2296, Dec. 2015.
[66] W. Guo et al., "LD-Man: Layout-driven multimodal attention network for online news sentiment recognition," *IEEE Trans. Multimedia*, vol. 23, pp. 1785–1798, 2021.
[67] W. Nie, R. Chang, M. Ren, Y. Su, and A. Liu, "I-GCN: Incremental graph convolution network for conversation emotion detection," *IEEE Trans. Multimedia*, vol. 23, pp. 1785–1798, Oct. 8, 2021, doi: 10.1109/TMM.2021.3118881.
[68] D. Liu et al., "Towards unified surgical skill assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 9517–9526.
[69] M. Redi, N. O'Hare, R. Schifanella, M. Trevisiol, and A. Jaimes, "6 seconds of sound and vision: Creativity in micro-videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 4272–4279.
[70] B. Xia, X. Wang, T. Yamasaki, K. Aizawa, and H. Seshime, "Deep neural network-based click-through rate prediction using multimodal features of online banners," in *Proc. IEEE 5th Int. Conf. Multimedia Big Data*, 2019, pp. 162–170.
[71] J. Ikeda, H. Seshime, X. Wang, and T. Yamasaki, "25th international conference on pattern recognition (ICPR)," in *Proc. 25th Int. Conf. Pattern Recognit.*, 2021, pp. 2995–3002.
[72] S. Oyama and T. Yamasaki, "Visual clarity analysis and improvement support for presentation slides," in *Proc. IEEE 5th Int. Conf. Multimedia Big Data*, 2019, pp. 421–428.
[73] S. Yi, J. Matsugami, and T. Yamasaki, "Assessment system of presentation slide design using visual and structural features," *IEICE Trans. Inf. Syst.*, vol. E105-D, no. 3, pp. 587–596, 2022.
[74] C. Liu, J. Wu, P. Kohli, and Y. Furukawa, "Raster-to-vector: Revisiting floorplan transformation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2214–2222.
[75] Z. Zeng, X. Li, Y. K. Yu, and C.-W. Fu, "Deep floor plan recognition using a multi-task network with room-boundary-guided attention," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9095–9103.
[76] C. Mura, R. Pajarola, K. Schindler, and N. Mitra, "Walk2map: Extracting floor plans from indoor walk trajectories," *Comput. Graph. Forum*, vol. 40, no. 2, pp. 375–388, 2021.
[77] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
[78] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.
[79] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2016, pp. 770–778.
[80] X. Yan and J. Han, "gSpan: Graph-based substructure pattern mining," in *Proc. IEEE Int. Conf. Data Mining*, 2002, pp. 721–724.
[81] G. Dong and J. Li, "Efficient mining of emerging patterns: Discovering trends and differences," in *Proc. 5th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 1999, pp. 43–52.
[82] K. Ohara, T. Yamasaki, and K. Aizawa, "An intuitive system for searching apartments using floor plans and areas of rooms," in *Proc. 78th Nat. Conv. IPSJ*, 2016, vol. 2016, pp. 311–312.
[83] J. J. McGregor, "Backtrack search algorithms and the maximal common subgraph problem," *Softw.: Pract. Experience*, vol. 12, no. 1, pp. 23–34, 1982.
[84] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[85] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[86] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[87] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.
[88] P. ITU-T, "Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models (p. 1401)," 2020. [Online]. Available: http://handle.itu.int/11.1002/1000/14159
[89] A. L. Maas et al., "Rectifier nonlinearities improve neural network acoustic models," in *Proc. 30th Int. Conf. Mach. Learn.*, vol. 30, no. 1, 2013, p. 3.

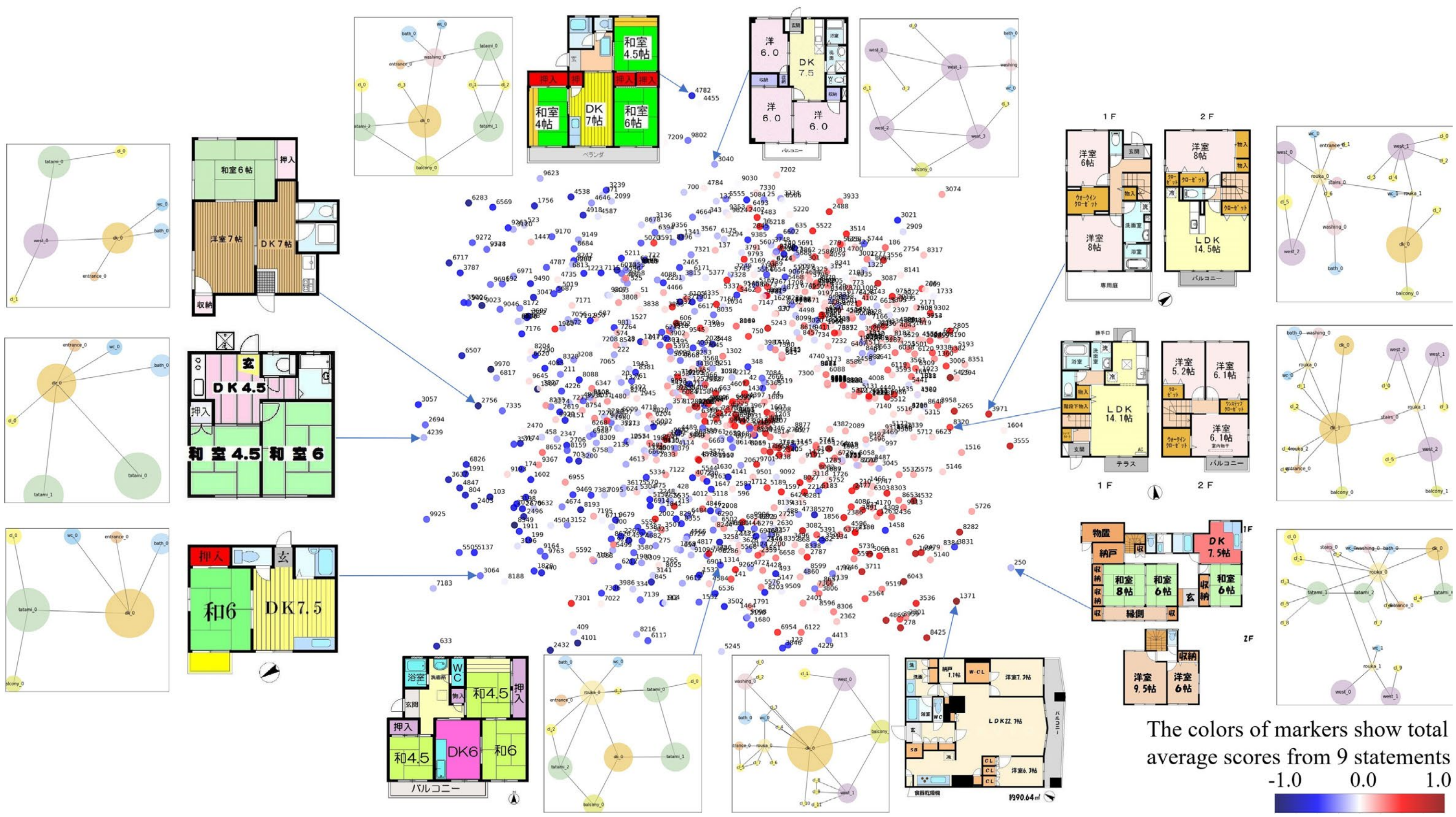
Taro Narahara (Member, IEEE) received the B.S. degree in mathematics from Waseda University, Tokyo, Japan, the M.Arch. degree from Washington University, St. Louis, MO, USA, the M.S. degree in design and computation from the Massachusetts Institute of Technology, Cambridge, MA, USA, and the Doctor of Design degree from Harvard University, Cambridge, MA, USA. He is currently an Associate Professor with the Hillier College of Architecture and Design, New Jersey Institute of Technology, Newark, NJ, USA. Between 2018 and 2020, he was a Visiting Scholar with the ETH Zurich's Institute of Technology in Architecture, Switzerland, and with the Graduate School of Information Science and Technology, The University of Tokyo, Japan. His research interests include architectural design and machine learning. Dr. Narahara is a member of ACM, SIGGRAPH, CAADRIA, and IEICE. He worked on award-winning projects such as the Mori Art Museum while associated with Gluckman Mayner Architects and Skidmore, Owings & Merrill LLP as a Licensed Architect.

Toshihiko Yamasaki (Member, IEEE) received the B.S. degree in electronic engineering, the M.S. degree in information and communication engineering, and the Ph.D. degree from the The University of Tokyo, Japan, in 1999, 2001, and 2004, respectively. He is currently a Professor with the Department of Information and Communication Engineering, Graduate School of Information Science and Technology, The University of Tokyo. He was a JSPS Fellow for Research Abroad and a Visiting Scientist with Cornell University, Ithaca, NY, USA, from February 2011 to February 2013. His research interests include attractiveness computing based on multimedia Big Data analysis, pattern recognition, and machine learning. Dr. Yamasaki is a member of ACM, AAAI, IEICE, ITE, and IPSJ.

Subjective Functionality and Comfort Prediction for Apartment Floor Plans and Its Application to Intuitive Online Property Searches, Narahara T, and Yamasaki T., IEEE Transactions on Multimedia (TMM) 2022. [IEEE, PDF, Video]



(Additional figures from my recent research) Examples of our generated functionality and comfort evaluation dataset using real estate floor plan images. Each example shows a floor plan image, a bar graph for nine evaluation measures, a graph, and a semantically segmented color-coded floor plan from left to right.

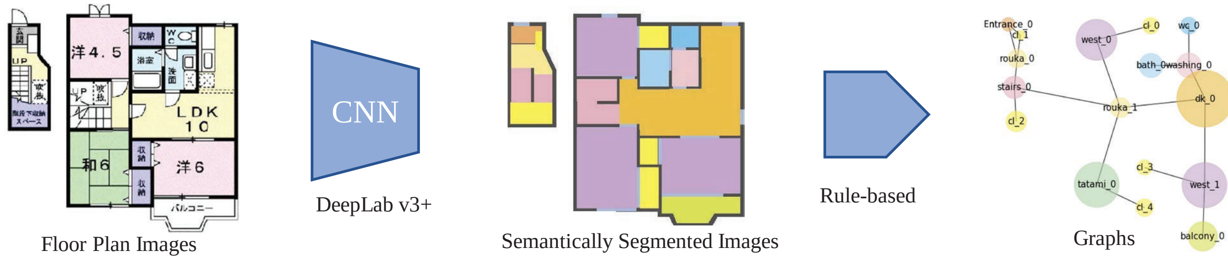


A 2-D mapping of graphs using Multidimensional Scaling (MDS) and Maximum Common Subgraph (MCS) as a similarity measure. The colors of markers are based on the "Total Score" of each floor plan image. Some examples of floor plan images and graphs are located with arrows pointing to the map's corresponding markers.

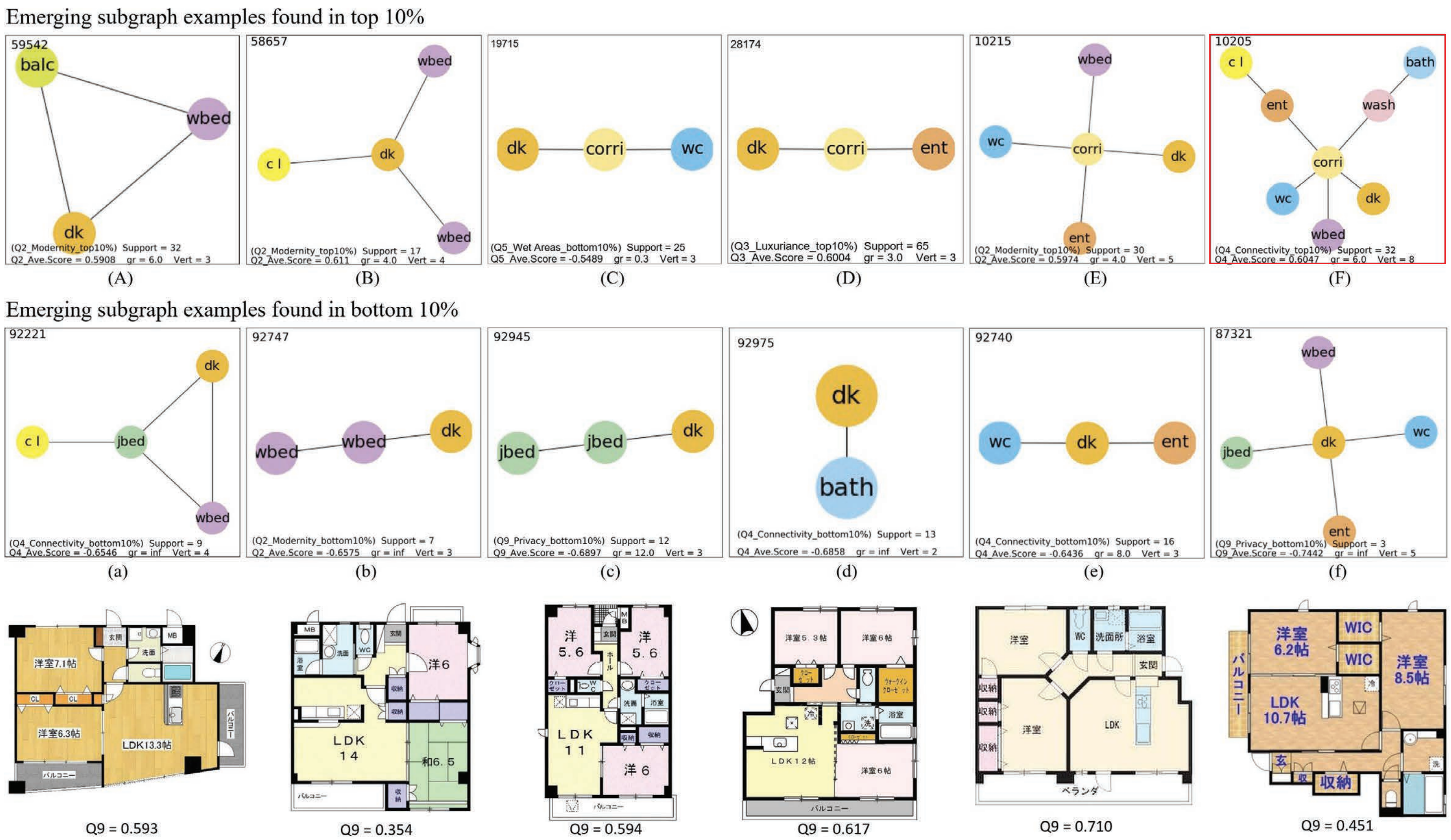
PROJECT VIDEO

Segmentation and Graph Extraction:

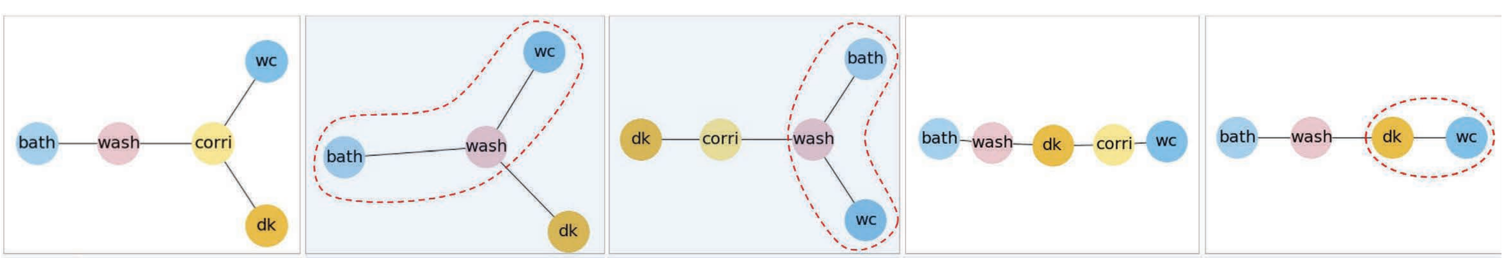
PROJECT VIDEO



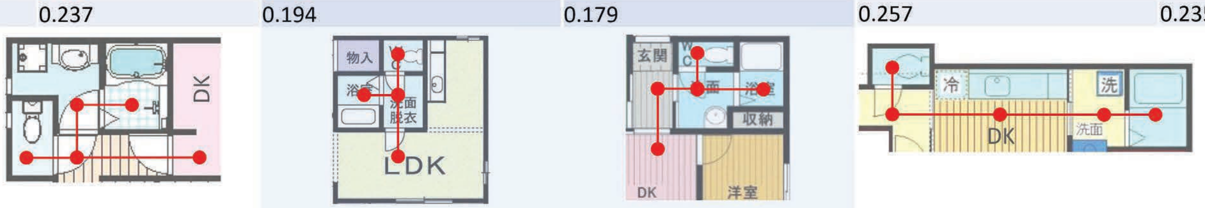
To automatically extract corresponding graphs from floor plan images, we first trained a segmentation prediction network in DeepLab v3+ using a manually annotated dataset that was color-coded and semantically segmented into 15 room types. Second, we used the rule-based method to extract 1,000 graphs from the inferred segmented images automatically.



Examples of emerging subgraphs frequently found in floor plans. Subgraphs on the top row have high average scores in classes indicated in each, and those in the middle row have low average scores. Each example shows numbers of support, average score, growth rate (gr), and vertices. The top left (A) appears also in Figure 1a (circled in red) having a high Q2 (modernity) score. The top right subgraph (F) is included in all floor plans in the bottom row with high Q9 (privacy) scores.

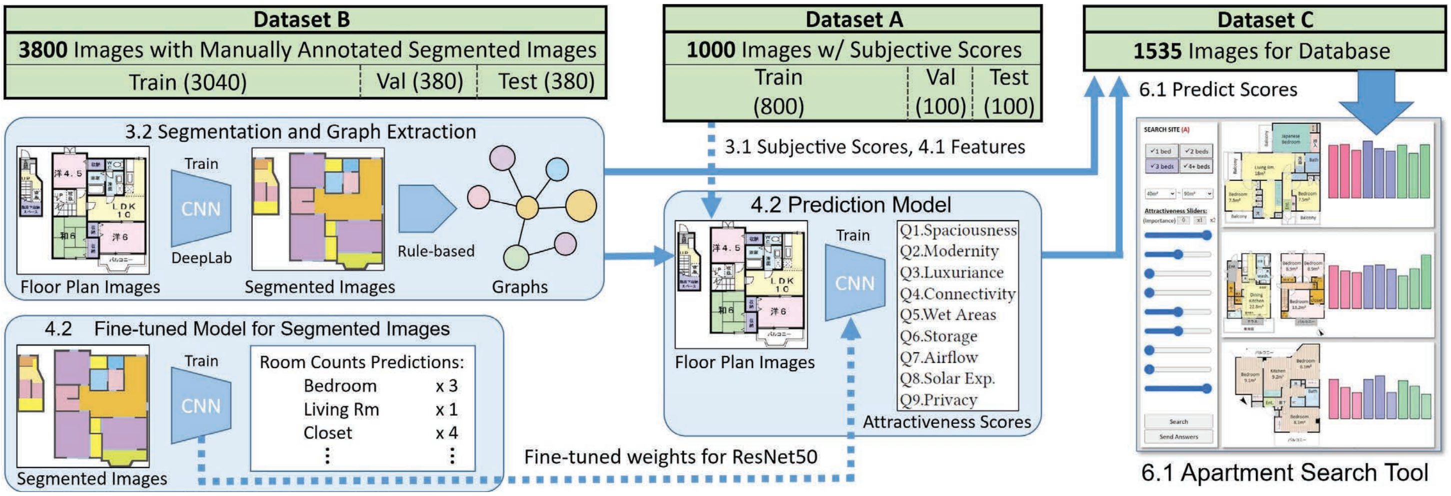


count	308	189	85	80	68
mean	0.330	0.208	0.171	0.296	0.249
Std.	0.237	0.194	0.179	0.257	0.235

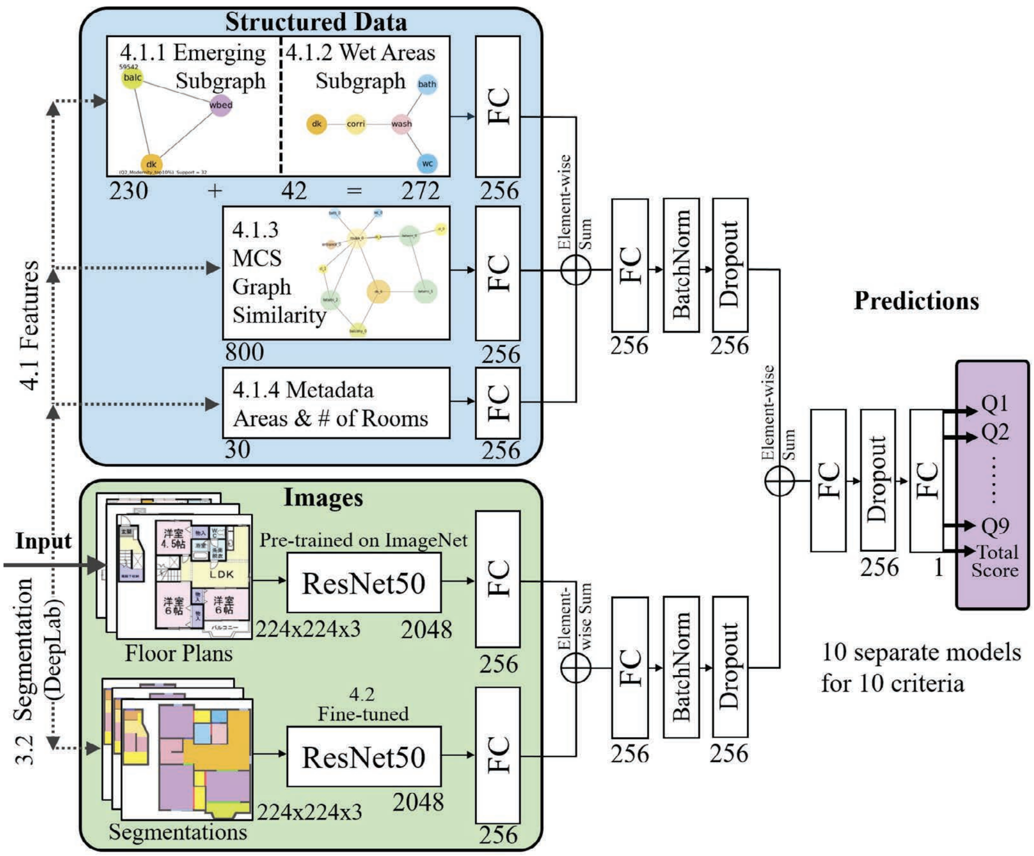


The main five types of subgraphs for wet areas that make up 76% of all floor plans. Two subgraphs that include the linkage (wc)– (wash)– (bath), circled in red, have lower mean scores for Q5 than the others.

Additional figures from my recent research:



An overview of our framework. Dataset A was used for the dataset with subjective scores and to train our prediction model. Dataset B was used to obtain the network that outputs segmented images from the floor plan images as input. We predicted subjective scores by our model from a separate set of floor plan images in Dataset C which was used for our proposed apartment search tool.



Our proposed multimodal neural network-based framework to predict subjective apartment scores via their floor plan images, graph representations, and various metadata (see the paper for details) [IEEE Xplore, PDF].

Kitabayashi, R., Narahara, T., & Yamasaki, T., Graph Neural Network Based Living Comfort Prediction Using Real Estate Floor Plan Images, short paper. [ACM, MM Asia'22].

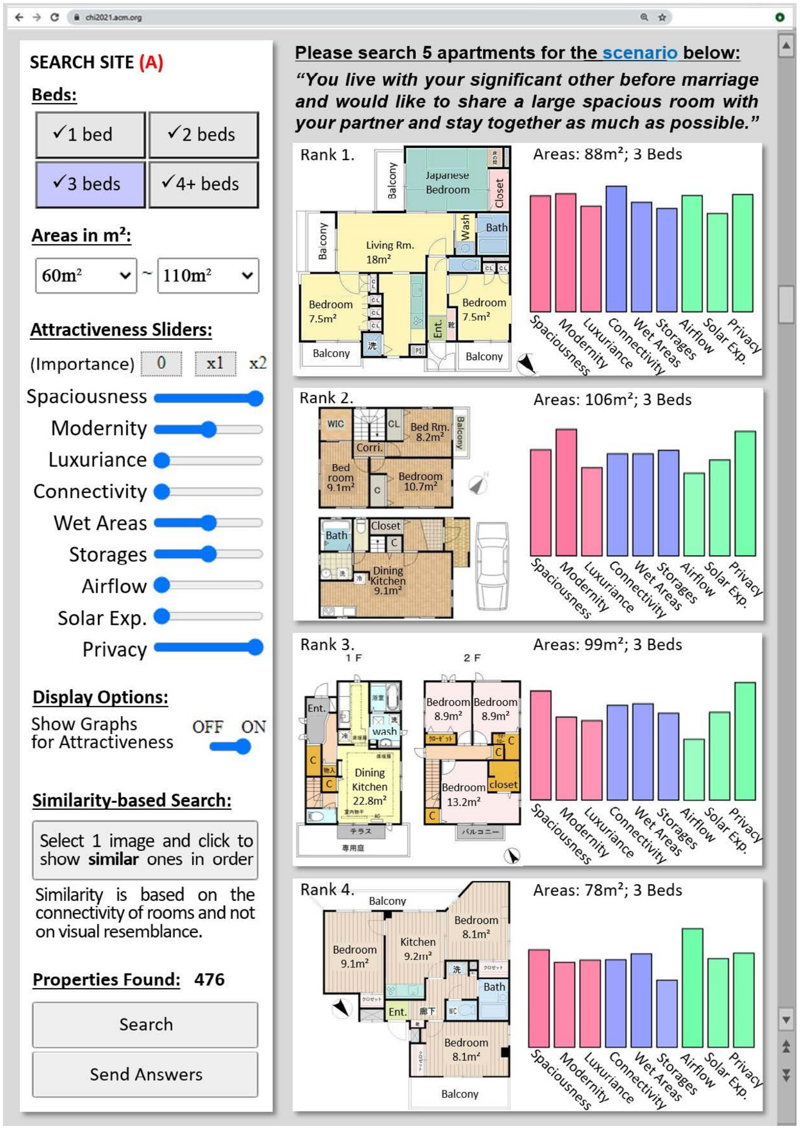
Our recent results using **Graph Neural Networks (GNNs)**, a type of neural network that can directly handle graphs more specifically, **Gated Graph Neural Networks (GGNN)**, shows prediction accuracies close to my original method using emerging subgraph features.

Model Features w/o	w/o SubG	w/o MCS	w/o Meta	w/o Img	w/o Segm	IV-C Prop.
Q1. Spaciousness	0.669	0.730	0.707	0.637	0.648	0.721
Q2. Modernity	0.782	0.787	0.772	0.771	0.765	0.793
Q3. Luxuriance	0.762	0.753	0.751	0.749	0.747	0.776
Q4. Connectivity	0.542	0.620	0.592	0.546	0.571	0.637
Q5. Wet Areas	0.477	0.452	0.422	0.439	0.499	0.525
Q6. Storage	0.715	0.751	0.728	0.713	0.733	0.751
Q7. Airflow	0.494	0.514	0.573	0.446	0.592	0.607
Q8. Solar Exp.	0.520	0.571	0.531	0.525	0.564	0.591
Q9. Privacy	0.764	0.780	0.786	0.791	0.748	0.816
Total Score	0.727	0.740	0.763	0.712	0.755	0.794
Average	0.645	0.670	0.662	0.633	0.662	0.701

The experimental results show that we can predict the scores with a correlation coefficient of 0.701 on average. Our proposed model outperformed all of the comparative baseline models which lacks one of the input features from the proposed.

Additional figures from my recent research:
(Confidential/Unpublished Materials)

Click this link to watch a project video.



The figure shows the proposed apartment search tool that can query the functionality and comfort using importance sliders with adjustable weights in three levels for nine subjective criteria in addition to a common search interface based on user selection of the number of bedrooms and a range of areas.

TABLE II
THE DIRECT COMPARISON QUESTIONS WERE ASKED ON A 5-POINT LIKERT SCALE. A HIGHER SCORE INDICATED A PREFERENCE FOR OUR PROPOSED TOOL, WHILE A LOWER SCORE INDICATED A PREFERENCE FOR THE BASELINE TOOL. A SCORE OF 3 INDICATED NO PREFERENCE.

Question	Mean	CI	p-value
Which tool helped you find more desirable floor plans?	4.03*	[3.90, 4.17]	4.9×10^{-34} < 0.001
Which tool did you enjoy better while searching?	3.97*	[3.83, 4.11]	1.6×10^{-30} < 0.001
Which tool was faster for you to search floor plans?	3.84*	[3.67, 4.01]	1.1×10^{-18} < 0.001
Which tool was easier to search apartments?	3.76*	[3.60, 3.93]	2.5×10^{-16} < 0.001
Which tool was more intuitive for you to search floor plans?	3.43*	[3.25, 3.62]	7.0×10^{-6} < 0.001

* Significantly different based on 95% confidence interval.

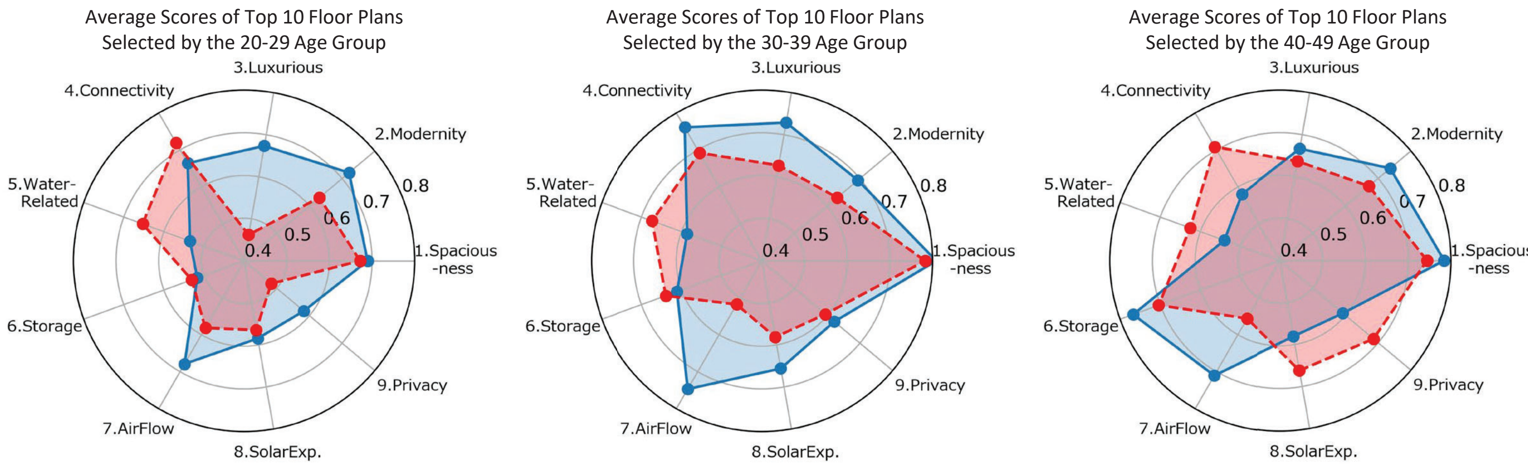
TABLE III
THE FIVE-GRADE RATINGS OF ALL FLOOR PLANS SELECTED BY PARTICIPANTS USING TWO METHODS

	Baseline	Proposed	p-value
Mean scores of the 5-grade ratings of all floor plans selected	3.75	4.01	1.67×10^{-40} < 0.001

A t-test was conducted on two related scores; the average scores of five floor plans for five scenarios from all participants using both tools.

Our user study showed that our proposed tool could provide a better user experience compared to a baseline tool with the common search interface without the importance sliders.

Example of analysis of floor plan dataset based on user attributes and their preferences



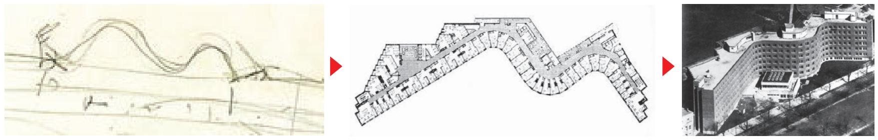
Floor plans selected from the top 1 to 7 by generation, in order from top to bottom in each column.

Radar charts show the average value of each attractiveness item out of top 10 floor plans for each generation, which indicates that there are differences in floor plan preferences among generations.

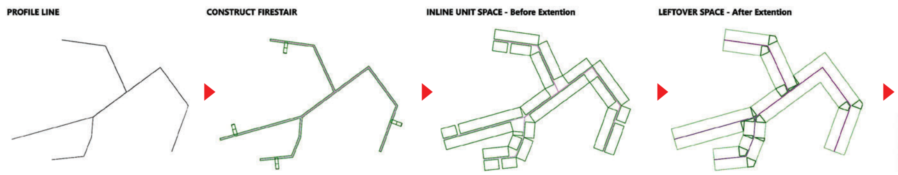
Narahara T, Wang X, Yamasaki T., “Construction and Analysis of a Large-scale Attractiveness Dataset for Real Estate Floor Plans,” Proceedings of the Annual Conference of JSAI 2020 (The Japanese Society for Artificial Intelligence) (0), 2P6GS1305, 2020, [Link, PDF].

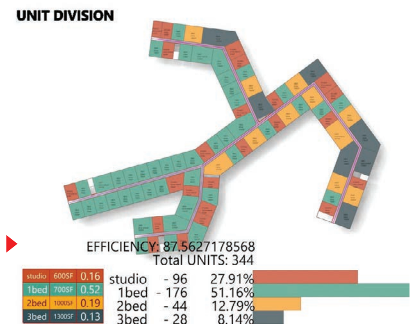
Sketch to Build: An Intuitive Design Platform for Sustainable Housing Complexes [Demo Video]
Zhang Z, and Narahara T., Annual Modeling and Simulation Conference (ANNSIM 2022), July 19, 2022. [IEEE Xplore, PDF]

Initial sketches influence the final built form. MIT Baker House, Cambridge, MA by Aalto. A sketch (left: 1947), A photo (right; 1949).

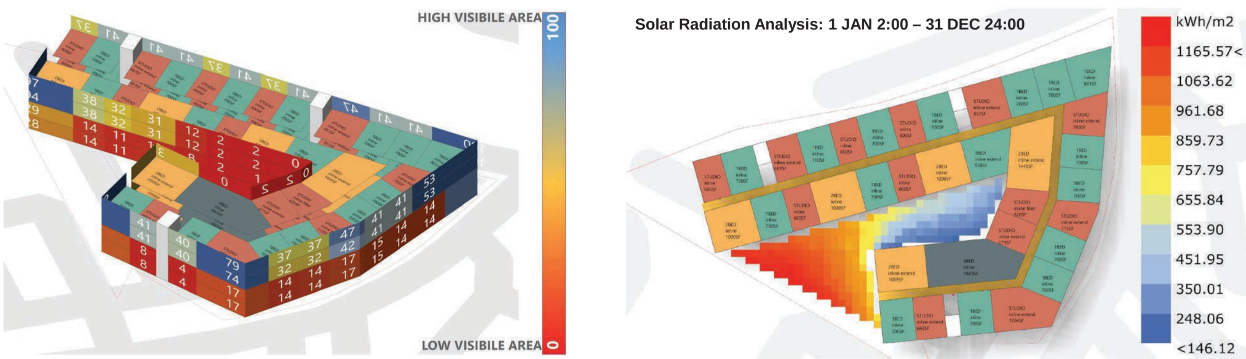


Sequential steps to generate a floor plan from a user-drawn sketch (left) to the final floor plan layout (right).

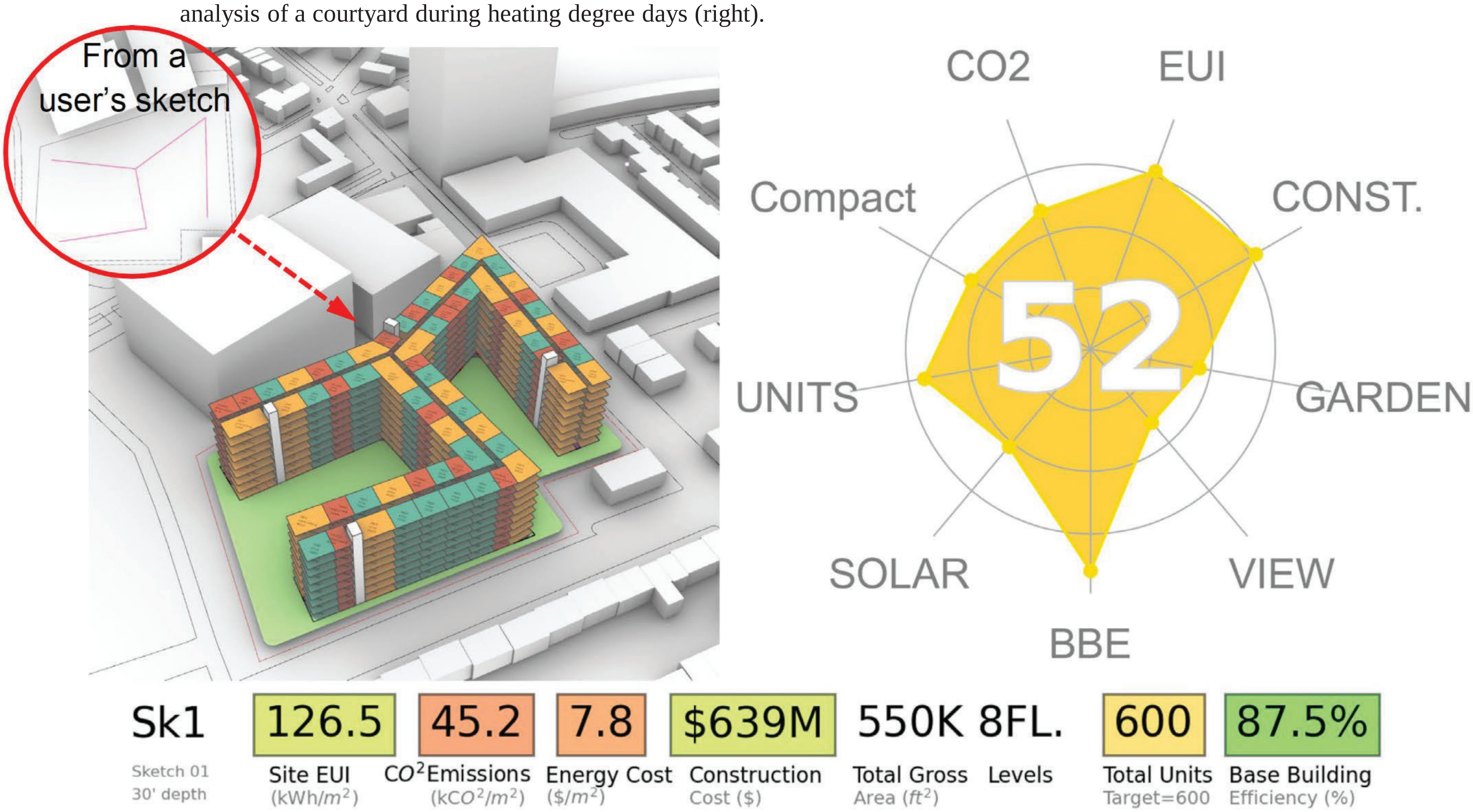




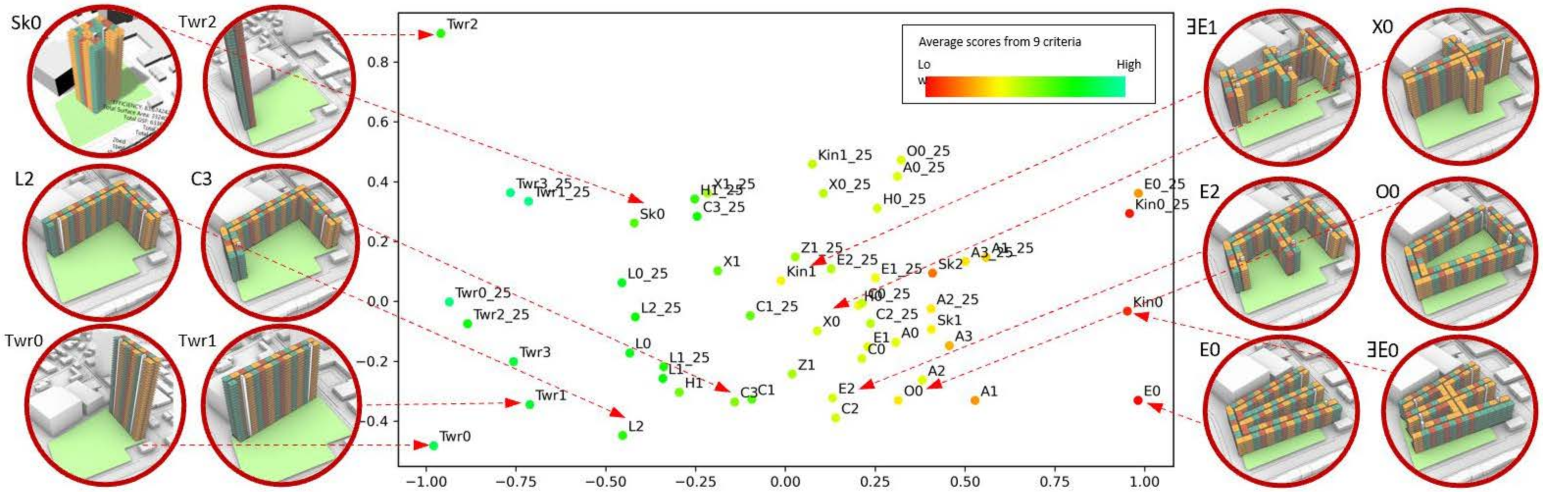
Abstract: Today, there is a growing demand for housing complexes due to rapid urbanization in metropolitan areas. While architects must meet new sustainability standards, they are expected to demonstrate creative solutions for humanizing mass housing for the well-being of residents. This paper proposes an intuitive platform for users to visually study possible housing designs and their potential performance in energy use intensity (EUI), environmental, and some financial criteria based on preliminary sketches drawn by users. Before users start sketching, our program auto-generates basic layouts with performance results. With this knowledge, users will be able to visually grasp intrinsic relationships between built forms and performance characteristics and reflect on their new design. Our goal is to provide a platform that enables designers to effectively incorporate qualitative contributions from early exploratory stages into advanced design stages, allowing architects to focus on creativity.



View Analysis based on 2D Isovist for each unit with a normalized score ranging from 0 to 100 (left), solar radiation analysis of a courtyard during heating degree days (right).



Examples of generated performance evaluation datasets using our program's auto-generated pre-defined built forms. Each example shows a perspective of a building layout (left), a radar chart for nine criteria with an overall score (right), and values from the analysis used for the radar chart prior to normalization (bottom). (Notes: The results are based on the test site in the temperate climate zone on the East Coast of the United States, with the conditions set based on 600 units with percentages for unit types of 20% studio, 36% 1-bedroom, and 44% 2-bedroom apartments.)



PCA visualizations of 52 layouts based on their performance scores in nine criteria from radar charts as their features. The color of a marker is based on the average score of all criteria from each layout (green being higher and red being lower).

SPATIAL ANALYTICS OF HOUSING PRICES WITH USER-GENERATED POI DATA, A CASE STUDY IN SHENZHEN

by Muxin Jia (Ph.D. in Urban Systems, 2020-), Dr. Taro Narahara (Dissertation Adviser)

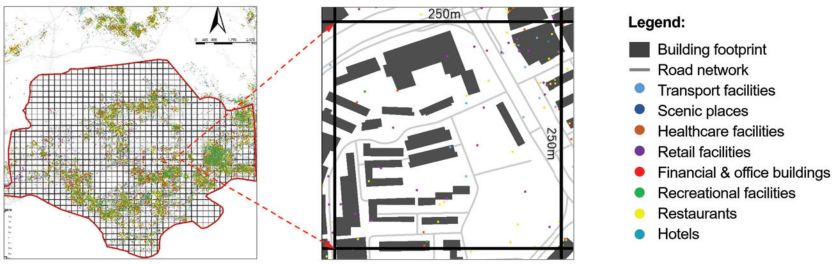


Figure 1. Spatial distribution of eight categories of user-generated POIs.

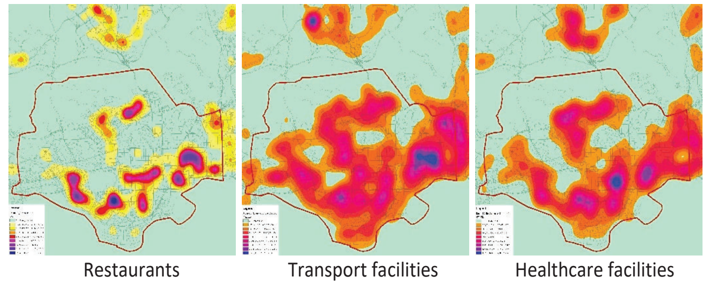
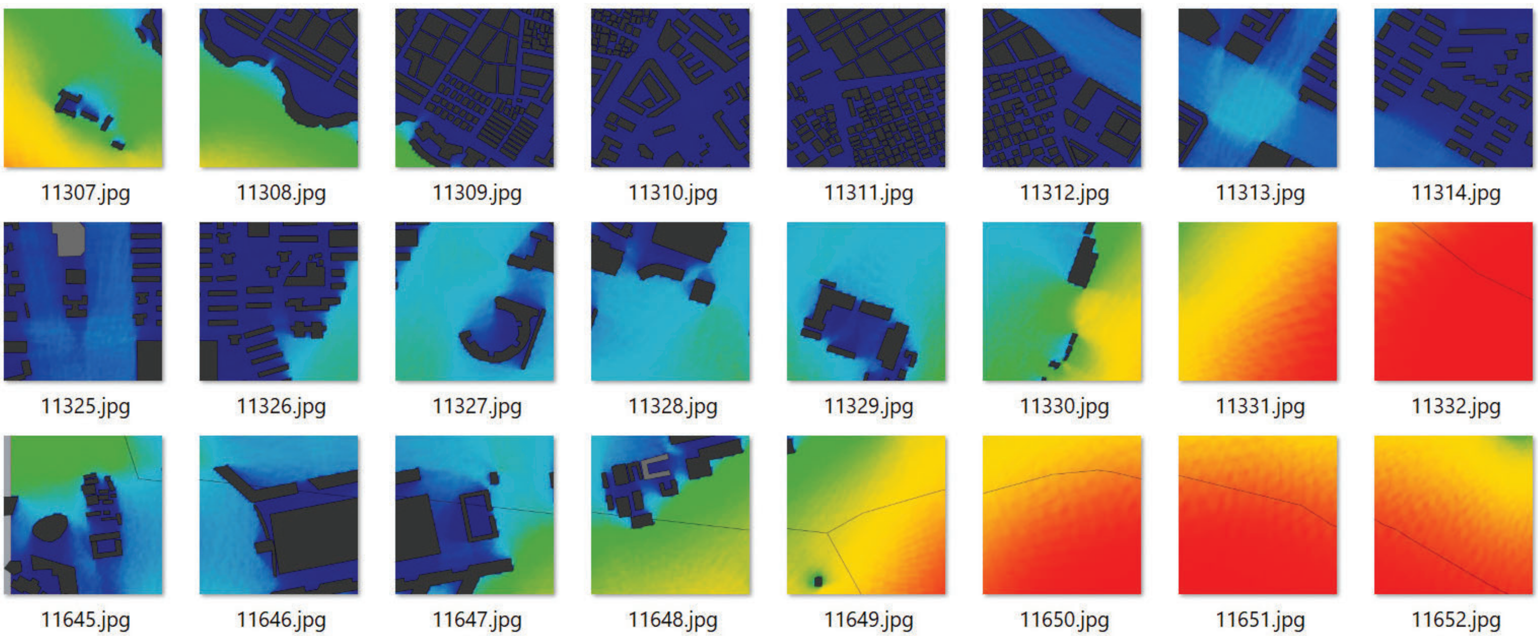
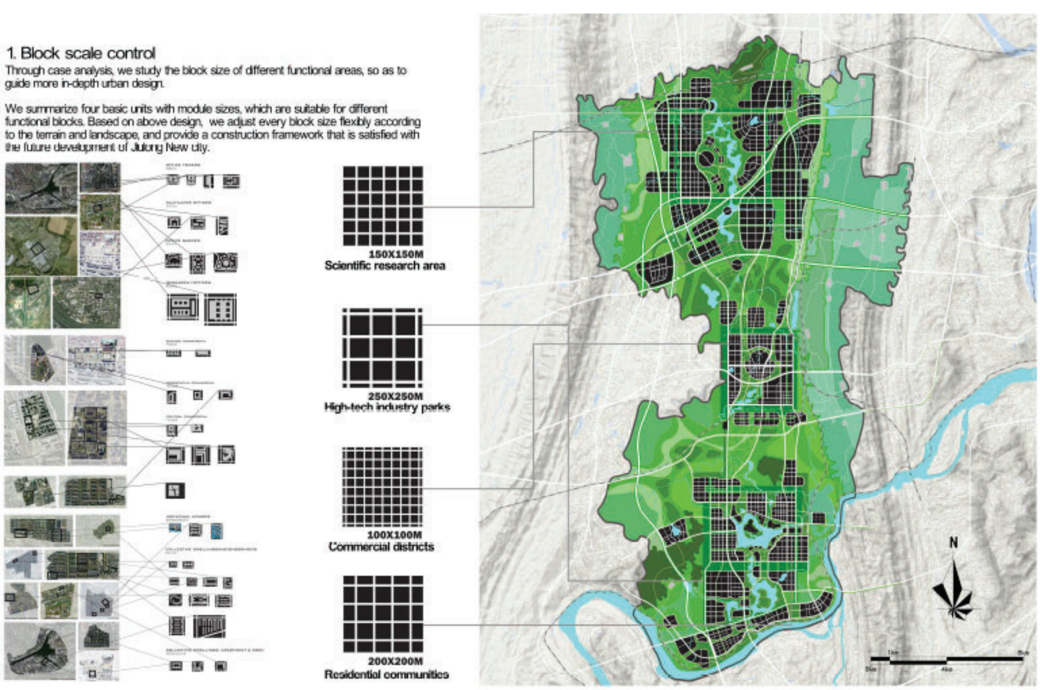
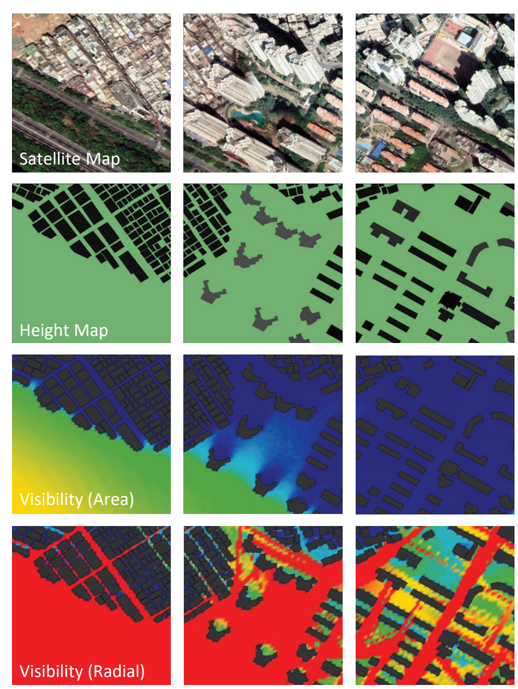


Figure 3. Heatmaps of eight types of check-in POIs

Housing is among the most pressing issues in China. Researchers are eager to identify housing property's internal and geographic factors influencing residential property prices. However, few studies have examined the relationship between social media users' **point of interest (POI) data** and house prices using big data. This paper presents a machine learning model for regression analysis to reveal the relationship between housing prices and check-in POI density in Futian District, Shenzhen. The results show that our proposed price prediction model using additional features based on POI data proved to provide higher prediction accuracy. Our results indicate that incorporating POI features based on current feeds from location-based social networks can provide more up-to-date estimates of housing market price trends [CAADRIA'23 PDF].







Multimodal datasets containing satellite images, building height maps, i**sovist visibility maps**, and various metadata are currently used as inputs for **multimodal deep neural networks** to forecast real estate prices and popularity of places, including point of interest data.

CHARACTERIZING RESIDENTIAL BUILDING PATTERNS IN HIGH-DENSITY CITIES USING GRAPH CONVOLUTIONAL NEURAL NETWORKS Muxin Jia (Ph.D. in Urban Systems, 2020-), Dr. Taro Narahara (Dissertation Adviser) [CAADRIA2024]

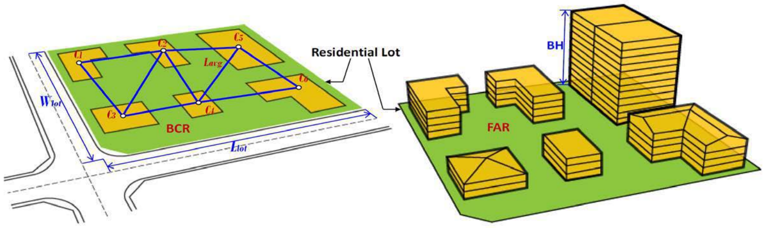


Figure 1. Illustration of indices for the clustering buildings at the block scale.

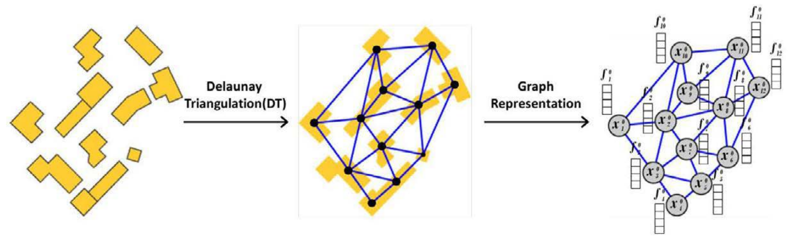
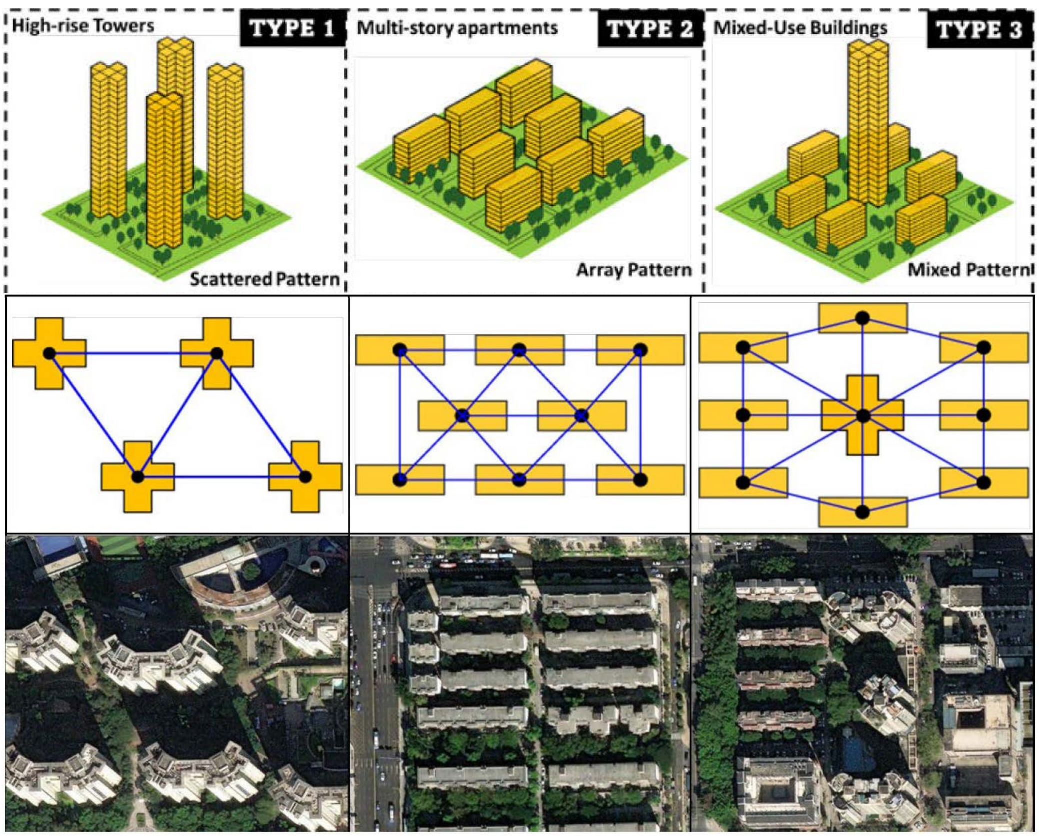


Figure 2. Graph representation of a building group based on Delaunay Triangulation.

In urban morphology studies, accurately classifying residential building patterns is crucial for informed zoning and urban design guidelines. While machine learning, particularly neural networks, has been widely applied to urban form taxonomy, most studies focus on grid-like data from street-view images or satellite imagery. Our paper provides a novel framework for graph classification by extracting features of clustering buildings at different scales and training a spectral-based GCN model on graph-structured data. Furthermore, from the perspective of urban designers, we put forward corresponding design strategies for different building patterns through data visualization and scenario analysis. The findings indicate that GCN has a good performance and generalization ability in identifying residential building patterns, and this framework can aid urban designers or planners in decision-making for diverse urban environments in Asia.



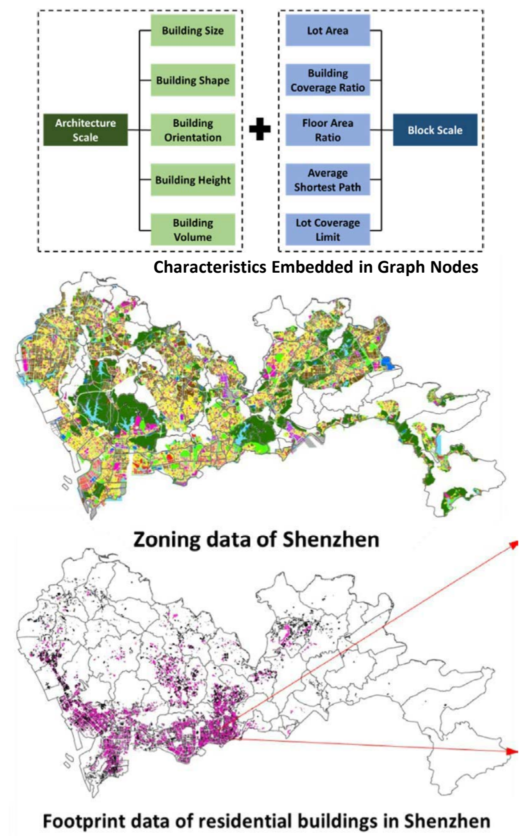
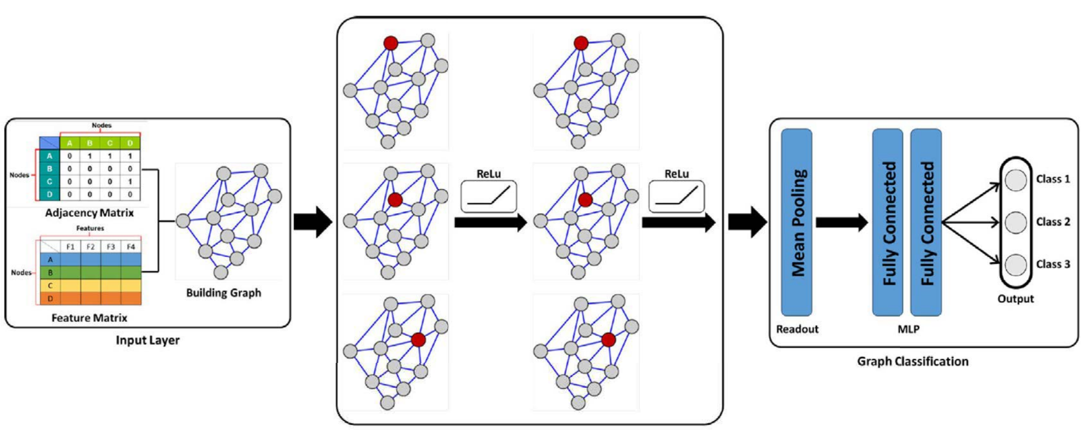


Figure 5. Three typical types of residential building layout in China.



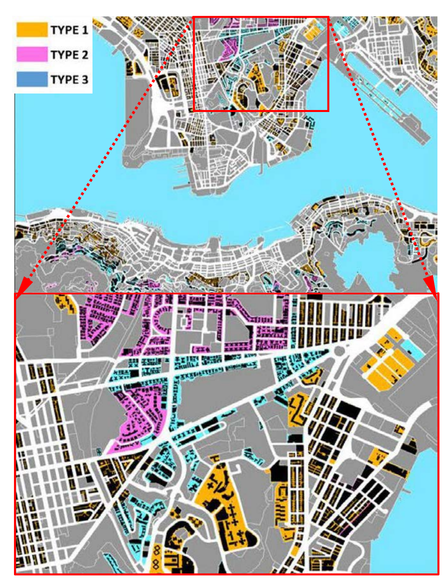


Figure 6. Proposed network architecture using GCN for predicting 3 types of urban conditions.

Figure 9. Visualization of prediction result of Hong Kong dataset (right). The prediction accuracy of 93.77% was achieved using the model trained on Shenzhen data, which has 94.91% accuracy based on its test data.

Kurashiki Viewer: Qualitative Evaluations of Architectural Spaces Inside Virtual Reality, [PDF, Link to a Demo].
Narahara T., International Conference for The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2022.

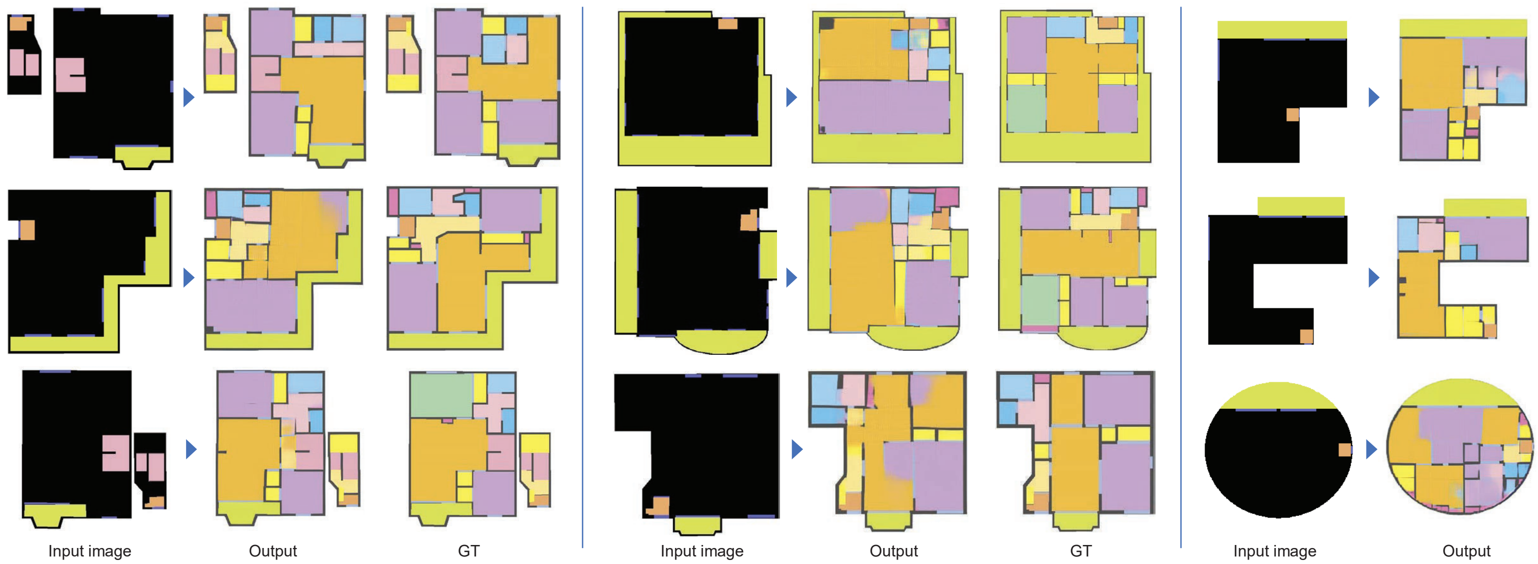
Abstract. This paper discusses how virtual reality (VR) environments can be employed as a data collection tool beyond visualization and representation. Using a VR model that runs on a web browser based on an existing historic town in Japan called Kurashiki, the experiment asked 30 recruited participants to freely walk around and leave ratings on a 5-point scale on any buildings or objects appealing to them. The proposed system can display points of interest using heatmaps on a map visualizing statistical preferences among them. The project provides a quantitative means for qualitative values of architectural and urban spaces, making such data more shareable. Such a platform could help multiple stakeholders reach better consensuses and possibly collect training datasets for machine learning models that could extract features related to the attractiveness in architecture and urban spaces.

ML Application Possibilities:

From cloud-based models, datasets on perceptive values evaluated by online participants can be used for training networks using Bounding Box Regressors, such as Fast R-CNN to predict estimated popularity

A Comparative Study of Data-driven Approaches for the Generation of Floor Plans in Japanese Apartments

Narahara, T., Wang, X., and Yamsaki, T., The Tenth International Workshop on Image Media Quality and its Applications [IMQA 2020].



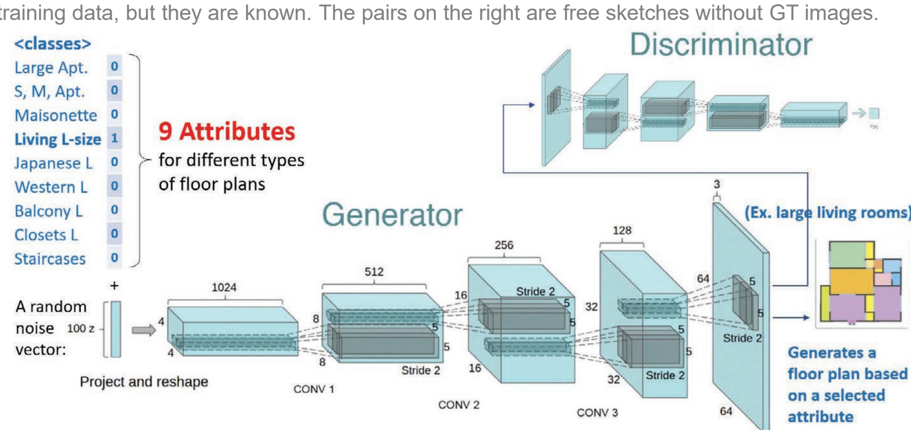
A pair of datasets (input and ground truth (GT) images of floor plans) were used to train the pix2pixHD network so that the empty areas of input images (with four initial elements only) can be filled with possible architectural plans (outputs). The GT images for the three pairs on the left are included in the training data. The GT images for the three pairs in the middle are not included in the training data, but they are known. The pairs on the right are free sketches without GT images.

Color-coded floor plans obtained using semantic segmentation and graph extraction (left column). Synthetic variations created by applying cGAN on the color-coded and categorized floor plan images (right column). Plans on each column represent the attributes indicated at the top row (e.g. the red column has more living rooms with a Japanese tatami mat floor).

Table 1: Comparison of Different Methods.

Network	*DCGAN*	*DCGAN*	*pix2pixHD*	*cGAN*
Dataset images	Raw images from the Home's dataset	Color-coded semantically segmented images	Color-coded semantically segmented images	Color-coded semantically segmented images
Image quality	- Low	✓ Medium *	✓✓ High	✓ Medium*
Generate diverse outputs based on selected attributes	- No	- No	- No	✓ Yes
Applicable to selected regions (building footprints)	- No	- No	✓ Yes	- No
Connectivity and adjacency among rooms are considered	- No	✓ Medium	✓ Medium	✓ Medium

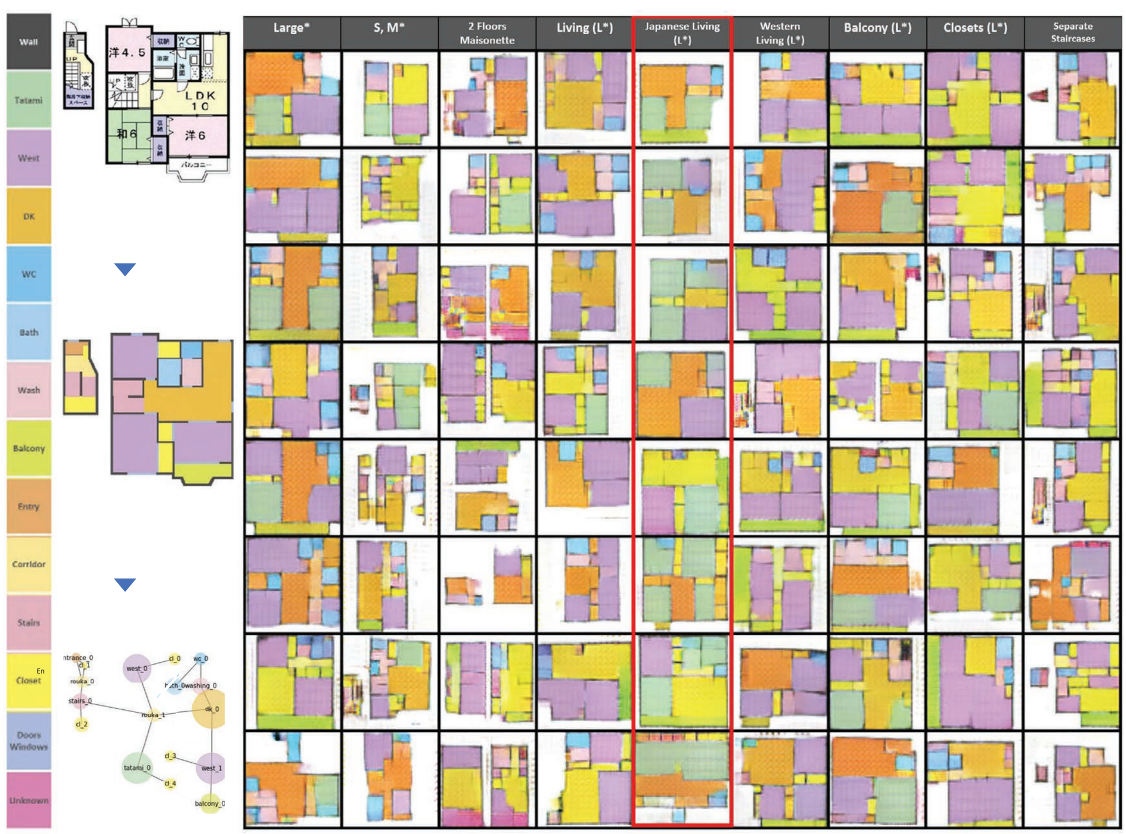
*can be improved with the post-processing with the use of pix2pixHD together



Architecture of conditional GAN for generating specific floor plans based on user-specified attribute.

[2018-2019: My earlier work using Deep Neural Networks]

Abstract: Recent approaches based on generative adversarial networks (GANs) have synthesized photorealistic facial expressions with user-specified attributes using big data of human faces. In this study, we investigate the feasibility of these techniques for generating architectural floor plans using the dataset of the floor plans of Japanese apartments. We have compared four approaches based on different GANs, including our new approach based on conditional GAN, which can generate floor plans with user-selected attributes such as the relative size of a specific room. We observed that the use of floor-plan images with segmentation labels yields better results than that obtained by using raw images from the dataset. Further, we found that the use of pixel- wise image inputs only can occasionally lead to the generation of dysfunctional floor layouts. Therefore, we believe that the use of graph structures for the floor plans is a promising approach for future studies.



Narahara, T., "Megastructure: Past, Present, and Future", Architecture in the Age of the 4th Industrial Revolution, The eCAADe + SIGraDi 2019 Conference, 2019, p637 - 644. [PDF, Video]

Megastructures: Past, Present, and Future

Taro Narahara[1]
[1]*New Jersey Institute of Technology*
[1]*taronarahara@gmail.com*

During the '60s, theorists and architects such as Yona Friedman proposed visions for megastructures where residents could freely come and build their units with individual variations. However, there were no technological means to build such structures, and these visions appeared to be unrealistic at the time. This paper discusses how those visions could be re-envisioned through the use of anticipated new technologies and speculates about possible structures and their impacts on our living.

Keywords: *Metabolism, modular systems, space elevators*

1. BACKGROUND

Since more than half a century ago, architects and theorists such as Yona Friedman have proposed ideas for mobile architecture such that the inhabitant should be the sole conceiver of his own living premises within a structure that would allow individual variations (Friedman 2016).

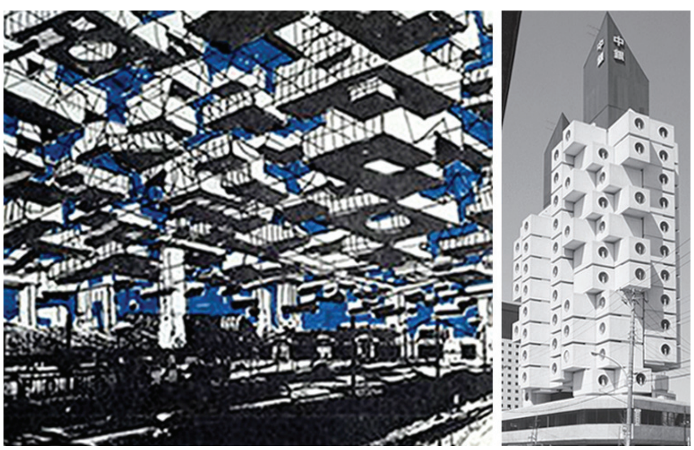


Figure 1: The Spatial City by Yona Friedman, 1959 (left) and Nakagin capsule tower by Kurokawa, Tokyo, 1971 (right).

Friedman sketched floating space-frame-like superstructures over existing cities that provide flexibility for inhabitants to construct their dwellings freely while maintaining physical integrity of the community. His vision appeared to be unrealistic, as there were no technological means for suspension of such structures or transportation of individual dwelling units (Figure 1). During the '60s, megastructures with pluggable prefabricated pods on the infrastructural core, proposed by Japanese Metabolists such as Kisho Kurokawa (Figure 1), had a practical limitation due to their need for transportability for reconfiguration and the obsolescence of their infrastructural systems for adaptation (Yatsuka and Yoshimatsu 1997).

However, anticipated new technologies for the 21st century-hyper-strength materials such as carbon nanotubes; space elevators, wireless communications, and energy transfer; autonomous drones and artificial intelligence-could realize a transformative vision for a new kind of living for future generations. The paper introduces possible ideas for speculative structures for living supported by anticipated new technologies and discusses how new possibilities for visualization and validation can help us envision such structures that have yet to exist.

2. EMERGING TECHNOLOGIES

Following in the footsteps of these predecessors, this paper introduces some possible visions for flexible

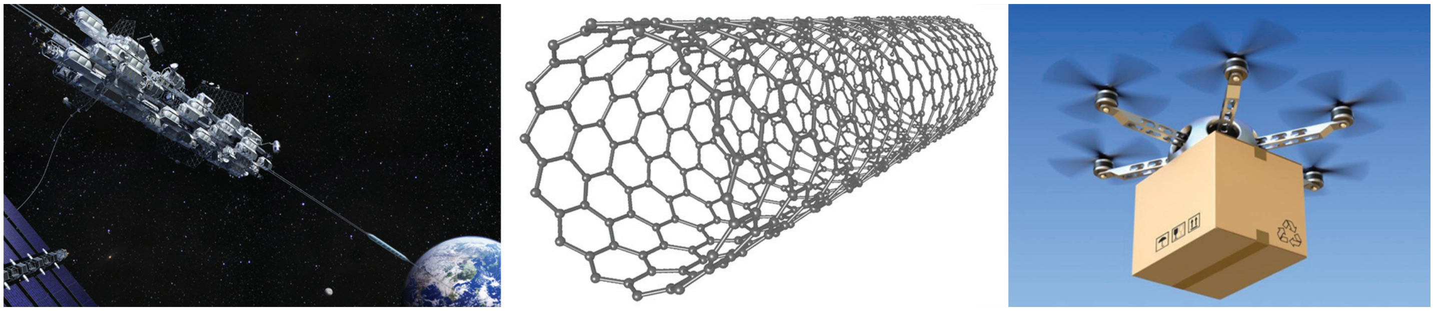


Figure 2: Obayashi Corporation's space elevator (Ishikawa 2016) (left), Carbon nanotube (middle), Amazon's drone delivery [1] (right).

habitable structures that can be constructed using innovative technologies and ideas that had yet to be fully introduced in their times. Some theoretical physicists foresee that the invention and realization of a space elevator could provide gravity-free spatial structures supported by the balance between the centrifugal force and gravity of the earth and connected through space elevators. The introduced visions will be based on the premise that high-strength lightweight materials such as carbon nanotubes will become available for manufacturing the required length of a space elevator (Figure 3). In theory, such a gravity-free structure, with auxiliary uplift support from helium-based high-altitude solar platforms, would establish a second "ground" for future residents of Earth without harming natural and artificial resources on the existing ground (Figure 3).

In fact, the Japanese general construction company Obayashi Corporation has announced that they will have a space elevator constructed by 2050 due to the advances in carbon nanotubes (Ishikawa 2016), and witnessing such structures might not be too far away from our time. The elevator will reach 96,000km into space and will transport people and cargo to a new space station. Unlike in the era of the Metabolists in the '60s, transfer of resources, including energy, can be done wirelessly using the conductivity of carbon nanotubes or laser power beaming, which reduces the heavy reliance on infrastructures that has been preventing faster updating of systems. Metabolists did not modularize mechanical, electrical, and plumbing systems for their infrastructural cores, unlike the way they did for their replaceable housing units. The cores that structurally and mechanically support housing units were not replaceable and updatable. Thus, the life spans of their buildings became shorter than expected. Luna Ring, proposed by Shimizu Corporation, has demonstrated schematic ideas for power generation using a ring of solar cells around the moon and transmission of the power using microwave laser beams to earth (Shimizu 2009). In terms of innovations on mobility, the SkyPod project by PLP Architecture has demonstrated an innovative idea to externalize and free up vertical transportation for ultra-tall buildings (Hesselgren et al. 2018), and some companies, including Uber, are developing drones that can be used for daily transportation. These proposals by professional practitioners in recent years indicate that those ideas that were once regarded as sci-fi stories have become feasible for planning. The paper further speculates about what is possible based on speculative yet thoroughly scientific (not sci-fi) studies.

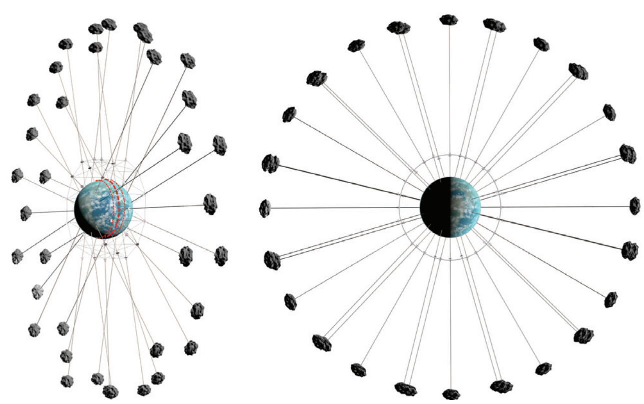


Figure 3: Space elevators constructed repeatedly around the equator at equal intervals inspired by Polyakov (1977) (diagrams not in scale)

3. PROPOSED STRUCTURES

This paper introduces speculative ideas for habitable modular pods that can be aggregated, rearranged, and disconnected based on each resident's needs, and that can be transported and reconfigured using advanced autonomous drones working as assemblers of the whole system (Figures 7). Inventions of hyper-strength materials could suspend portions of structures and allow them to use more lightweight materials for members only under tension. Thus, more dynamic reconfigurations of pods can be done.

Technical Details

As stated in the previous section, space elevators using carbon nanotubes could provide gravity-free spatial structures supported by the balance between the centrifugal force and gravity of the earth. In theory, these structures can be constructed repeatedly around the equator at equal intervals (Polyakov 1977). It is also known that the location of these base stations could be offset from the equator up to 35-degree latitude in north and south, which could allow us to construct arrays of space elevators in several loops around the earth, braced diagonally, forming further stable 3-D structures (Figure 3). The current estimated payload for such a space elevator is roughly 20 tons each (Edwards and Ragan 2006) (Beletsky and Levin 1993) and, with additional uplift from helium-based high-altitude solar platforms, in principle, it is possible to suspend lightweight habitable modules on such structures (Figure 4).

Such above-ground structures could provide additional areas for possible habitations and harvesting of resources without harming natural and artificial resources on the existing ground. As some theorists in the 60's such as Friedman suggested, floating structures could be a way to respond to issues such as the exponential increase in populations and pollution in the future. The appropriate altitude for such structures' height location needs to be carefully considered based on the environmental conditions and the proximity to other key stations above and below the residents. Silica fiber tiles, an effective insulator that was used for the Space Shuttle's thermal protection

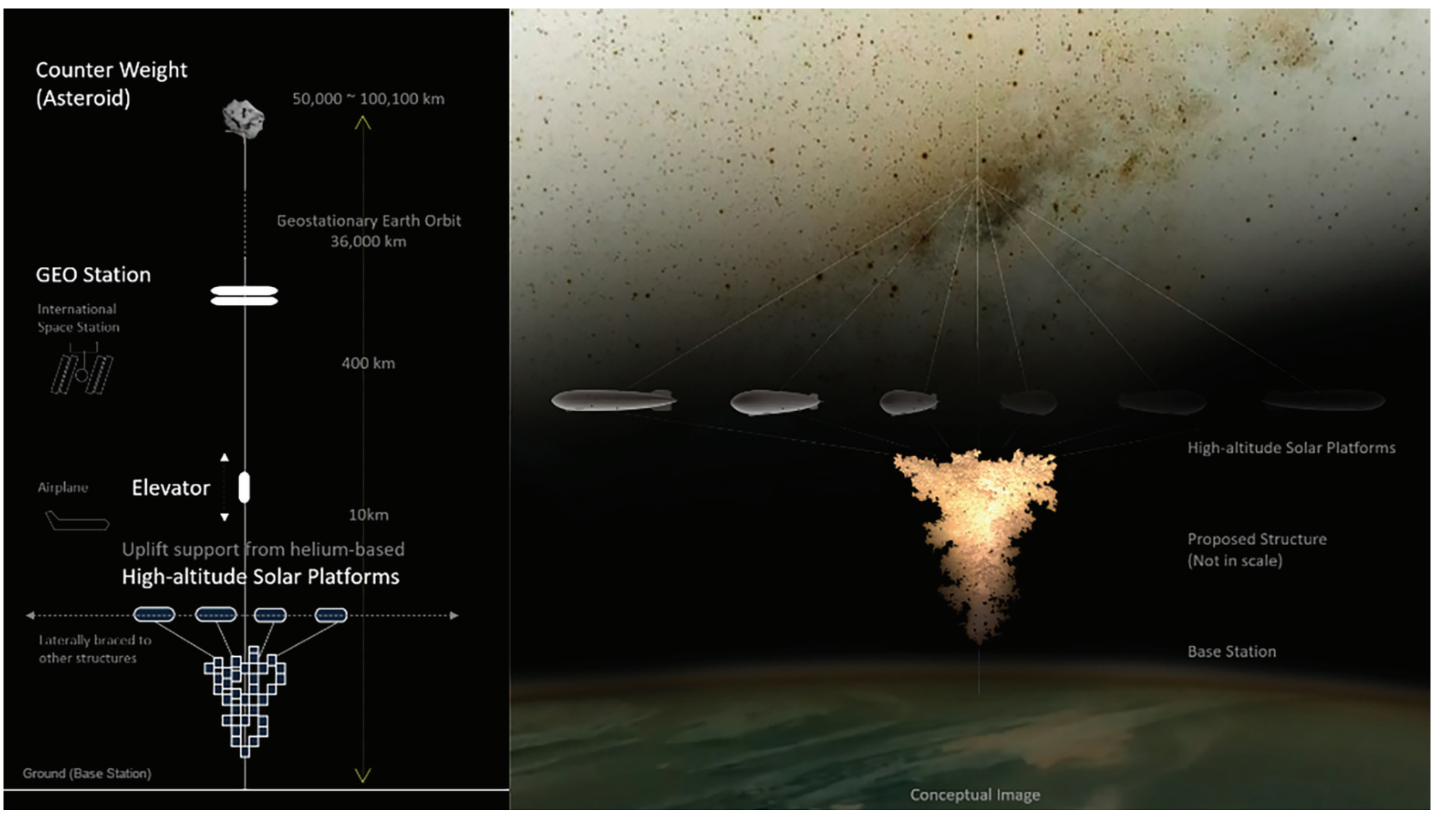


Figure 4: Proposed structures using a space elevator

Figure 5: Proposed structures using a space elevator and a reconfigurable module transported by a drone

system, could be used to tolerate anticipated acute temperature differences for living up in the air at different altitudes. They are indicated with rough textures in Figure 11 and prevent heat transfer.

The composite material using carbon nanotubes could be theoretically at least ten times stronger than steel and half the weight of aluminum if successfully manufactured and provides necessary properties for such lightweight structures. The paper proposes a 9m-cube module with approximately 180m² habitable space which can house one to two families. Each unit has six operable square openings at the middle of all six faces, and they can be connected through the operable mechanical joints that provide air-tight physical connections among adjacent units with sufficient structural strength. Modules can be attached to the space elevators' platforms and can form clusters by attaching themselves to each other and to suspension cables from the space elevators. Modules have internal 3-D corridors that internally connect residents and circulate physical resources such as fresh water and air using internal omnidirectional transportation cabs, while other non-physical resources such as communications and energy can be wirelessly transported-for example, using laser-power beaming technology (Shimizu 2009). Worth noting here is that there is no infrastructural core, unlike the Metabolists' buildings, and infrastructural functions are instead integrated inside the units, which allows for more variations for clusters' configurations.

Autonomous drones would allow us to live, work, and travel anywhere we like by functioning as transporters and assemblers of the habitable mobile units that can be nested to the floating structures (Figure 7). Tiltrotor drones with a crane and grippers, which are already near existing technologies, as seen from Amazon's drone delivery (Amazon.com, Inc. 2019) and Bell Nexus (Bell Helicopter Textron Inc. 2018), can move and plug our units into the best locations based on our preferences, proximity to elevator platforms, environmental factors such as solar radiation,

Figure 6: Possible ideas for speculative structures for living supported by anticipated new technologies.

structural stability, and so on. Each module can be an origami-like foldable structure made of hinged carbon nanotube composite panels covered by inflatable multi-layer Ethylene tetrafluoroethylene (ETFE) for further rapid reconfigurations for multiple units. A single drone can transport multiple modules to the site in a folded state and can unfold them on-site using its gripper and crane for the installation. ETFE is a lighter material for glazing than glass, robust, and a good insulator. Multiple drones can work collectively to transport larger clusters to realize faster reconfigurations (Figure 10).

Visualization and validation methods

Visualizing, validating, and communicating such speculative visions and ideas could be quite challenging and time-consuming if it required the use of multiple separate application platforms. However, the recent development of game engines such as Unity 3D and Unreal Engine using real-time shaders allows us to create and explore unbuilt and unseen speculative structures almost on the fly inside visually stunning immersive environments without requiring us to spend hours on rendering frame by frame. Such production platforms allow for quick iterations and reshaping of three-dimensional constructs and enable us to virtually walk through inside of them by becoming avatars. Component-based design, commonly practiced in game design, allows for procedural generation of structures composed of building blocks with certain behaviors, characteristics, and logic through custom scripts (Figure 8). For example, aggregations of modules by incoming inhabitants were visually assimilated using a Diffusion-Limited Aggregation (DLA) algorithm with custom cellular rules to provide enough open faces for each unit (for details on algorithms, see Narahara 2010). The structure's structural stability was studied computationally using custom codes in C# by the author to visually represent its deformation. Although it was not fully implemented inside a game engine, at the very schematic level, the stability of the structure could also be estimated using computational physics features inside game engines with custom scripts.

Figure 7: A reconfigurable module transported by a drone

5. CONCLUDING REMARKS

The author wanted to illustrate a possible vision where the architecture could be built based on decisions by multiple individuals. As the pattern language shows (Alexander et al. 1977), having commonly identifiable and sharable design elements is an approach often practiced by researchers in order to ease collective design activities by multiple participants. In this proposed vision, the modular unit with identical joints has become a common interface to connect to neighbors. This conceptual vision may appear to be too systematic and dull for some readers. However, the system can have multiple types of modules with different functionalities and visual variations. Also, by connecting multiple modules, we can introduce functionalities beyond what a single module can serve for, for example, larger footprints for connected modules can provide communal spaces such as multi-purpose halls, conference rooms, and roof gardens.

In a society with the anticipated technologies, we are not bound to a specific location for living anymore, and there is no fixed resulting form for the proposed clustering structures. The autonomous drones can transport our living spaces anywhere while maintaining necessary virtual connections to conduct business tasks and various involvements with others in remote locations. The proposed structures' configurations can be constantly changed over time based on multiple individuals' collective decisions. In such

Figure 8: Procedural generations of structures (right), and Dynamic structural behaviors (left).

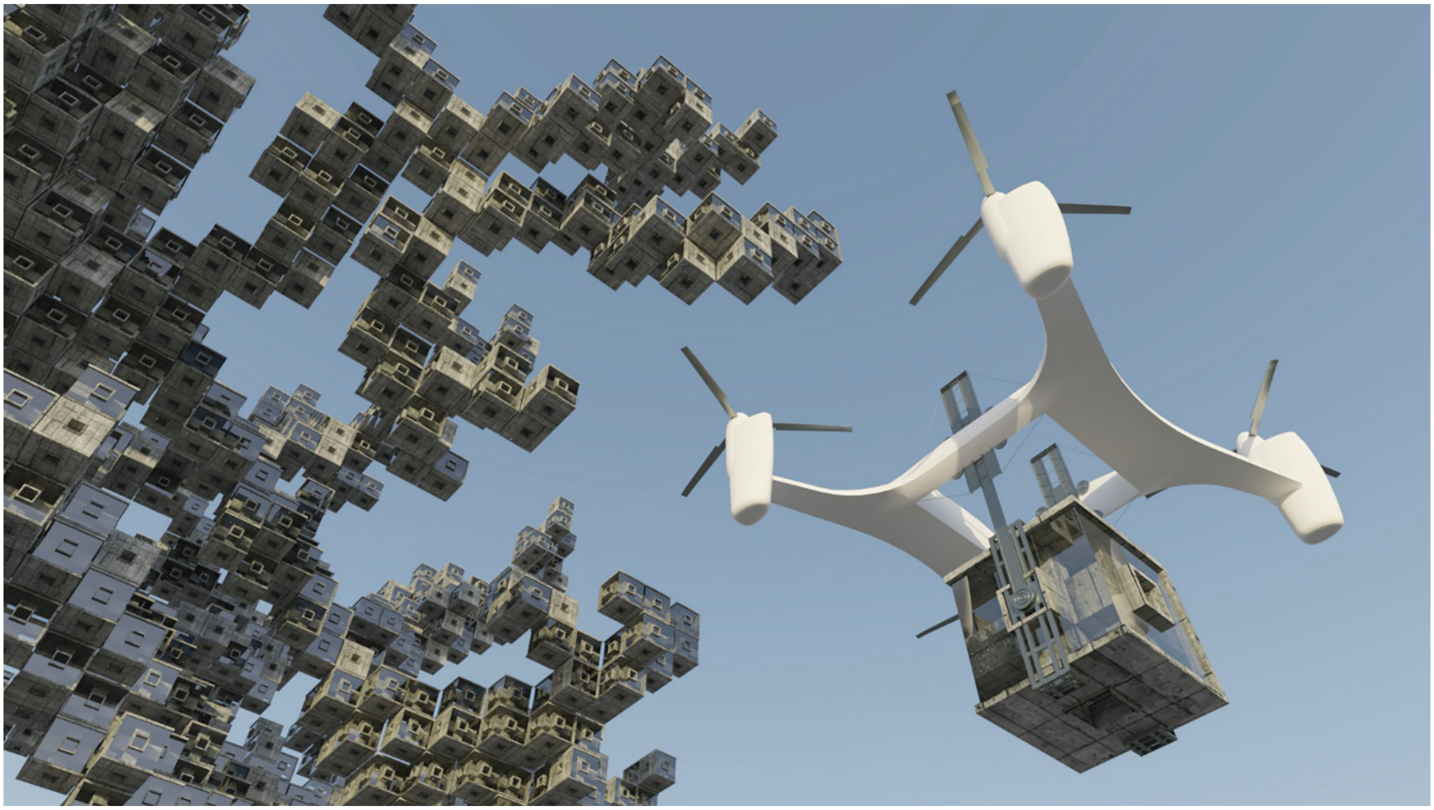


Figure 9: Reconfigurable modules transported by a drone.

hypothetical circumstances, we will develop and acquire very different lifestyles, values, ethics, and social structures. It might take many years to witness those changes. However, as a designer and architect, the author's intention was to take a small yet positive step toward these society-changing technologies by investigating feasibilities of such hypothetical structures based on speculative scientific (yet not sci-fi) studies.

Through the reinterpretation of utopian visions offered by theorists in the 60's, what we expect to see as architecture in the near future could be something much more indeterminate in terms of a physical form - formless and ephemeral with the possible use of anticipated technology. We could be virtually extremely closely connected with each other using the expected advances in digital communication technologies with a high level of augmentations for reality. Yet, using anticipated materials and engineering technologies, we could be simultaneously physically far apart and extremely differentiated from each other compared to the current standard. In a society where everything in the current physical world could be simulated and experienced virtually (including all human senses, possibly using brain-machine interfaces), what we expect to prioritize to have as an experience in the physical real world will depend on future residents of the earth beyond our imagina-

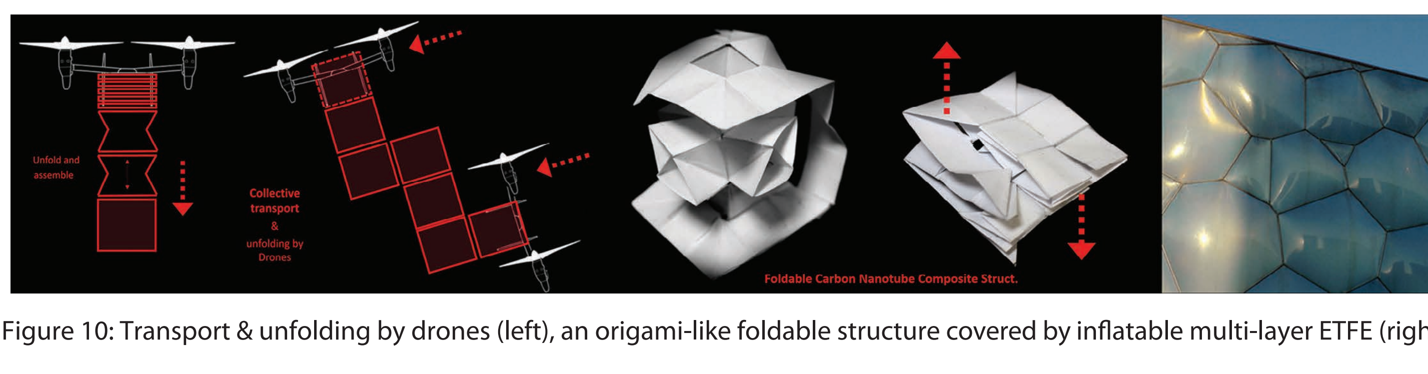


Figure 10: Transport & unfolding by drones (left), an origami-like foldable structure covered by inflatable multi-layer ETFE (right)

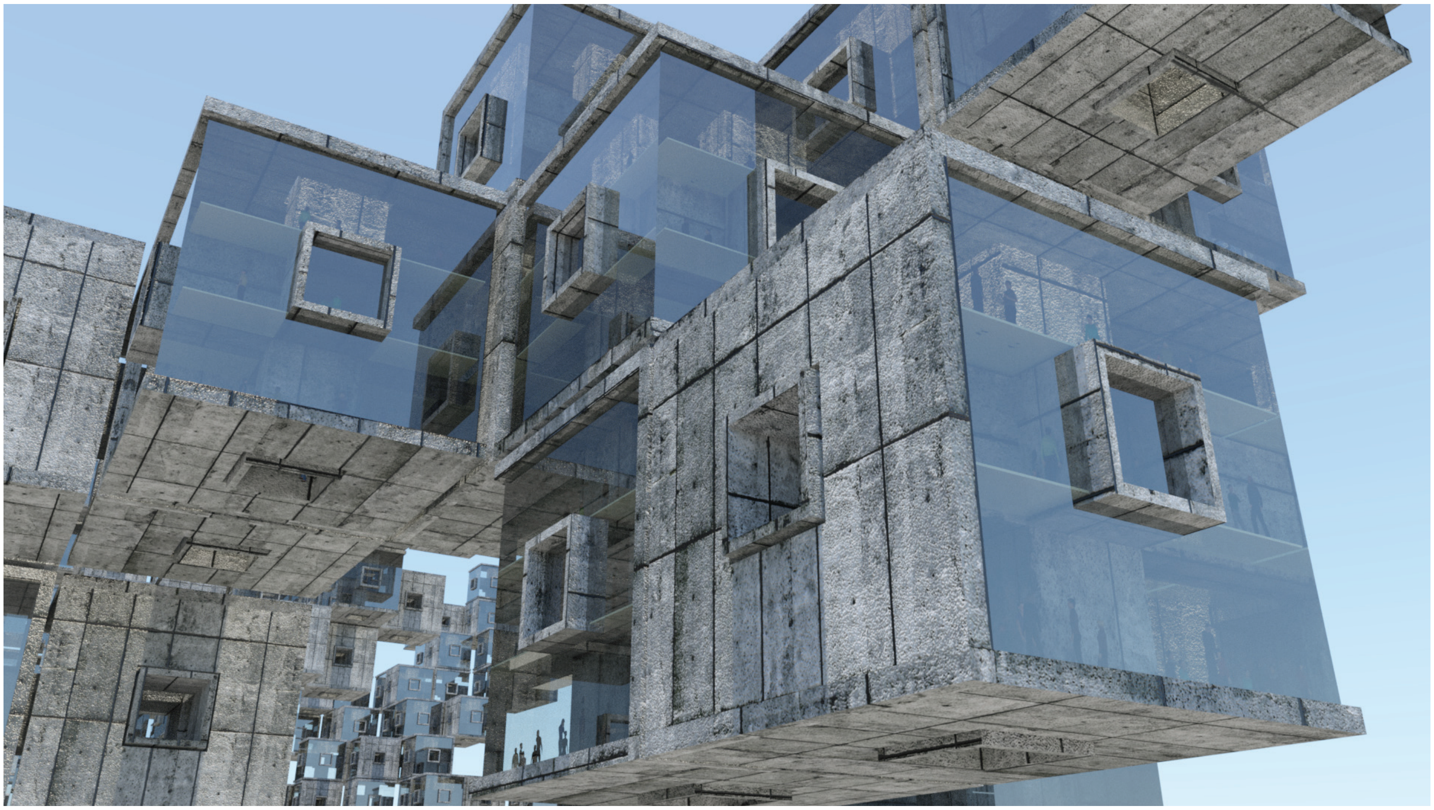


Figure 11: A view of the housing modules. Silica fiber tiles, an effective insulator indicated with rough textures.

tion. Conceptual visions created by our predecessors could be realized in unimaginable ways in a distant future, and they could be very different from the originally proposed vision. The author believes that the advances in and widespread use of visualization and validation tools will trigger an exponential increase in our creativity in the virtual domain, resulting in the realization of our physical constructs faster than ever.

REFERENCES

Alexander, C., Ishikawa, S. and Silverstein, M. 1977, *A Pattern Language: Towns, Buildings, Construction*, Oxford University Press, New York

Beletsky, V. and Levin, E. 1993, *Dynamics of Space Tether Systems*, An American Astronautical Society Publication, San Diego

Edwards, B. C. and Ragan, P. 2006, *Leaving the Planet by Space Elevator*, LuLu.com, Seattle

Friedman, Y. 2016, *Imaginary Cities*, Quodlibet, Macerata, Italy

Hesselgren, L., Aladag, D. and Cruise, R. 2018 'The Future of Integrated Transport in the Digital Age', *The 22nd International Congress on Vertical Transportation Technologies*, Berlin, pp. 276-289

Ishikawa, Y. 2016, 'Obayashi Corporation's space elevator construction concept', *Journal of the British Interplanetary Society*, 69(6/7), pp. 227-239

Narahara, T. 2010, *Self-organizing Computation: A Framework for Generative Approaches in Architectural Design*, Ph.D. Thesis, Doctor of Design, Harvard Graduate School of Design

Polyakov, G. 1977, 'An Earth's Space Necklace', *Technika-Molodyegi*, No. 4, pp. 41-43

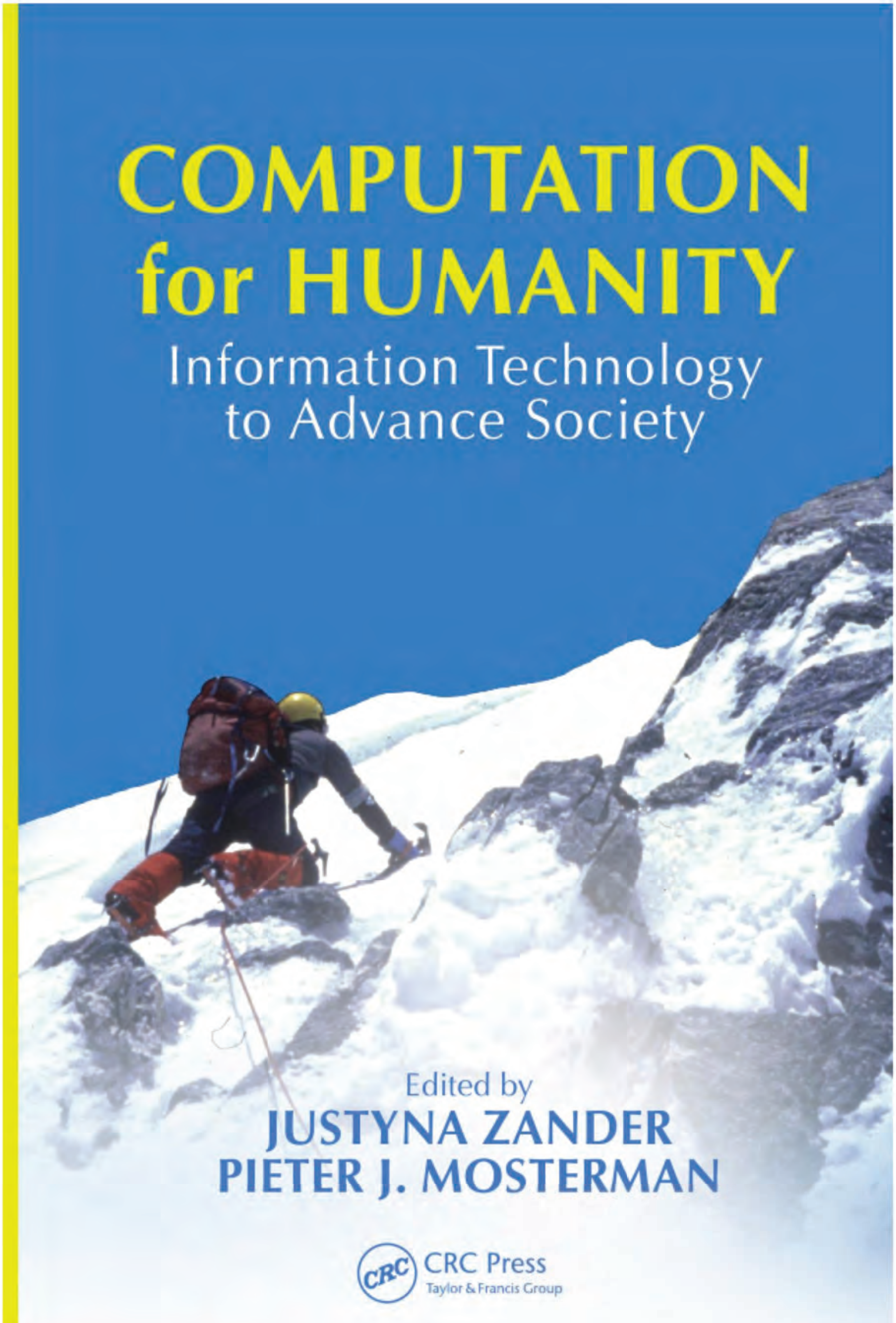
Yatsuka, H. and Yoshimatsu, H. 1997, *Metabolism: senkyūhyaku rokujū-nendai Nihon no kenchiku avangyarudo*, Inakkusu Shuppan, Tokyo

[1] Amazon.com,Inc.2019,AmazonPrimeAir,https://www.amazon.com/Amazon-Prime-Air

[2] BellHelicopterTextronInc.2018,https://www.bellflight.com/company/innovation/nexus

[3] ShimizuCorporation,2009,LUNARING,SolarPowerGenerationontheMoon,https://www.shimz.co.jp/en/topics/dream/content02/pdf/lunaring_e.pdf

Narahara, T. (2014). The computer as a tool for creative adaptation: Biologically inspired simulation for architecture and urban design. In Zander, J., & Mosterman, P. (Eds.), *Computation for Humanity: Information Technology to Advance Society* (1st ed., pp. 69-94). CRC Press, Taylor & Francis Group, LLC. (Peer-reviewed in 2 stages) https://doi.org/10.1201/9781315216751
[link, PDF, Website].



4

Computer as a Tool for Creative Adaptation: Biologically Inspired Simulation for Architecture and Urban Design

Taro Narahara

CONTENTS

4.1 Introduction

In today's design methodologies in architecture and urban design, we normally attempt to anticipate all current and future design requirements and potential changes for buildings prior to construction and endeavor to resolve all issues in a single (relatively static) solution. However, this solution may not always be able to respond to ongoing radical population growth and environmental changes. Moreover, the physical scale of buildings and the complexity involved in building programs have been increasing to unprecedented levels. In such conditions, increased use of process-based four-dimensional design strategies (space + time) can be anticipated.

This four-dimensional design thinking is not only promising for developing more flexible and adaptable architecture but also for expanding the territory of design to systems issues. This chapter investigates the potential of architecture and urban design to produce systems that can grow over time and introduces computer-based approaches in architecture with particular focus on computational emulation of natural processes in design through conceptual models. The main focus is not on providing immediate solution methods to resolve any specific professional problems in architecture but rather on investigating the emergent characteristics of these simple models' methods that can potentially evolve new design solutions over time and on showing how tools employing the methods can be used beyond passive evaluation and analysis tools for given architectural instances.

In recent years, many scientists have started to gain the advantages of self-organizing systems in nature through their computational models in areas such as telecommunication networks and robotics. The main advantages of such systems are robustness, flexibility, adaptability, concurrency, and distributedness. With limited space, the author has chosen to write about how this specific type of computation can influence architectural design methods. After a brief overview of computer-based methods in architecture, the feasibility and application area of extensible systems in architecture will be reviewed, and the chapter speculates as to the possibilities of open frameworks for design using computational methods through relatively simple yet explicit model examples.

4.2 Early Applications of Computation in Architecture and Urban Design

Early in the 1970s, one of the pioneers in the field, William Mitchell of MIT, foresaw the possibilities and applicability of computers to architectural design and set the agenda for computer-aided design (CAD) and its education (Mitchell 1977). Around the same time, Nicholas Negroponte founded MIT's Architecture Machine Group, which was later expanded into the MIT Media Lab. The group's original focus was to create an architecture machine to help users design buildings without architects (Negroponte 1970). This unique approach set the primary motive and foundation for the current development in architectural computing, often referred as *generative design*. Negroponte's group conceived automated architectural plan generator software based on parameter-inputs by users.

However, research in the field remained largely theoretical until the drastic increase in desktop computational power in the mid-1980s allowed for practical application by architects. In the 1970s, it was hard to draw a line in a computer; by the 1980s, many architecture firms were starting to adopt CAD software (such as AutoCAD) for the documentation of drawings from manual drafting. By the late 1990s in the United States, most of the drawings by architects, including construction documents and conceptual renderings, had been digitized. Since then, beyond merely being a representation and visualization tool, computation tools have been actively used for the evaluation of given instances in architecture in order to analyze their performance. Finite Element Analysis software has been widely used by engineers and even architects for structural analysis, and environmental analysis software is available for simulation and evaluation of day lighting, solar radiation, thermal performance, energy analysis, and so on (Autodesk Ecotect Analysis 2012).

More contemporary application areas of computer-based methods are building information technology (BIM) and generative design. A BIM is a process involving the generation and management of digital representation of a facility (Eastman 2008; NBIMSP 2012), and BIM tools provide a comprehensive 3-D model that is parametrically adjustable based on attributes such as construction cost, time, materials, and product-specific information from manufactures, without users having to reconstruct models every time a change in design occurs. This type of software is often also referred to as Parametric Modeling software among architects. In principle, 2-D construction drawings such as plans and sections can be instantly acquired by slicing the 3-D model, and the use of the latest crowd computing technologies starts to allow data sharing of building information. On the other hand, generative design has been widely used in the schematic phase of architectural design since the 2000s, after scripting capabilities in common CAD software matured (AutoLisp 2012; MEL 2012; Rhinoscript 2012). The method generates possible design iterations based on a set of rules or an algorithm, normally using a computer program, and one of the pioneering works by Stiny and Gips (1972) has developed into a Shape Grammar, which is a widely accepted approach within this area.

Today, many architects write their own project-based custom code for geometrical operations using scripting capabilities. The prime example of this new attitude—the architect simultaneously being a designer and toolmaker—is Frank Gehry, who developed his own technical consulting firm, Gehry Technologies (2012), and software, Digital Project (2012), to resolve complex issues on geometry and fabrication for his projects such as the Guggenheim Museum Bilbao (Shelden 2002). The firm and the tool contributed to rationalizing advanced geometry and construction issues for projects by others such as Beijing National Stadium, known as the Bird's Nest, by Herzog & deMeuron. In terms of application in a more social context, the Dutch firm MVRDV (1999, 2005) has actively incorporated computer programs to solve political and economic issues such as zoning, and their unique computational interfaces include a form of digital game with architects and mayors as its players. In academia, the Centre for Advanced Spatial Analysis (CASA) has done extensive research on city planning, policy, and architecture using digital technologies in geography and the built environment. The Space Syntax group (2012) has developed human-focused techniques for the analysis of spatial configurations using the connectivity of urban spaces based on cognitive factors such as visibility (Hillier 1999) and has served as consultant for important firms' projects, most recently the London Olympic Legacy design proposal. Lastly, much effort has been made to lay a scientific foundation for the field by scholars from various interdisciplinary areas, including Gero (2012).

In terms of more conceptual implications, recent theories of form in architecture have focused on computational methods of formal exploration and expression, and prominent works include Novak's "Computational Compositions" (1988), Mitchell's *Logic of Architecture* (1990), Eisenman's "Visions Unfolding" (1992), Frazer's *An Evolutionary Architecture* (1995), Lynn's *Animate Form* (1999), and Kostas Terzidis's *Algorithmic Architecture* (2006).

4.3 Complexity, Time, and Adaptation in Architecture and Urban Planning

One of the main concerns in architecture today is the increasing quantity of information to be processed during design and the level of complexity involved in most building projects. As globalization and economic development increase, large-scale urban development has become ever more essential. Complex threads of relationships among buildings and urban infrastructures are intertwined to produce inseparable connections. Dependencies among these structures are extremely intense, not only pragmatically but also aesthetically. Nearly half a century ago, Alexander (1964) already foresaw these conditions and stated the following:

> In any case, the culture that once was slow-moving, and allowed ample time for adaptation, cannot keep up with it. No sooner is adjustment of one kind begun than the culture takes a further turn and forces the adjustment in a new direction. No adjustment is ever finished. And the essential condition on the process—that it should in fact have time to reach its equilibrium—is violated.

Today most buildings and infrastructure designs require dynamic and collaborative engagements by multiple professionals, and a conventional knowledge-based approach alone may not be able to respond to emerging building types. For example, housing projects for thousands of people have been emerging in urban areas, and demands for planning and design of buildings with multiple occupancies and complex programs are becoming a challenge. The social impact of such buildings can completely alter

(a) (b)

FIGURE 4.1
In these high-rise residential towers in Hong Kong, housing thousands of people, identical floors are simply stacked one on top of another (a). Algorithmically optimized 3-D clusters with maximized opening areas (b). (From Narahara, T., Self-organizing computation: A framework for generative approaches to architectural design, Doctor of Design dissertation, Harvard University Graduate School of Design, Cambridge, MA, 2010.)

the behavioral dynamics and physical conditions of local environments and often leads to redefinitions of transportation systems and infrastructures at far greater scales. Moreover, lack of flexibility and adaptability to ever-changing environments has resulted in the necessity to carry out expensive and invasive operations of demolition and reconstruction.

For instance, conventional solutions for multifunctional large-scale complexes, high-rise office towers, and housing complexes are often to design a plan resolving all the problems within a single floor and simply stack one on top of another (Figure 4.1). This approach may satisfy a number of initial requirements on a temporary basis. However, there are always other potential spatial configurations worthy of investigation. Selecting optimal solutions that can accommodate unpredictable additions and renovations for future adaptation has become a difficult task for architects. Our ultimate goals for successful design have shifted to seeking solutions for satisfying more long-term needs in a flexible manner.

On the other hand, some buildings do not include a comprehensive solution from the outset for all the potential scenarios of the future. Instead, some of those buildings possess systems that allow them to adapt to future changes over time by altering their designs spontaneously based on simultaneous feedback from a number of simple entities (or agents) inside the system. These feedback systems can be effectively distributed to formulate globally satisfactory working solutions as a collective result. These methods do not always guarantee the best solution in a deterministic sense; however, they may prove effective where there is no deterministic and analytical means to derive solutions. As a natural consequence of adapting to radical population growth, sometimes these characteristics can be seen in low-cost housing developments in less regulated zones with no supervision by professionals (Figure 4.2). These developments from human designs do not provide

(a) (b)

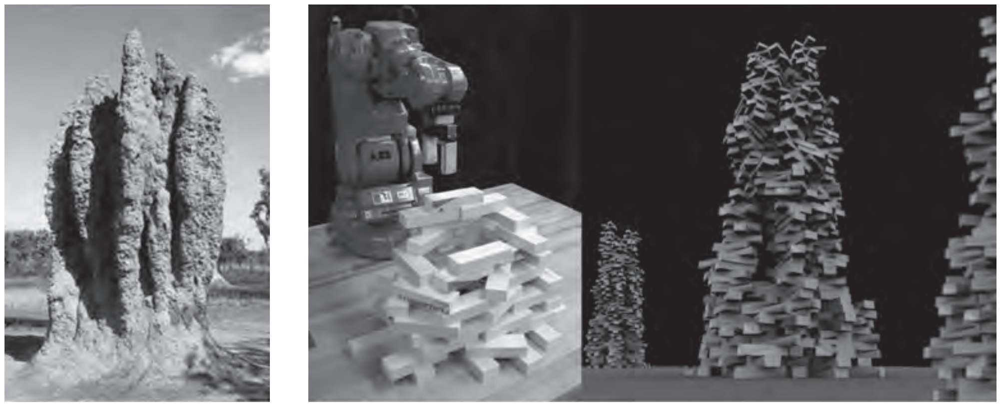
FIGURE 4.2
Extreme results of bottom-up design. (a) Favela in Rio today called Rocinha. (Photo by Ciaran O'Neill, http://irishabroad.blogspot.com) (b) Kowloon Walled City in 1994 before its demolition. (Photo by Dan Jacobson. Kowloon Walled City, Hong Kong, 1989, available at http://commons.wikimedia.org/ wiki/ File:Kowloon_Walled_City.jpg)

completely positive results, however, characteristics of dynamic adaptations seen in these examples suggest ideas for future computational implementations. This type of design approach is often referred to as *bottom-up* and is found in many natural systems.

As a promising means to respond to the aforementioned complexities in recent architecture culture, this chapter explores the active adoption of computational methods inspired by biological systems in the next section. Several pioneering works are reviewed with particular emphasis on computational emulation of natural processes in design, and the chapter, more specifically, investigates advantageous use of computation inspired by *self-organizing* systems.

4.4 Computational Emulation of Natural Processes in Design

Self-organization is a characteristic that can be found in systems of many natural organisms: flocking of birds, pigmentation of cells in animal skin patterns, and collective building behaviors by social animals and insects such as termites. These behaviors are often referred to as "emergent" behaviors, and *emergence* refers to "the way complex systems and patterns arise out of a multiplicity of relatively simple interactions" (Camazine et al. 2002). Original theories of self-organization, developed in the context of physics and chemistry, are defined as the emergence of macroscopic patterns out of processes and interactions defined at the microscopic level (Nicolis and Prigogine 1977; Haken 1983). Such systems' behaviors display many characteristics that are similar to the bottom-up approach in some artificial



(a) (b)

FIGURE 4.3
(a) Collective constructions by social insects (termites). (Photo taken and supplied by Brian Voon Yee Yap, Cathedral Termite Mounds in the Northern Territory, 2005: This image is from Dan Jacobson and is freely available at //commons.wikimedia.org/ wiki/ File:Termite_Cathedral_DSC03570.jpg under the creative commons cc-by-sa 3.0 license.) (b) experiment on construction automation by the author at Design Robotics Group at Harvard.

Not only in natural systems have we witnessed self-organizing growth processes, but also in some artificial systems. Beyond the scale of buildings, we have witnessed self-organizing growth processes in many formations of cities on large scales over long spans of time. Although results of these processes are not always successful in all aspects of design, the processes display heuristic and almost trial-and-error types of approaches that are robust and flexible enough to dynamically adapt to ever-changing environments. As the cities deliberately created by designers and planners rarely display the level of flexibility seen in these spontaneous city growth patterns, it is worth investigating their characteristics. In this area, Batty (2006) and Batty and Longley (1994) at CASA have published seminal work on urban growth modeling using algorithms such as fractals. Further rigorous computational reinterpretation and application of the principles underlying artificial self-organizing phenomena will possibly enhance and elaborate the advantages of these systems to a more practical level.

In recent years, many scientists have started to obtain the advantages of self-organizing systems in nature through their computational models (Bonabeau et al. 1999). The main advantages of such systems are robustness, flexibility, adaptability, concurrency, and multiplicity. In the bio-inspired computation field, application of ants' foraging behaviors to telecommunication networks (Schoonderwoerd 1996; Di Caro and Dorigo 1997) and control of multiple robots (swarm robotics; Lipson 2005; Murata 2006) are a few examples of applications that use distributed controls to gain more flexibility and robustness in their systems. *Self-organizing computation* is a computational approach that brings out the strengths of the dynamic mechanisms of self-organizing systems: *structures appear at the global level of a system from interactions among its lower-level components*. In order to computationally implement the mechanisms, the system's constituent units (*subunits*) and the rules that define their interactions (*behaviors*) must be described. The system expects emergence of global-scale spatial structures from the locally defined interactions of its own components. Development of adaptive design systems may benefit from active implementations of self-organizing logic by gaining its characteristics, such as flexibility, adaptability, and tolerance for growth.

In architecture, following and synthesizing natural principles as a basis of building design date back to the work of Otto (1985) in Stuttgart, Germany, in the 1970s, and much of the effort has been spent on translating principles found in that decade into our contemporary professional practice that requires computational methods for the complete process of design. Most recently in this area, Hensel et al. (2010) have reported their theories and experiments regarding biomimetics and emergence in architecture.

There were several inspirational projects built by architects in the past, and obviously, the development of flexible and adaptable architecture has been a perennial theme among practitioners. During the 1960s in Japan, metabolists introduced megastructures that could constantly grow and adapt by plugging prefabricated pods onto the infrastructural core. Similar reconfigurable systems have been proposed by others including Archigram (Cook 1999). However, original visions of metabolic growth and adaptation were rarely realized physically, as the sizes and weights of the pods were practically very difficult to reconfigure. In the 1990s, construction automation by general construction companies in Japan shed light on the concept of self-reproduction in architecture: architecture that can produce architecture (Shiokawa et al. 2000). However, there was still a clear division between assembler and assemble relationships. Mechanical components that could produce buildings were far from actual livable architectural spaces. Thus they could only repeat, producing an identical or similar building at a time, and no future adaptation was available. These precedents indicate the difficulties of designing universal subunits that could tolerate technological, environmental, and circumstantial changes associated with structures and the differences in scale between artificial and natural systems (Figure 4.4).

Most artificial modular systems' building configurations are predetermined by designers and tend to require longer periods to actively organize themselves into a meaningful form or to require assistance from other assembler systems such as cranes. Reconfigurable swarm robotics systems devised by computer scientists are a more advanced concept that affords active behaviors to their subunits. In order to fulfill this gap between natural and artificial systems, it may be inferred that subunits of artificial systems need to gain more active dynamic behaviors while maintaining a fine granular size relative to global systems. Recent interests in nano/micro robotics all point in a biomimetic engineering direction, and development of such nano devices may be one way to assimilate emergence. The graph in Figure 4.5

(a)

(b)

FIGURE 4.4
(a) Nakagin capsule tower by Kurokawa, Tokyo, 1971 (Photo by Tomio Ohashi, courtesy of Kisho Kurokawa Architect & Associates.) (b) Habitat 67 by Moshe Safdie, Montreal, Canada, 1967 (Photo by Wladyslaw, Montral:Habitat67, 2008; image is from Wladyslaw, available at //commons.wikimedia.org/wiki/File:Montreal_-_QC_-_Habitat67.jpg)

shows the relative size of a subunit over the entire system on the *x*-axis and the lifetime or time for a growth period (in years) on the *y*-axis. Regardless of the sizes of entire systems, the natural systems tend to show smaller ratios for the relative size in the *x*-axis compared to those of many artificial systems, especially ones that were deliberately designed by professionals. Termite nests are composed of varied debris transported by termites, and their sizes are much finer than replaceable pods proposed by metabolists. When we get reduction in this relative size of a subunit, systems start to display more prominent characteristics of emergent behaviors.

However, recent advances in technologies may allow architects to envision more actively responsive structures in the near future. For example, new construction materials including ultra-high-strength concrete can span far greater lengths and expand architectural scope and possibilities. Moreover, recent advancements in sensor technologies have opened possibilities for buildings to have active adaptable mechanisms, and series of kinetic structures by Calatrava (2012) and Hoberman Associates (2012) foreshadow artistic integration of sensor technologies, robotics, and architecture. Application of robotics in architecture has already become a trend in academia, and Gramazio and Kohler at ETH Zurich (2012) and Design Robotics Group at Harvard (2012) have actively used industrial robotic arms for their design experiments since 2005 (Figure 4.3a).

In line with Alexander's prediction half a century ago (Alexander 1964) about the rapidly changing culture of his day, the author speculates that

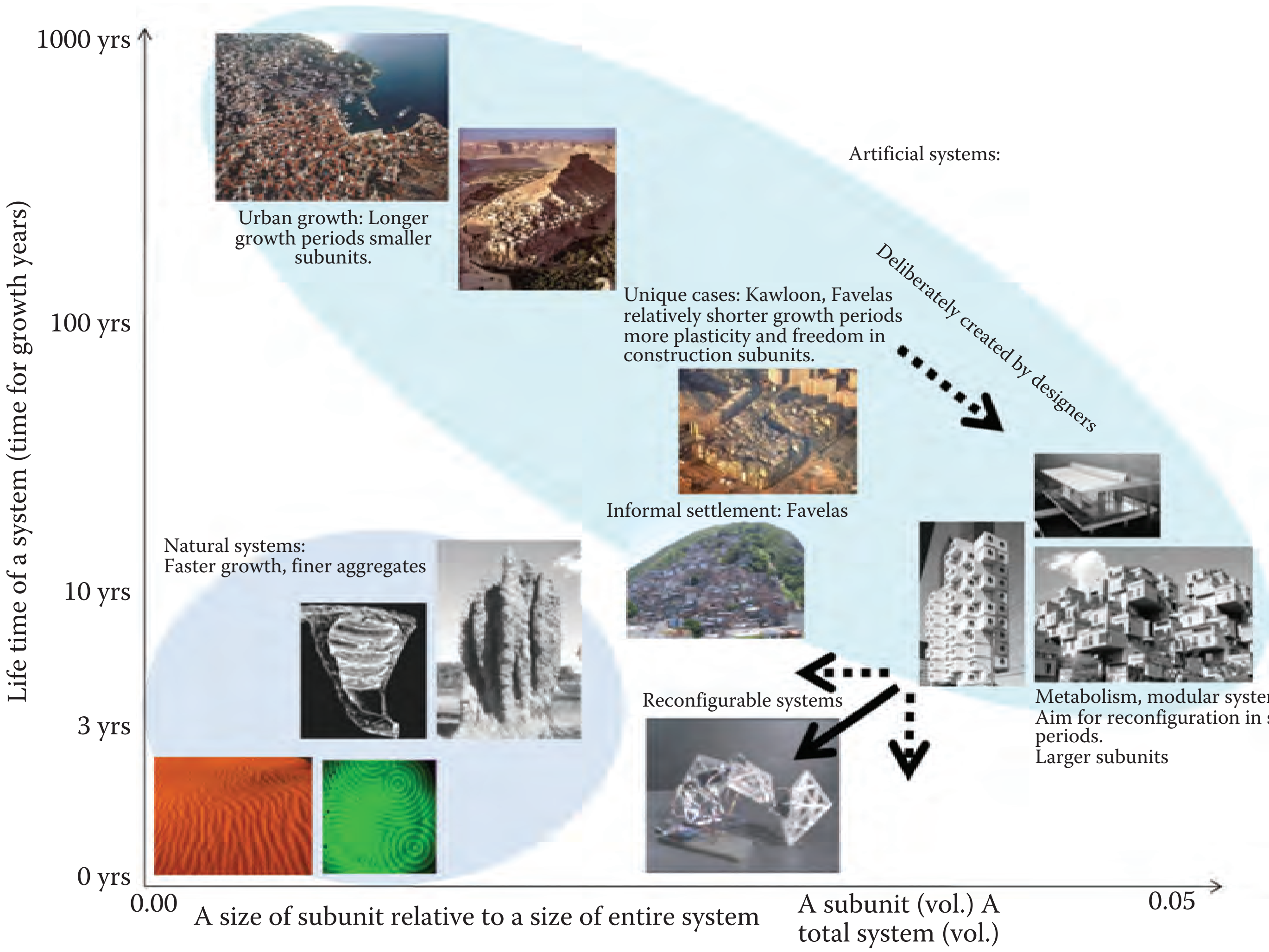


FIGURE 4.5
Various systems' lifetime and their subunit scales.

there will be more demands for buildings to be able to adapt to newly emerging needs for different qualities and quantities of architectural and urban-scale components. Within the limited allowable growth areas in dense urban settings, the desire to be able to accommodate new needs will increase demand for more adaptable buildings that can avoid future demolitions or drastic reconstructions. The hypothesis of this chapter is that the computational implementation of self-organizing principles is one potentially valuable strategy to improve the design of systems that are required to change over time.

4.5 Emergent Formations: Examples of Self-Organizing Computation

The following section presents relatively simple yet explicit examples of self-organizing computation in order to clarify the preceding more conceptual and abstract discussions about design and self-organization. The main focus is not on providing immediate solution methods to resolve any specific professional problems in architecture, but rather on investigating the emergent characteristics of these simple models' methods that can potentially evolve new design solutions over time, and on showing how tools employing the methods can be used beyond passive evaluation and analysis tools for given architectural instances.

Lane formation is a fascinating emergent phenomenon we can observe from simple agent-based pedestrian simulation (Figure 4.6). In crowds of oppositely walking pedestrians, the gradual formation of varying lanes of pedestrians moving in the same directions are observed. This is an empirically observed collective phenomenon and has been recorded in many real-life locations such as crowded pedestrian streets crossing in the city of Tokyo (Katoh et al. 1980). The emergence of this spatiotemporal pattern is a result of nonlinear interactions among pedestrians, and groups of pedestrians can find efficient walking formations solely from locally embedded individual behaviors without imposing any global geometry. Each pedestrian's embedded behavior is simply avoiding others, blocking his or her way in local neighborhood conditions, yet the global collective behavior that emerges from interactions of simple behaviors shows self-organized characteristics of a crowd.

Pedestrian agents as in Figure 4.7 are further developed (Narahara 2007), displaying more cognitive behaviors in reaction to spatial characteristics such as transparent surface, opaque surface, and furniture, in a system called the Space Re-Actor (Figure 4.7). Each agent is assigned a psychological profile with a different degree of sociability and reacts to proximity and visibility of others in the same space. Behaviors of agents are stochastically applied

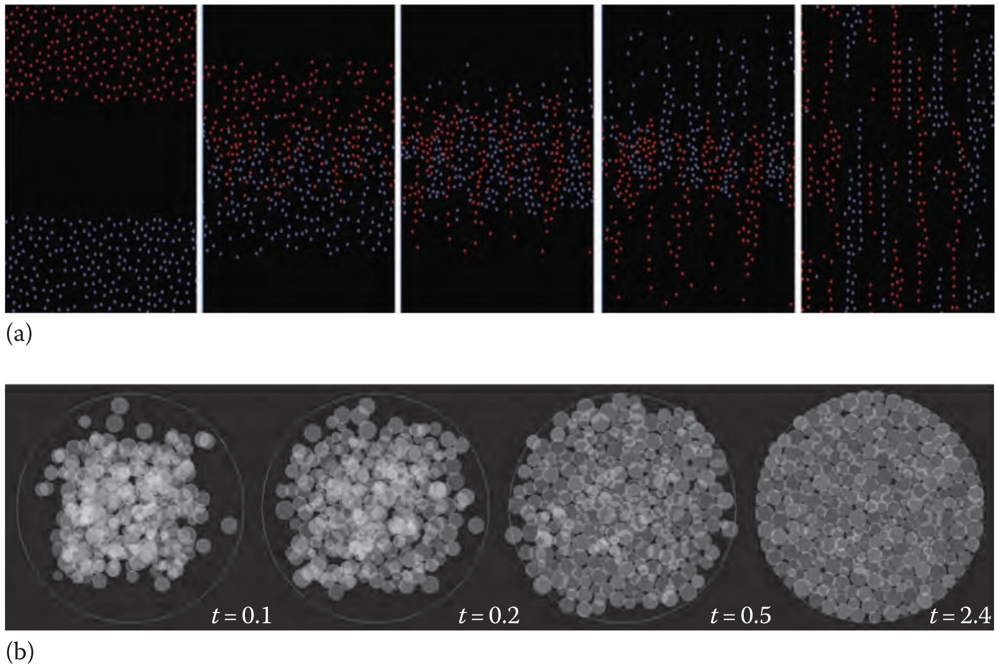


FIGURE 4.6
Pedestrian simulation by the author and emergence of lanes (a). Circle packing using a bubble mesh method (simulation by the author) (b).

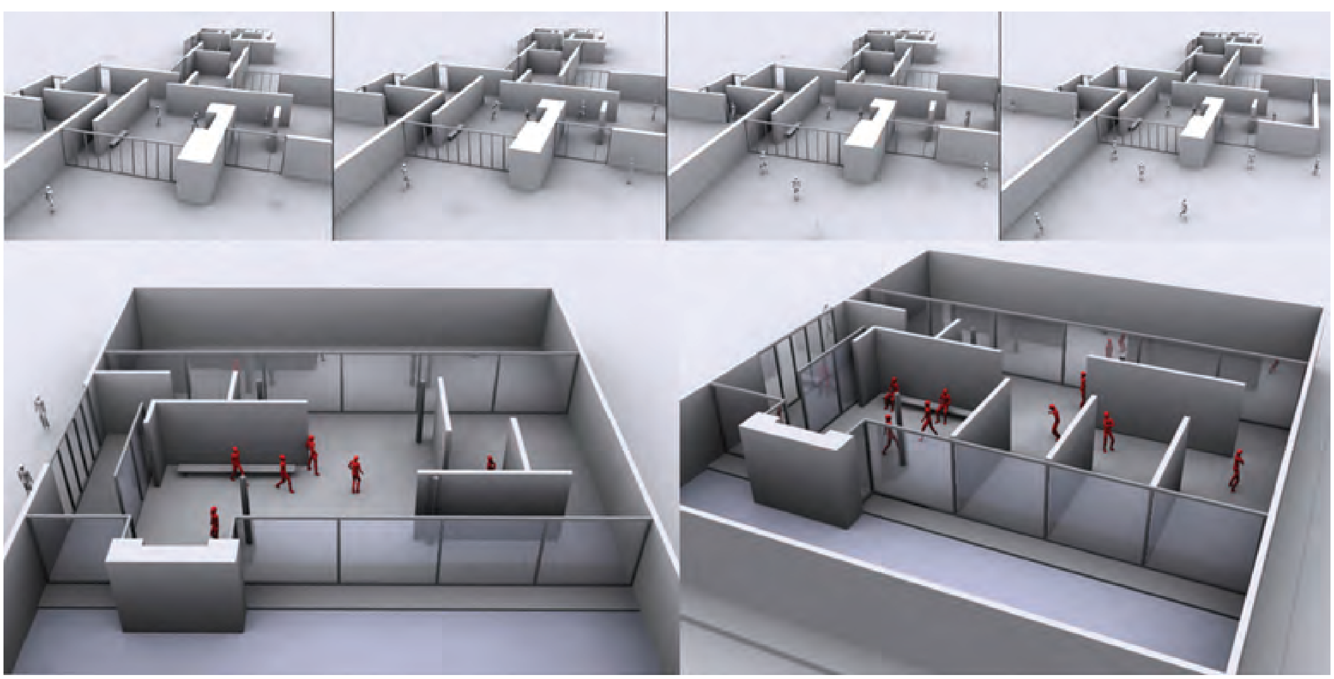


FIGURE 4.7
The Space Re-Actor: Simulation and visualization of human behavior. (From Narahara, T., The space re-actor: Walking a synthetic man through architectural space, MS thesis, Massachusetts Institute of Technology, Cambridge, MA, 2007.)

based on statistical information about human behaviors under certain circumstances, which implies that every run of the simulation can be slightly different, and normally, analyses based on stochastic simulations require multiple trials to verify their results. Though this model does not directly create new design, its real-time editing feature for spatial plans allows users to select spatial configurations with desired pedestrian flows and behaviors.

Another example of self-organizing computation using local interactions by agents is close packing of circles within a circle. Circle packing is one typical case where simulation using dynamics excels the performance of any analytical means, and it can be implemented by relatively straightforward code. Simple, locally implemented physical motions of bubbles—pushing and squeezing against each other—can eventually lead a group of bubbles to form a globally cohesive packing layout.

A physical implementation of previous examples in a more architectural context is the following robotic device with locally embedded sensors and microcontrollers (Figure 4.8). Unlike pluggable pods by metabolists being inert objects, subunits here are scaled architectural components that can reconfigure themselves into globally functional configurations based on feedback from locally distributed intelligence. Bottom-up control strategies allow the device to optimize its orientation with respect to a light source, independent of how and where the unit is placed using a multidimensional optimization algorithm. The project aims to demonstrate a design system that can respond to a dynamically changing environment over time without imposing a static blueprint of the structure in a top-down manner from the outset of design processes (Narahara 2010b).

In architecture, difficulties of developing computational design systems include the multiple objectives typical of an architectural problem and the considerable size of the search space that contains possible formal and programmatic architectural solutions. In the case of adaptive growth models, objectives are also ever-changing dynamic properties of the models, and the computational design systems of such models are expected to possess certain solution-seeking behaviors that can maneuver through the vast dynamic “solution-scape.”

The earlier examples have yet to yield reliable tools for practice, though they display generative characteristics. It is worth noting that all the preceding methods rely on a heuristic. A heuristic method is a solving of a problem by iterative processes of trial and error and is intended to find optimal solutions rather than to find a single deterministic solution. As can be seen in the last project, calculations of dynamic reconfigurations can be fairly extensive if you search all one by one (i.e., brute-force search). This fact implies that the deterministic analytical means are less adequate where we need concurrent solutions for dynamically changing conditions, and we may need to rely on heuristic search as the complexity of the project increases.

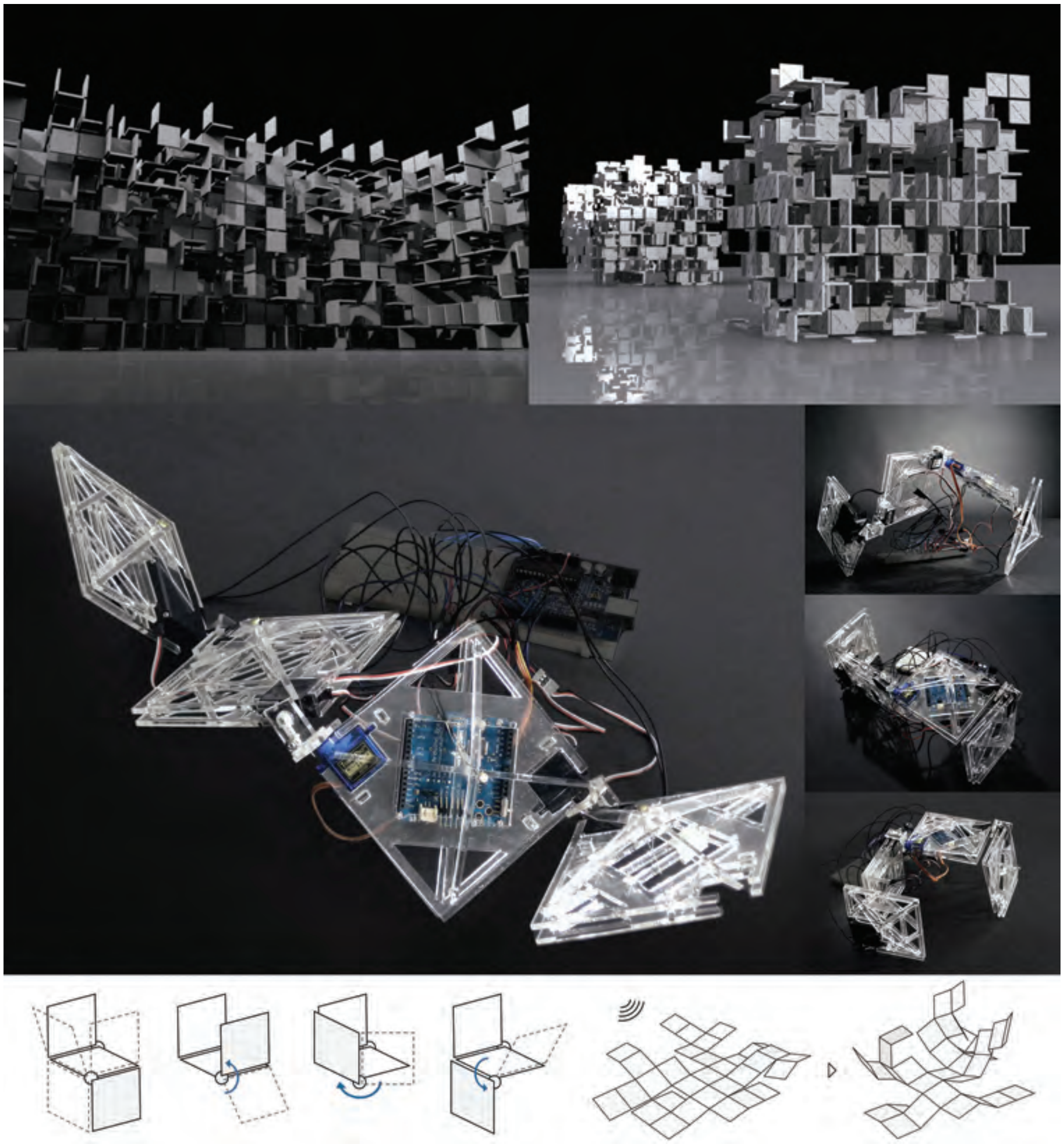


FIGURE 4.8
Robot tries to find better configurations to receive more light exposure on its four panels using Nelder–Mead multidimensional optimization. (From Narahara, T., *Int. J. Archit. Comput*., 8, 1, 29, 2010.)

4.6 Program for Simulating the Origins of Urban Form

In this section, the chapter further explores a generative aspect of agent-based models on an urban scale and proposes a computational method that simulates growth processes of settlement patterns.

There is an earlier work that introduced methods to produce spatial patterns using the behavior of spatial agents. Helbing et al. (2001) have done computational simulations using their "active walker model" and simulated trail formations observed within a human trail system. Their agents use a marking behavior that leaves modifications of the ground that make it more comfortable to walk on. This implementation is similar to formation of trail systems by certain ant species using chemotaxis, and an algorithmic implementation of chemotaxis was also introduced by Dorigo (1992). In architectural research, Schaur (1991) investigated empirical precedents of a human trail system, and these precedents indicate that many computational simulation results are able to obtain characteristics of pattern formations in a human trail system.

The system proposed in the following text is a computer program written in C# and is applicable to any terrain found in Google Earth by exporting geographical data. The system's three primary components are terrain, agents, and buildings. The system assumes that the development starts from unoccupied empty terrain, and behaviors of wandering settlers are simulated by the computational agents. The terrain surface is subdivided into a grid of triangulated patches that can store dynamic local information about traffic intensity of agents (settlers) and local acuteness of a terrain (slope). The buildings and streets (trails) will be gradually generated by the agents as a part of their behaviors. The primary behavioral characteristics of agents include physical mobility such as hill-climbing ability, their attraction toward environmental conditions, their selection of paths based on local traffic density, and presence or absence of global destinations.

These primary factors that govern the heading direction vectors of agents are named as attraction to gentle slope, attraction to traffic intensity, and attraction to destinations. The earlier three attractions can coexist simultaneously in many real-life scenarios. A normalized sum of three vectors weighted by three factors, slope-factor (*s*-factor), traffic-intensity-factor (*t*-factor), and destination-factor (*d*-factor), produces the heading vector of an agent. By manipulating the proportions of these three weights, the behavior of agents can be directed. These values are also dynamically changeable parameters based on changing environmental potentials.

$$V\text{agent} = s * V\text{slope} + t * V\text{traffic} + d * V\text{dest}$$

$$(s + t + d = 1.0;\ 0 \le s, t, d \le 1.0)$$

The system advances every finite step of time. The value for attraction—intensity of traffic—is directly related to the visibility of the trails. Traces of trails by agents are perceptible by others in the region at adjacent patches and attract them. Trails that are frequently used by agents become attractors for agents in their neighborhoods. Consequently, the frequency of these trails' usage is amplified. In contrast less used trails will eventually fade away, and this is implemented by applying a reduction rate to the traffic intensity value of all subdivision surface areas every step of a simulation. The positive and negative feedback work as a pair to produce organic gradual transformations of street patterns such as branching, bundling, and emergence of hierarchy among them (Figure 4.9).

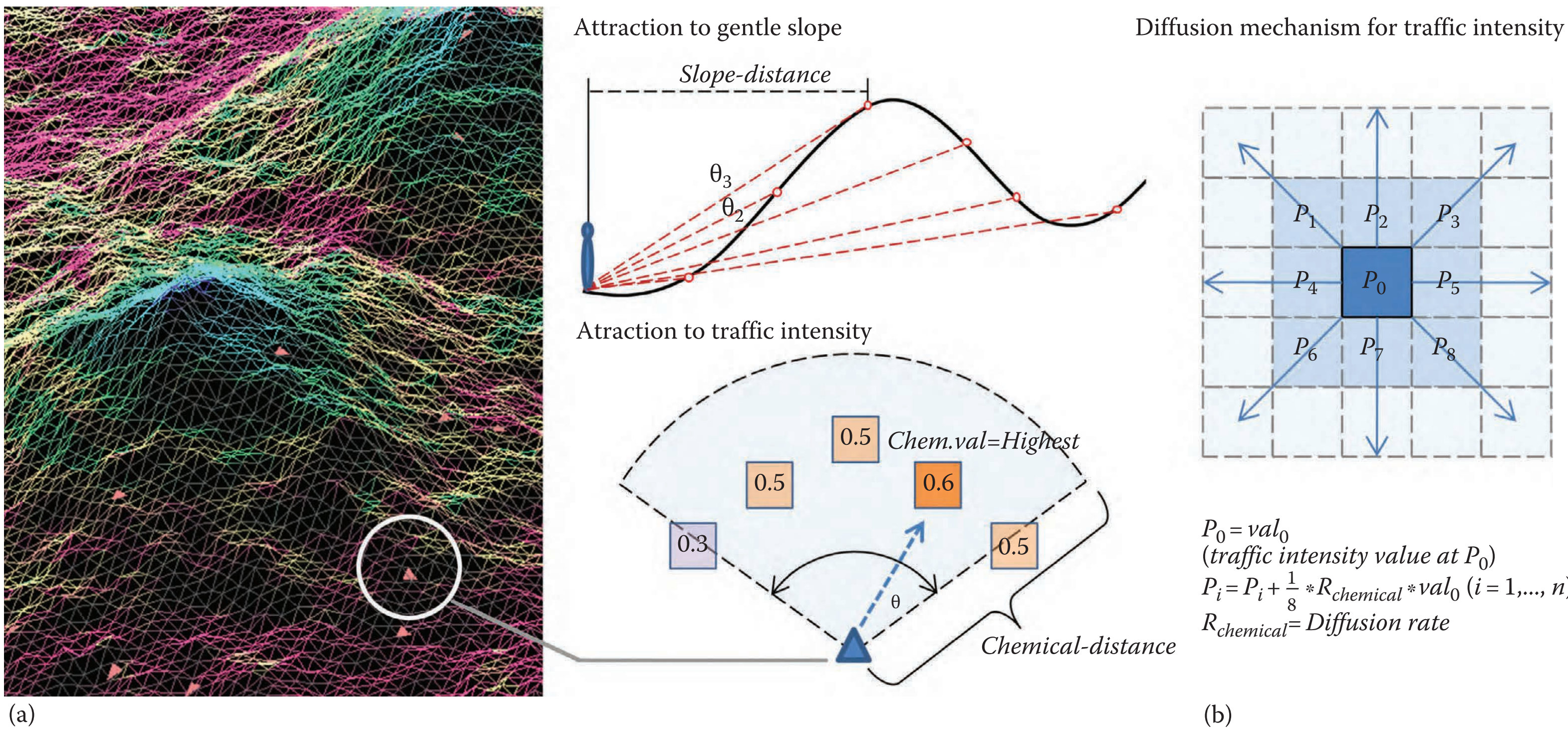


FIGURE 4.9
Agents on a wire-framed terrain with color-coded traffic frequencies (a). Agent's cone of vision and chemical value checking mechanism (b).

As a preliminary scenario with fixed behaviors of agents throughout the run of simulations, the results are categorized in roughly three emergent patterns, *direct paths, minimal ways*, and *detours*, and other in-between patterns. The high value for slope-factor induces agents to find a comfortable walking path. By avoiding climbing or descending a steep hill, agents produce *detours*. In this simulation, the given terrain has two steep hilltops at the south and north, and detours around these hilltops are recognized from the results. The high value for destination-factor induces agents to find the shortest path to form a *direct path* system. The high value for traffic-intensity-factor stimulates agents to minimize overall length of circulation. When agents have more frequent trips between destinations, shorter overall length of the system is beneficial because of its lower construction costs of roads: *minimal ways*. These three factors can be applied in various different proportions to find compromise solutions among three different motivations (Figure 4.10).

One of the biggest merits of the system is that purely microscopic behaviors can produce a global configuration without requiring any macroscopic information to be input into the system. When numbers of cities are small, resulting configurations are rather predictable. However, deriving a street configuration from a large number of cities at arbitrary locations on irregular

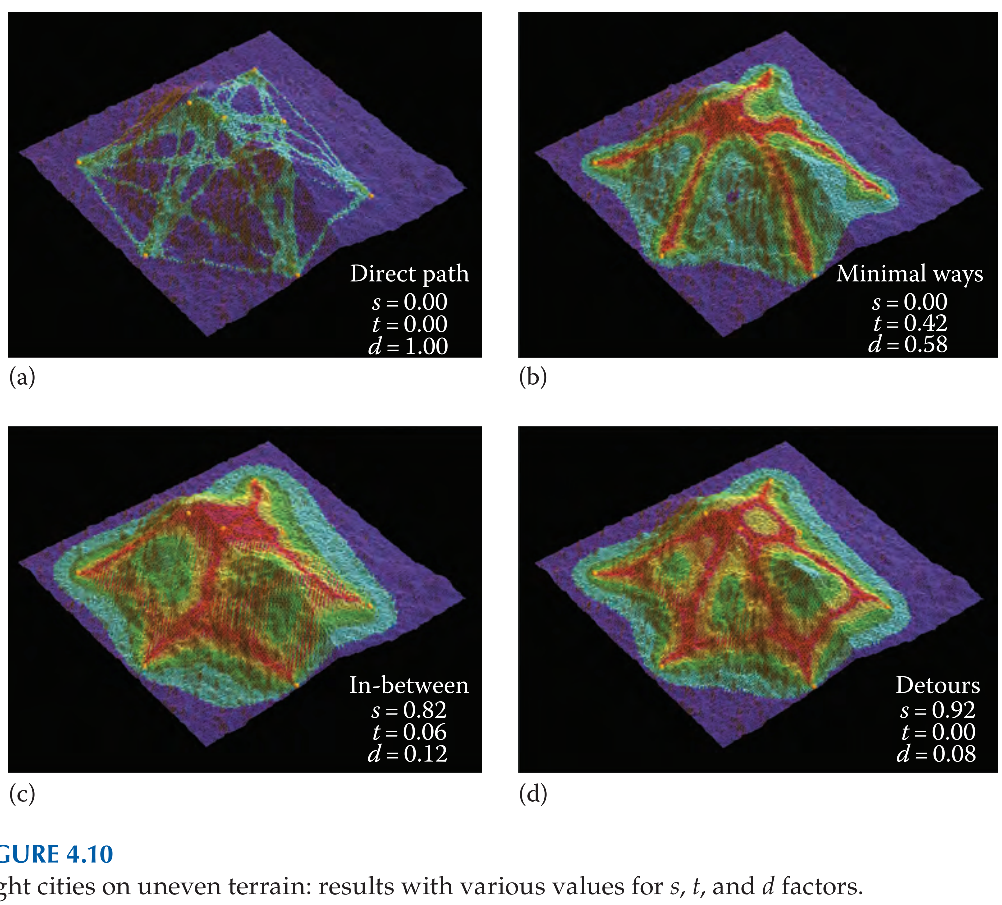


FIGURE 4.10
Eight cities on uneven terrain: results with various values for *s*, *t*, and *d* factors.

terrain geometry is a challenge. Multiagent simulation is not necessarily the fastest method of derivation, but it is a reasonably robust method for deriving street configuration as it requires only microscopic behaviors as input information.

4.7 Growth Simulation

The system can change behaviors of subunits over time based on stimuli from the changing environment. In the following section, shifts in behaviors of agents stimulated by environmental changes are considered. Once the activity level of an entire environment reaches a certain maturity, agents start to rely on information that already exists in environments. Agents start to follow higher traffic frequency areas instead of trails that no one has been taking (i.e., *t*-value increases). The earlier behaviors were implemented by setting a threshold value to shift behaviors of agents. As the number of terrain patches that possess traffic intensity value exceeds the threshold, agents start to check the traffic frequency around them to make decisions about their heading directions (Figure 4.11).

After the emergence of street networks, some of the intersections of several arteries become population concentration areas, and these areas have the potential to grow into cities. The algorithm finds peak areas of traffic intensity value above a certain threshold value and finds places where these peaks are forming clusters. Traffic-intensive patches that are within a certain distance from each other are read as one island. If these islands are larger than a certain minimum size, they are considered as city areas. After sufficient careful trials, numbers that produce results that represent a natural scale of development relative to the scale of terrain are empirically adopted. Once cities are registered by the system, all agents will travel around these cities.

Emergence of buildings is dependent on environmental potentials that are mainly frequencies of traffic and topographical conditions of a terrain. The frequencies of traffic at each patch of the terrain are considered as an indication of population density. When this value is higher than a certain threshold value, the site becomes a potential location for buildings under a certain probability. The sizes and heights of the buildings are defined by a negative exponential function of traffic intensity value to reflect upper-bound heights found in existing settlements.

Slopes of the site are another criterion for building sites. If the maximum slope of the site is above a certain degree, it is quite likely that no settler is willing to build any structures at such a steep slope. This maximum value can be dependent on regional tectonic cultures, available materials and technologies, and climatic conditions. Buildings are modeled to have an ability to

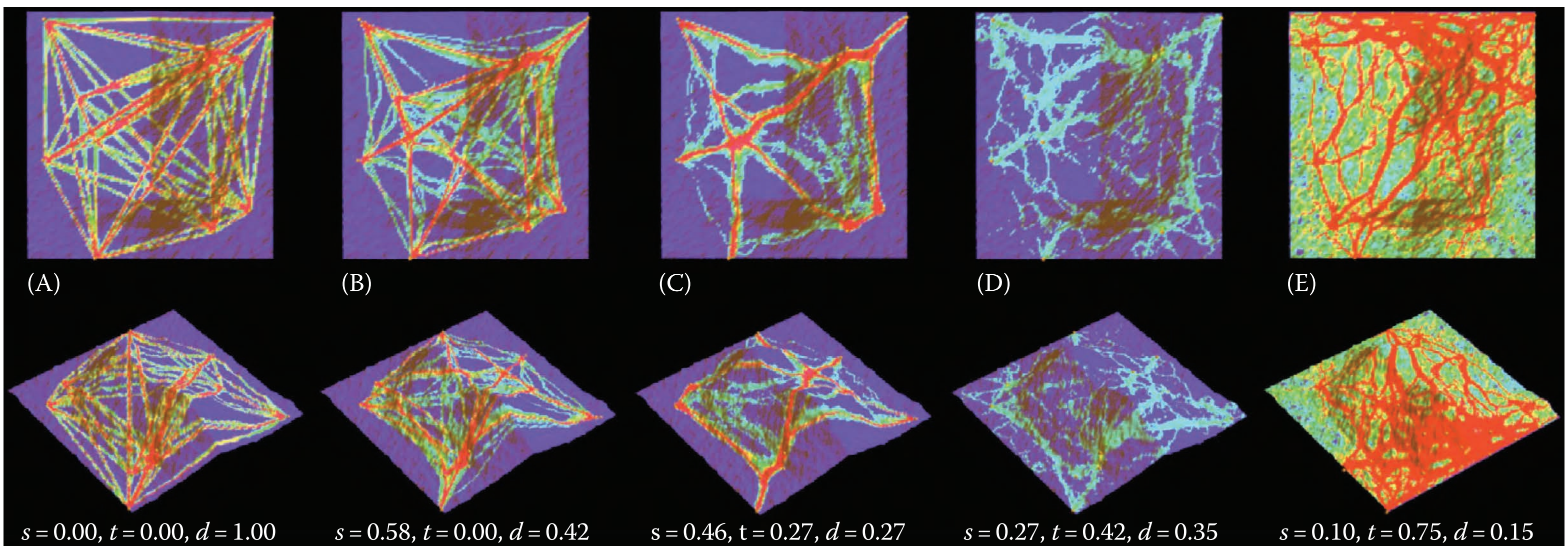


FIGURE 4.11
Results of emerging patterns on a terrain with eight predefined stationary destination points and two hills with various values for *s*, *t*, and *d* factors.

align themselves to the heading directions of agents and shift locations of buildings to avoid direct collisions. Buildings also calculate the average rotation angle of others within a certain distance away and attempt to align them to the average angle. This is a self-organizing behavior often seen among a flock of birds. Synchronization among the rotation angles of buildings is eventually expected to produce natural arrays of buildings.

Once the overall environmental potential grows above the aforementioned minima, cities will have emerged. At this stage, settlers are no longer randomly wandering migrants seeking temporary shelter. Instead, they act as heterogeneous self-driven agents based on clear objectives of their own and travel through these newly emerging cities. Spatiotemporal developments of the site are organized by the behaviors of settlers. However, the environmental changes induced by the settlers also simultaneously influence their behaviors.

4.8 Results of Simulation of an Existing Site

Figure 4.12 shows results obtained from the proposed system tested on topographic data from an existing site, San Miniato in the region of Tuscany in Italy. The site was employed due to its unique correlations between its landform and urban settlements. The original environment of the system is completely vacant: unoccupied. The early phase of the experiments has captured pat-terns seen in some existing settlements (Figure 4.12, top). The formal results of the simulation are induced by a repertoire of a few behaviors imple-mented in agents, and some of the behaviors spontaneously emerged from the growth of a system itself. This inductive characteristic of the system implies that some tendencies found in human design activities are, to some extent, captured or approximated by a few repertoires of major behaviors by agents.

Figure 4.12 (bottom) shows the hypothetical results where incoming population and traffic intensities of the site area continue to increase. The current site, the San Miniato area, has a moderately settled, relatively rural condition, yet some of the results from later phases show developments close to the typical density of metropolitan-class cities. Various parameters that govern the local behaviors of agents changed the later patterns of the developments of the cities. Schemes from five different parameter settings display various extreme urbanistic future scenarios for San Miniato in a speculative domain. One can decide which schemes to select by finding which schemes induce global behavioral patterns to emerge that one prefers agents to enact.

Of course, configurations based on decisions by groups of humans are often results of complex negotiations and superimpositions of multiple results over time, and complete descriptions of these processes may not

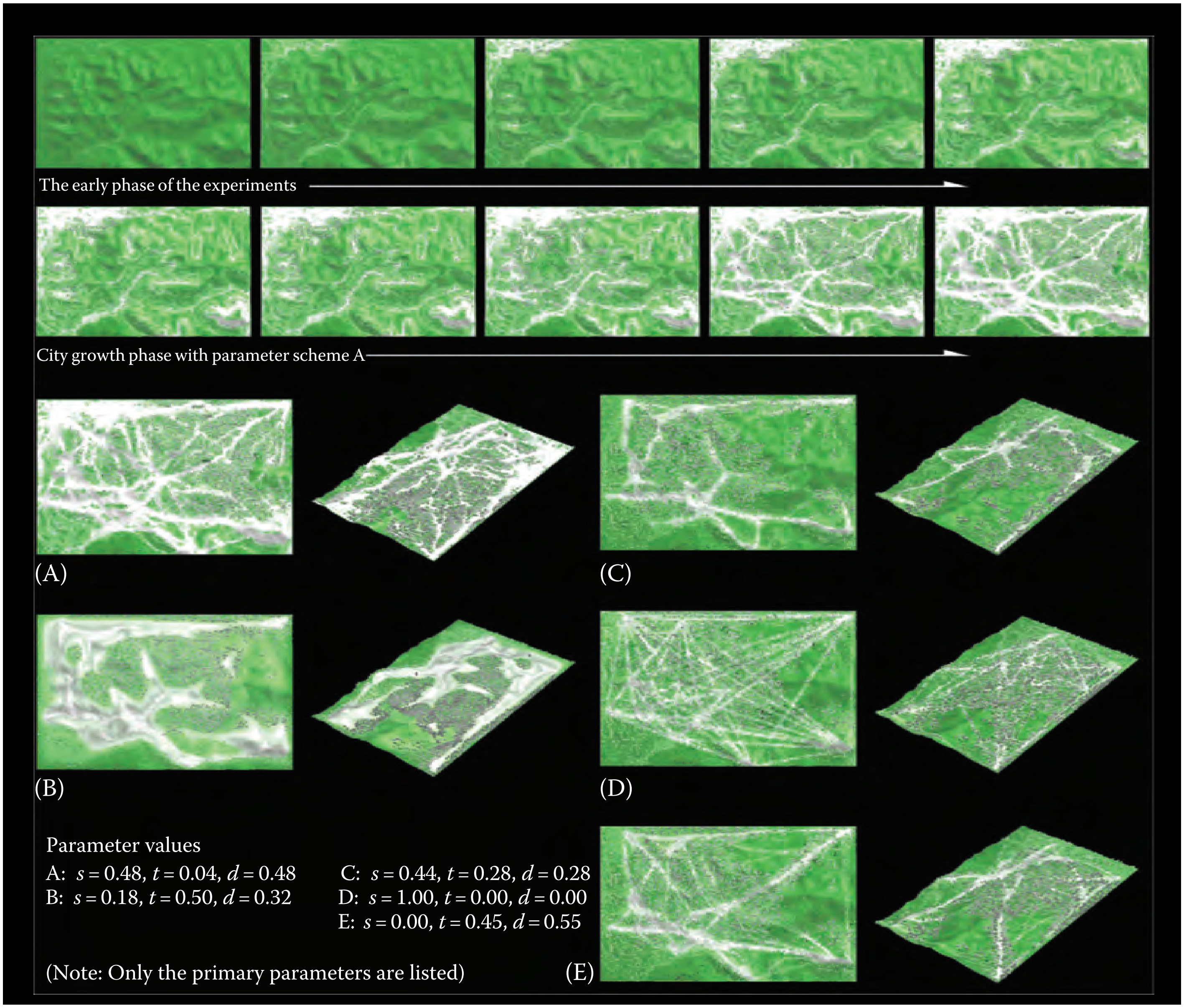


FIGURE 4.12
City growth phase with scheme A: gradual growth of paths and buildings (top). City growth phase with five different parameter settings (bottom).

easily be reduced to a simple set of parameters. The system has yet to be equipped with social, political, and economic considerations. However, in principle, dynamics associated with such considerations can be computationally described by implementing more complex behaviors and contextual constraints, in addition to the three primary behaviors, in order to move toward a more practical application.

In this system, the behaviors of agents are updated accordingly as new paths and buildings are generated. This coevolutionary process between agents and environments is known to exist in many self-organizing systems. One obvious advantage is its ability to *represent and generate growth processes over time*. This means that the approach can design a system in transition, and it is a necessary feature for simulating decentralized dynamics of settlements.

Environmental growth of the system and behavioral changes of agents lead the entire system to gradually adapt itself to emerging states of the system, such as "building allocations that can maintain functional traffic patterns" or "developing shortest path patterns for agents to travel around cities." These objectives are ex post interpretations of the results and are not provided directly as *globally* defined initial requirements or goals for the system. During the course of the system's run, these changes toward satisfaction of certain objectives emerge as a result of the system's ability to adjust its response to stimuli according to the state of the environment, solely from *locally* defined individual behaviors of agents. This characteristic of the system leads us to speculate regarding the possibility of creating a growth system that can discover new unknown objectives (in a sense, the emergence of new architectural programs) by further developing this computational approach, and such a system would be potentially applicable to planning and strategic development of architectural projects in earlier stages.

Finally, one of the unique characteristics of self-organizing computation is its *nonreliance on any external knowledge*. Unlike some existing city generation software, the system does not impose specific design templates such as grid, radial, or branching patterns. Reliance on a preexisting template may preclude the possibility of discovering what original inputs naturally turn into. With a few primary inputs related to the site's geographical information as initial conditions, the system can spontaneously produce all design components using self-organizing computation.

In this manner, the inherent characteristics of the resulting configurations are traceable back to several parameter values that govern the behavior of the system, and certain sets of parameters that lead to characteristics similar to existing urban phenomena can be studied. The goal of the experiment is to derive forms from behaviors instead of supplying a formal knowledge of design patterns at the outset.

4.9 Conclusion

Computational design application tools in architecture are on the brink of transition from being mere analytical tools to becoming more creative tools that can induce emergence of new solutions, or at least serve as "coevolvers" of design solutions for humans. This transition in the role of computational tools will have a big impact on our design communities and will pose a question about what the actual roles of human design experts in the field are. Further, the rise of collective design interface platforms may change our current value systems in architecture and design. This chapter has investigated computational strategies that could advance this transition and proposed generative approaches through conceptual experiments. An obvious next step is to find real-life scenarios to which these approaches are applicable and develop feasible systems that go beyond conceptual desktop experiments to become practical off-the-shelf solutions.

Acknowledgments

Firstly, I would like to thank Dr. Justyna Zander and Dr. Pieter J. Mosterman for this great opportunity. I would also like to sincerely thank my former academic advisers, Professor Martin Bechthod and Professor Kostas Terzidis at Harvard University, and Professor Takehiko Nagakura at the Massachusetts Institute of Technology, for their insightful guidance and constant support. Finally, I would like to thank my current employers, Dean Urs Gauchat and Professor Glenn Goldman at New Jersey Institute of Technology for their generous academic support.

References

Alexander, C. (1964). *Notes on the Synthesis of Form,* Harvard University Press, Cambridge, MA.

Autodesk Ecotect Analysis. (2012) http://usa.autodesk.com/ecotect-analysis/ (accessed: August 8, 2012).

Batty, M. (2006). *Cities and Complexity: Understanding Cities with Cellular Automata, Agent based models, and Fractals*, MIT Press, Cambridge, MA.

Batty, M. and Longley, P. (1994). *Fractal Cities: A Geometry of Form and Function*, Academic Press, San Diego, CA.

Bonabeau, E. et al. (1999). *Swarm Intelligence: From Natural to Artificial Intelligence*. Oxford University Press, New York.
Camazine, D., Franks, S., Theraulaz, B. (2002). *Self-Organization in Biological Systems*, Princeton University Press, Princeton, NJ.
Centre for Advanced Spatial Analysis (CASA), University College London. (2012). http://www.bartlett. ucl.ac.uk/casa (accessed: August 8, 2012).
Cook, P. (1999). *Archigram*, Princeton Architectural Press, New York.
Design Robotics Group. Material processes and systems research, Graduate School of Design, Harvard University, (2012). http://research.gsd.harvard.edu/drg/ (accessed: August 8, 2012).
Di Caro, G. and Dorigo, M. (1997). AntNet: A mobile agents approach to adaptive routing, Technical Report IRIDIA/97-12. Universite Libre de Bruxelles, Belgium.
Digital Project (2012). http://www.gehrytechnologies.com/node/4 (accessed: August 8, 2012).
Dorigo, M. (1992). Optimization, learning and natural algorithms, PhD thesis, Politecnico di Milan, Italy.
Eastman, C. et al. (2008). *BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers*, Wiley, Hoboken, NJ.
Eisenman, P. (1992). Visions unfolding: Architecture in the age of electronics media, *Domus,* 734, 17–24.
Fanelli, G. (1990). *Toscana*, Francesco Trivisonno Firenze, Cantini.
Frazer, J. (1995). *An Evolutionary Architecture*, Architectural Association, London, U.K.
Gehry Technologies. (2012). http://www.gehrytechnologies.com/ (accessed: August 8, 2012).
Gero J. S. (2012). http://mason.gmu.edu/~jgero/ (accessed: August 8, 2012).
Gramazio & Kohler, Architecture and digital fabrication, ETH Zurich Department Architecture. (2012) http://www.dfab.arch.ethz.ch/ (accessed: August 8, 2012).
Haken, H. (1983). *Synergetics,* Springer-Verlag, Berlin, Germnay.
Helbing, D., Farkas, I. J., and Bolay, K. (2001). Environment and Planning B, 2001, Self-organizing pedestrian movement. *Environment and Planning B: Planning and Design*, 28, 361–383.
Hensel, M., Menges, A., and Weinstock, M (2010). *Emergent Technology and Design—Towards a Biological Paradigm for Architecture*, Routledge, London, U.K.
Hillier B. (1999). *Space Is the Machine: A Configurational Theory of Architecture*, Cambridge University Press, Cambridge, U.K.
Hoberman Associates, Inc. (2012). http://www.hoberman.com/home.html (accessed: August 8, 2012).
Katoh, I. et al. (1980). Characteristics of lane formation, *Nihon Kenchiku-gakkai Ronbun houkoku-shu*, 289, 121.
Lipson, H. et al. (2005). Self-reproducing machines, *Nature*, 435, 7038.
Lambot, I. and Girard, G. (1999). *City of Darkness: Life in Kowloon City*, Watermark Publishing, England, U.K.
Lynn, G. (1999). *Animate Form*, Princeton Architecture Press, New York.
Mitchell, W. (1977). *Computer-Aided Architectural Design*, Petrocelli/Charter, New York.
Mitchell, W. (1990). *The Logic of Architecture*, MIT Press, London, U.K.
Murata, S. (2006). *Modular Structure Assembly Using Blackboard Path Planning System*. *ISARC2006,* Tokyo Institute of Technology, Interdisciplinary Graduate School of Science and Engineering, Tokyo, Japan.

MVRDV. (1999). *FARMAX*, 010 Publishers, Rotterdam, the Netherlands.
MVRDV. (2005). *Spacefighter: The Evolutionary City Game,* Actar, Barcelona, Spain.
Narahara, T. (2007). The space re-actor: Walking a synthetic man through architectural space, MS thesis, Massachusetts Institute of Technology, Cambridge, MA.
Narahara T. (2010a). Design for constant change: Adaptable growth model for architecture, *International Journal of Architectural Computing (IJAC)*, 8, 1, 29–40.
Narahara, T. (2010b). Self-organizing computation: A framework for generative approaches to architectural design, Doctor of Design dissertation, Harvard University Graduate School of Design, Cambridge, MA.
NBIMSP (National Building Information Model Standard Project Committee). (2012). http://www.buildingsmartalliance.org/index.php/nbims/faq/ (accessed: March 2, 2012).
Negroponte, N. (1970). *The Architecture Machine: Towards a More Human Environment*, MIT Press, Cambridge, MA.
Nicolis, G. and Prigogine, I. (1977). *Self-Organization in Non-Equilibrium Systems*. Wiley & Sun, New York.
Novak, M. (1988). Computational compositions, *Computing in Design Education, ACADIA '88 Workshop Proceedings,* ed. P. J. Bancroft, University of Michigan, Ann Arbor, MI, pp. 5–30.
Otto, F. et al. (1985). Natürliche Konstruktionen: Formen und Strukturen in Natur und Technik und Prozesse ihre Entstehung (with a contribution by F. Otto) Deutsche Verlags-Anstalt, Stuttgart, Germany.
Santiago Calatrava LLC. (2012) http://www.calatrava.com/ (accessed: August 8, 2012).
Schaur, E. (1991). Non-planned settlements: Characteristics features—Path system, surface subdivision, ed. F. Otto, Vol. 39 of *Institute for Lightweight Structures* (IL), University of Stuttgart, Germany.
Schoonderwoerd, R., Holland, O., Bruten, J., and Rothkrants. L. (1996). Ant-based load balancing in telecommunications networks, *Adaptive Behavior,* 5, 169–207.
Shelden, D. R. (2002). Digital surface representation and the constructability of Gehry's architecture, PhD thesis, MIT, Cambridge, U.K.
Shiokawa, T. et al. (2000). Automated construction system for high-rise reinforced concrete buildings, *Automation in Construction*, 9, 229–250.
Stiny, G. and Gips, J. (1972). Shape grammars and the generative specification of painting and sculpture, in *Information Processing 71*, pp. 1460–1465. North-Holland Publishing Company, Amsterdam, the Netherlands.
Terzidis, K. (2006). *Algorithmic Architecture*, Architectural Press, Burlington, MA.
Weinstock, M. (2010). *The Architecture of Emergence—The Evolution of Form in Nature and Civilization*, Wiley Academy, London, U.K.
Yatsuka, H. and Yoshimatsu, H. (1997). *Metabolism: senkyūhyaku rokujū-nendai Nihon no kenchiku avangyarudo,* Inakkusu Shuppan, Tokyo.

SIGGRAPH2015 Xroads of Discovery

Narahara, T., Abbruzzese, K., and Foulds, R., "Haptic Collaboration: Biomedical Engineering Meets Digital Design", SIGGRAPH 2015 Talks. https://doi.org/10.1145/2785585.2792520 [Video].

Haptic Collaboration: Biomedical Engineering Meets Digital Design

Taro Narahara[†] Kevin M. Abbruzzese[*] Richard A. Foulds[*]
New Jersey Institute of Technology[**]

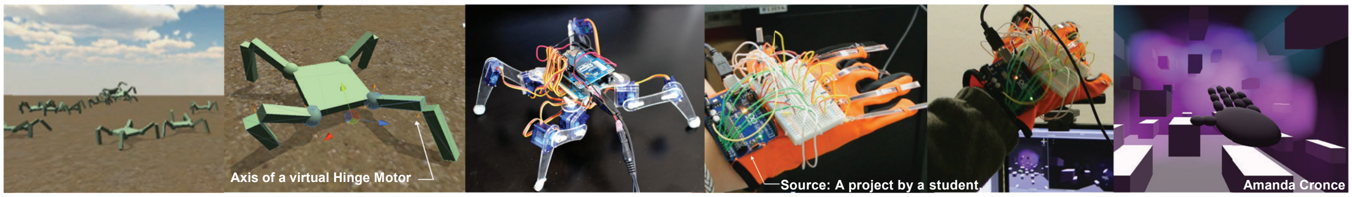


Figure 1: *Series of digital design projects interfacing physical prototypes with virtual game environments. A quadruped robot (right). A haptic glove (left).*

1. Introduction

This talk presents results of ongoing research and educational collaboration between the School of Art + Design (SoA+D) and the Department of Biomedical Engineering (BME) at New Jersey Institute of Technology. This collaboration began when researchers from BME became aware of a series of projects by digital design students from SoA+D producing virtual games that interface with fabricated physical design prototypes with microcontrollers through the use of the Unity 3D game engine as an application hub to connect the virtual and real worlds. The BME researchers had developed a novel admittance-controlled haptic robotic exoskeleton for assisting the upper extremity motions of people with stroke and cerebral palsy and were seeking to integrate it with an engaging and challenging virtual environment that can retain a user's interest. The result is a user-controlled haptic manipulator that allows individuals with neurological impairment to be therapeutically assisted by the exoskeleton (BME) while haptically interacting with virtual objects in a 3-D animated environment (SoA+D). The talk also introduces a new cross-disciplinary educational approach employing expertise of both academic units.

2. Our Approach

The SoA+D has developed a course for designers to create real-time interaction between physical prototypes and a digital application environment for games and simulation using readily available commodity hardware, including Arduino microcontrollers and Kinect sensors, and Unity 3D game engine software, with its computational physics. The course resulted in projects, such as 1) a haptic glove that allows a user to interact with virtual objects while returning a haptic sensation through vibrations of motors at the user's fingertips, and 2) a quadruped robot that can learn its own gait cycle virtually using the physics engine while physically testing motions. Unity and Arduino can be interfaced using the Transmission Control Protocol (TCP), which allows for implementation of GPU-accelerated computational physics, which is sufficiently fast to simulate user interaction with game objects and controls the physical prototype to provide a sense of virtual touch. While Unity animation operates at the computer screen refresh rate (~100 fps), its physics can operate at a higher frequency required for proper haptic perception.

[†] School of Art + Design, [*] Department of Biomedical Engineering
[**] e-mail: {narahara, kma26, foulds}@njit.edu

SIGGRAPH 2015 Studio, August 09 – 13, 2015, Los Angeles, CA.
ACM 978-1-4503-3637-6/15/08.
http://dx.doi.org/10.1145/2775280.2792520

3. Unity 3D: A Collaboration Hub

The versatility, usability, and adaptability of Unity have been proven as we have developed interfaces between Unity and MATLAB, which is the current primary means for BME researchers to control their haptic robots using TCP. Unity also serves as a collaboration hub between the BME and SoA+D due to its speed and accuracy for computational physics and compatibilities with various CAD model formats in common use among digital designers. This allows designers to provide original game logic with animated high-quality models using its real-time rendering capabilities without compromising accuracy and speed that are required by engineers. Unity's graphics capability is also compatible with immersive commercial stereo vision glasses and projection systems that can enhance the therapeutic effects.

4. Serious Games: Haptic Collaboration

Our approach couples the BME-developed exoskeleton with the SoA+D capacity to employ the graphics and the physics engines of game development systems to provide therapeutic effects of gaming using stereoscopic glasses and motion tracking systems in addition to the sense of touch to be provided by the exoskeleton. The BME's prototype supports 6 degrees of freedom plus a finger/thumb grip, which matches the flexibility of the human arm and surpasses any other admittance-controlled system of this type including the MIT Manus (only 3 haptic-degree-of-freedom with no haptic grasp). Admittance control allows the user to input a force and translates that force into motion, and the user will feel resistance in the robot when the virtual arm makes contact with the objects. Digital designers have started producing visually enhanced serious games for the system, and now the research collaboration has developed into a new cross-disciplinary educational platform. The talk will present and demonstrate the most recent results.

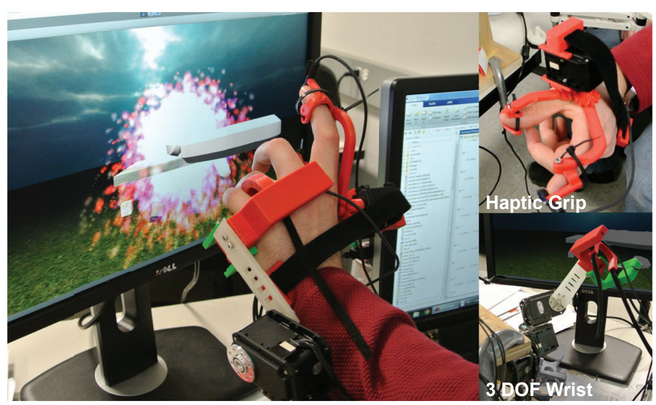
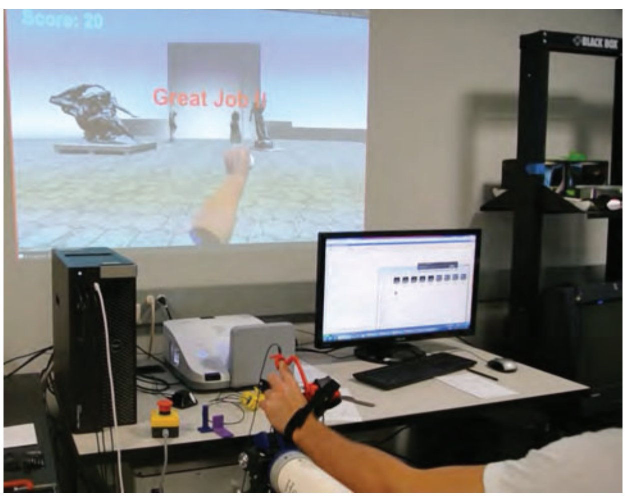


Figure 2: *An exoskeleton with admittance-controlled 3-degrees-of-freedom wrist and a finger/thumb grip haptically interacting with virtual objects in a 3-D animated environment.*



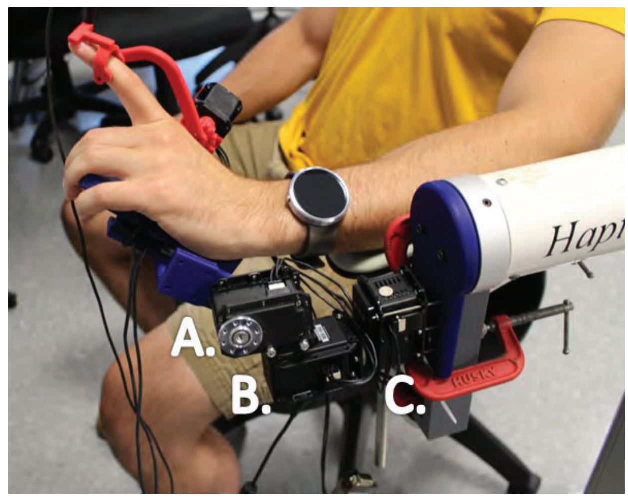


Figure. Left: Wrist exoskeleton with gripper interfaced with VE. Right: VE simulation in Unity of pinch task

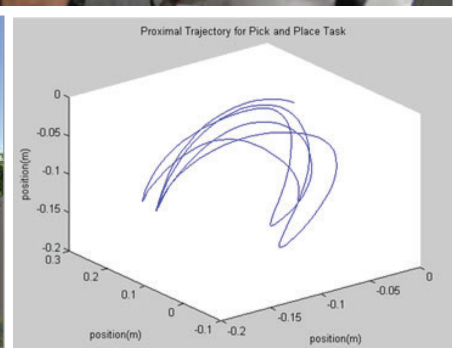


Figure. Proximal arm trajectory evaluated with the Haptic Master

Figure. Experimental groups for Pick and Place Task in order of increasing DOF. All groups use Haptic Master(HM) with the respective DOF.

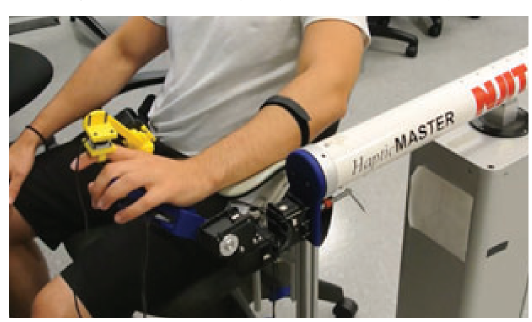


Figure. NJIT HandsOn System complete with 3 proximal DOF at the shoulder, 3 DOF at the wrist, and 1 DOF for pinch.

Narahara, T., and Kobayashi, Y., "Personalizing homemade bots with plug & play AI for STEAM education" SIGGRAPH Asia 2018 Technical Brief. https://doi.org/10.1145/3283254.3283270 [Video]

Personalizing Homemade Bots with Plug & Play AI for STEAM Education

Taro Narahara
College of Architecture and Design
New Jersey Institute of Technology (NJIT)
Newark, NJ, USA
taronarahara@gmail.com

Yoshihiro Kobayashi
School of Computing, Informatics, and Decision Systems Engineering, Arizona State University
Phoenix, AZ, USA
dr.yoshihiro.kobayashi@gmail.com

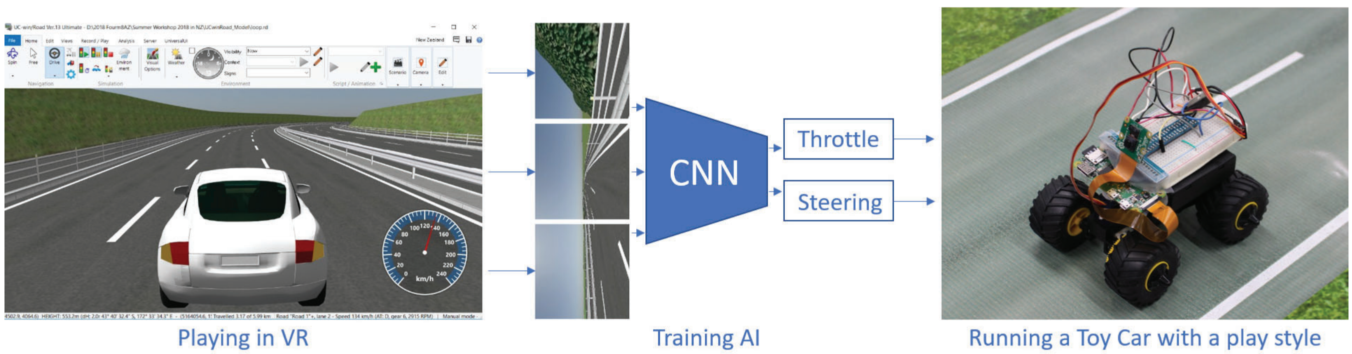


Figure 1: Three steps of the proposed module, 1) Playing and testing in the VR environment, 2) Training an AI model based on a dataset acquired from the virtual testing, and 3) Running a toy car using a trained AI model on a physical track.

ABSTRACT

In this study, we propose a new framework for hands-on educational modules to introduce ideas in AI and robotics casually, quickly, and effectively in one package for beginners of all ages in STEAM fields. Today, courses on introductory robotics are found everywhere, from K-12 summer camps to adult continuing education. However, most of them are limited to learning basic skills on sensor-actuator interactions due to their limited time and can rarely introduce what recent exciting AI can do, such as image recognition. As a case study to demonstrate the idea of the framework, an educational module to create a toy car with a camera controlled by Raspberry Pi is introduced. Our approach uses both physical and digital environments. Participants experience running their toy cars on a physical track using a convolutional neural network (CNN) trained based on how participants drive cars in a virtual game. The tested idea can be extensible as a framework to many other examples of robotics projects and can make ideas of AI and robotics more accessible to everyone. A proposed AI model is trained to assimilate the participant's game-play style in a VR environment which will be later re-enacted by the physical robot assembled by participants. Through this approach, we intend to demonstrate the AI's ability to personalize things and hope to stimulate participants' curiosity and motivation to learn.

SA '18 Technical Briefs , December 4–7, 2018, Tokyo, Japan

ACM ISBN 978-1-4503-6062-3/18/12...$15.00
https://doi.org/10.1145/3283254.3283270

CCS CONCEPTS

• **Computing education → Computing education programs; K-12 education; Adult education**

KEYWORDS

Physical computing, K-12, Adult Education, Machine Learning, AI, Robotics Education

ACM Reference format:

Taro Narahara and Yoshihiro Kobayashi. 2018. Personalizing Homemade Bots with Plug & Play AI for STEAM Education. In Proceedings of ACM In SIGGRAPH Asia 2018 Technical Briefs, 2018, Tokyo, Japan. ACM, New York, NY, USA, 4 pages.

1 INTRODUCTION

Today, courses on introductory robotics are found everywhere, from K-12 summer camps to adult continuing education. However, most of them are limited to learning basic skills on sensor-actuator interactions due to their limited time and can rarely introduce how recent exciting AI can be applied to their handmade robots. There are also courses on AI and programming but mostly focused only on coding without robotics. Authors have taught similar courses for K-12 to college students in various majors including design [Narahara 2015] and have experienced this challenge to introduce both robotics and AI in a short term.

The advantages including speed and safety of training robot control systems in simulation and transferring them to physical robots have been identified and practiced by engineers and developers, for example, the Isaac SDK by NVIDIA corporation [2018] for developers. There are some existing works such as ML-Agents for Unity3D [Unity Technologies 2018] that reveal the power of AI through gaming without overwhelming users with lengthy codes. However, there are not many effective teaching tools to connect instruction on electronics and fabrication for robotics and programming for AI.

In this on-going study, we propose a new framework for hands-on educational modules to introduce ideas in AI and robotics casually, quickly, and effectively in one package for beginners of all ages in STEAM (Science, Technology, Engineering, Art, and Mathematics) fields. As a case study to demonstrate our idea of the framework, we tested an educational module to create a toy car with a camera controlled by Raspberry Pi with a small number of adults at this stage as a preparation for more formal classroom settings.

2 METHOD

The following sections describe our pipeline for the proposed module. In the first half of the module, participants assemble a physical robotic prototype, a model of a toy car in this case. This part is similar to a typical existing workshop using microcontrollers, sensors, and actuators for beginning learners to acquire a basic knowledge and skills in programming and electronics through hands-on physical computing exercises.

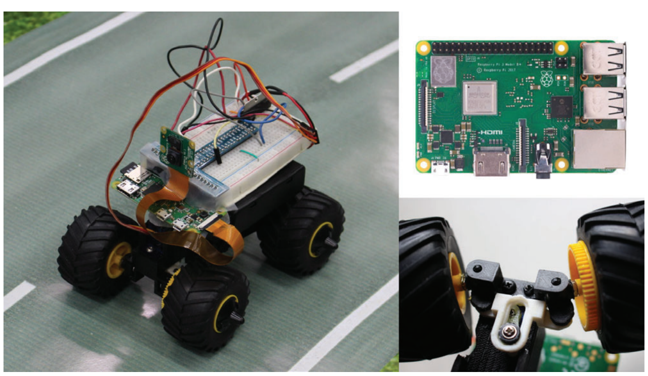


Figure 2: A toy car model (left), a Pi 3 Model B+ (top right), and a 3-D printed steering mechanism (bottom right).

2.1 Assembling and Programming a Toy Car

A single-board computer, Raspberry Pi was selected to be used for this project due to its ease of use, availability of online resources, and its higher processing speed and memory size compared to other microcontrollers such as an Arduino Uno (we tested with Pi 3 Model B+ and Pi Zero W) [Raspberry Pi 2018]. A Raspberry Pi uses a Python programming language which has many useful libraries including TensorFlow, an open source machine learning framework [Google LLC]. Nowadays, TensorFlow is widely used for AI applications including generation of deep neural networks, and, in principle, it is possible to plug and play (run and test) networks pre-trained in other higher-performing computer environments on a Raspberry Pi within the capacity that it can handle. We pre-loaded Raspberry Pi with TensorFlow for this project in order to extend the potential to introduce more AI aspects into robotics workshops for STEAM fields.

Figure 3: Game-play scene in the VR environment and captured three images

For a base design of a toy car, we used a chassis and gearbox for a DC motor from a commercially available 1/32-scale plastic car model by Tamiya, Inc. [2018]. The rotation speed of the rear wheels is controlled by an H-bridge driver with output voltage control, Toshiba TA7291P. The front wheels were modified with a custom-made 3-D printed steering mechanism by the authors, and its angle can be changed by an attached small 9g servo motor. A Python code on a Raspberry Pi controls speed and steering angle of the car by sending out appropriate signals from its GPIO (general-purpose input/output) pins to the motors. The Raspberry Pi Camera Module v2, which supports up to 3240x2464 pixels for static images and a 1080p30 video mode, is attached to the toy car for later uses for visual processing with a network on TensorFlow. The whole circuit was made on a solderless breadboard so that anyone can quickly complete it through learning, testing, and modifying simple template codes. We also provided a code with a mouse-based graphic user interface to control speed and steering angle of the car for fun at this warm-up phase.

After completing the part to produce a physical prototype, our approach has three steps, 1) Playing and testing in the VR environment, 2) Training an AI model based on a dataset acquired from the virtual testing, and 3) Running a toy car using a trained AI model on a physical track.

2.2 Playing in the VR Environment

In the first step, participants have fun driving 3-D representations of their toy cars (which was physically assembled previously) in the virtual reality (VR) environment using the off-the-shelf software, UC-win/Road by Forum8 Co., Ltd. [2018]. UC-win/Road has many easy-to-use user-friendly features to generate 3-D roads with cars and traffic for professional driving simulations. These features allow us to quickly create a custom driving track with waypoints marked and make this experience especially fun for K-12 students. Simply pressing three keys – throttle, left, and right for steering angle – can drive a car. Once a track is created, participants are asked to drive for a few minutes. The system built using its SDK saves three 128x128-pixel images from three virtual cameras facing front, left, and right of the car along with values for throttle and steering angle based on user's keystrokes and velocity of the car at each frame and recodes 4000 frames (about 2 minutes) in a CVS file format for the next training step. (Figure 4) (A custom plugin to capture the driving scene as a list of JPG images has been developed using its SDK as its built-in recording function was limited to exporting a movie file.)

1	left image	center image	right image	Speed	Steer	Throttle
2	Camera_Left\AVIJpg\000001.jpg	Main_View\AVIJpg\000001.jpg	Camera_Right\AVIJpg\000001.jpg	10.33106804	0.000185331	0.300000012
3	Camera_Left\AVIJpg\000002.jpg	Main_View\AVIJpg\000002.jpg	Camera_Right\AVIJpg\000002.jpg	10.36739731	0.000211761	0.300000012
4	Camera_Left\AVIJpg\000003.jpg	Main_View\AVIJpg\000003.jpg	Camera_Right\AVIJpg\000003.jpg	10.40373707	0.000238379	0.3125
5	Camera_Left\AVIJpg\000004.jpg	Main_View\AVIJpg\000004.jpg	Camera_Right\AVIJpg\000004.jpg	10.44168282	0.000265111	0.318750024
6	Camera_Left\AVIJpg\000005.jpg	Main_View\AVIJpg\000005.jpg	Camera_Right\AVIJpg\000005.jpg	10.48043346	0.000291917	0.321875036
7	Camera_Left\AVIJpg\000006.jpg	Main_View\AVIJpg\000006.jpg	Camera_Right\AVIJpg\000006.jpg	10.51958942	0.000216673	0.323437512
8	Camera_Left\AVIJpg\000007.jpg	Main_View\AVIJpg\000007.jpg	Camera_Right\AVIJpg\000007.jpg	10.55894947	8.30E-05	0.32421875
9	Camera_Left\AVIJpg\000008.jpg	Main_View\AVIJpg\000008.jpg	Camera_Right\AVIJpg\000008.jpg	10.59841442	-5.10E-05	0.336718738
10	Camera_Left\AVIJpg\000009.jpg	Main_View\AVIJpg\000009.jpg	Camera_Right\AVIJpg\000009.jpg	10.63948917	-0.00018554	0.342968762
11	Camera_Left\AVIJpg\000010.jpg	Main_View\AVIJpg\000010.jpg	Camera_Right\AVIJpg\000010.jpg	10.68137932	-0.00032067	0.346093774
12	Camera_Left\AVIJpg\000011.jpg	Main_View\AVIJpg\000011.jpg	Camera_Right\AVIJpg\000011.jpg	10.72367001	-0.000456356	0.344531268
13	Camera_Left\AVIJpg\000012.jpg	Main_View\AVIJpg\000012.jpg	Camera_Right\AVIJpg\000012.jpg	10.76577187	-0.000561867	0.34375
14	Camera_Left\AVIJpg\000013.jpg	Main_View\AVIJpg\000013.jpg	Camera_Right\AVIJpg\000013.jpg	10.80777168	-0.000441964	0.331250012
15	Camera_Left\AVIJpg\000014.jpg	Main_View\AVIJpg\000014.jpg	Camera_Right\AVIJpg\000014.jpg	10.84816742	-0.000321573	0.324999988
16	Camera_Left\AVIJpg\000015.jpg	Main_View\AVIJpg\000015.jpg	Camera_Right\AVIJpg\000015.jpg	10.88776302	-0.000200695	0.321874976
17	Camera_Left\AVIJpg\000016.jpg	Main_View\AVIJpg\000016.jpg	Camera_Right\AVIJpg\000016.jpg	10.92695999	-7.93E-05	0.3203125
18	Camera_Left\AVIJpg\000017.jpg	Main_View\AVIJpg\000017.jpg	Camera_Right\AVIJpg\000017.jpg	10.96595955	4.24E-05	0.319531262
19	Camera_Left\AVIJpg\000018.jpg	Main_View\AVIJpg\000018.jpg	Camera_Right\AVIJpg\000018.jpg	11.00486183	0.000164546	0.307031274
20	Camera_Left\AVIJpg\000019.jpg	Main_View\AVIJpg\000019.jpg	Camera_Right\AVIJpg\000019.jpg	11.04215431	0.000223824	0.30078125

Figure 4: A CVS file with a list of recorded images, speed, steering angle, and throttle.

2.3 Training and Testing the AI Model

The next step is to train a convolutional neural network (CNN) using the driving dataset acquired from the previous step. The CNN introduced by NVIDIA corporation [2016], which has been proven to work well in driving applications, is widely used as a base model for several open source projects and tutorials [Ravel 2018; Shibata 2018; Udacity 2017]. Our code and network model is based on such a project, "Self-Driving Car," offered at Udacity [2017]. Tracked records of how a participant controlled the virtual car under which visual condition of the road captured by a virtual camera – which replicates the physical camera on a Raspberry Pi – are used for the training of the CNN. The input data is a set of three 128x128-pixel images, and the output data is a set of steering, throttle, and speed. The training of the network involves 10 epochs with 20,000 updates in each epoch to minimize the error in the output, and we select the resulting network with least error rate. It takes approximately 20 minutes for each epoch using a conventional Windows 10 computer with an Intel i7 CPU and NVIDIA GTX1070 GPU. We normally obtained the best result at the 8^{th} or 9^{th} epoch and an increased error rate afterward due to the overfitting. The code outputs the data of the newly weighted network in the HDF5 binary data format which will be later used for test driving in both virtual and physical environments. For further details of the network, please refer to the reference [Udacity 2017].

First, after the training, we tested the trained network in the VR environment to see how it can drive a virtual car. The driving software sends an image from three cameras on the car in real time through TCP/IP socket to the AI server developed on Python. Then, the server returns the expected output values for throttle and steering based on the received input image back to the VR software. They work in tandem to control the vehicle using the driving instruction obtained from the trained network. The communication time between the VR environment and the AI server can be established in approximately 0.033 second per frame which is sufficient for our project.

2.4 Running a Toy Car using a trained AI model

The final step is to make a physical toy car run using the trained network. Participants transfer the HDF5 file containing the acquired trained network to the Raspberry Pi's memory on the robotic prototype for testing in the physical environment. The physical toy car is controlled by the network – which was trained to use the way the participant drove the virtual car – and was tested on a miniature track that carefully replicates the visual appearance of the virtual environment, yet different in terms of the course configuration. We plotted a physical miniature track with textures of the roads and background from the VR scene in a batch of eight 24"x48" sheets and laid them on the floor connected for the test run. A code in the Raspberry Pi adjusts the output values of the network for throttle and steering in the VR to the corresponding physical scale for the rotation angle of the servo motor and voltage output for the DC motor. The network outputs the steering angle in the range between -1 and 1, which is mapped to the rotation range of the servo motor between 75 and 105 degrees. The speed of the toy car is simply dictated by the rotation speed of the DC motor. At this stage, the calibration and mappings between the virtual actuator values from the network and the physical ones have been done based on trial and error, and further improvements are expected. We managed to run the car, staying on the track, using the networks trained based on good driving results in VR. In contrast, networks using bad driving results in VR occasionally went off the track.

Figure 5: A test run using a toy car model on the miniature track

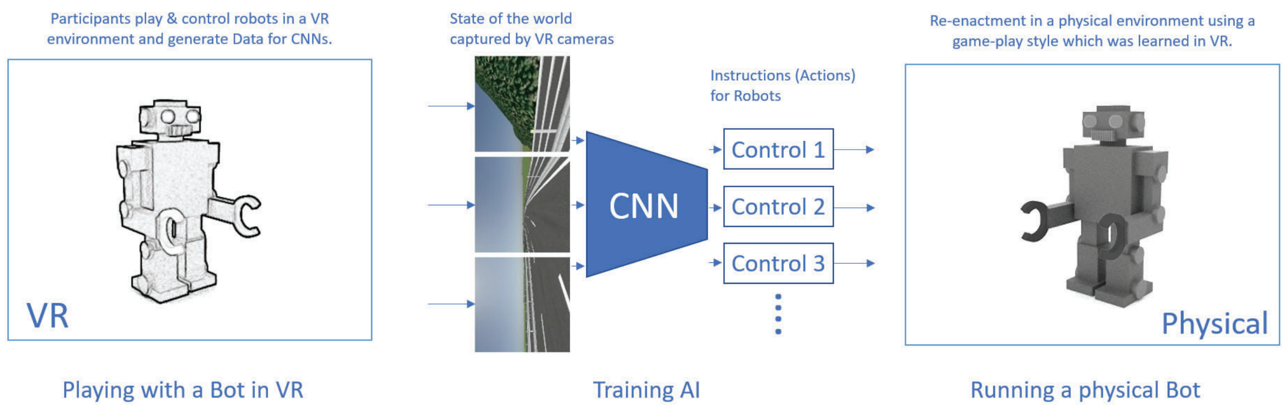


Figure 6: A The tested idea can be extensible as a framework to many other examples of robotics projects, such as controlling a soccer-playing robot. Our AI model is trained to assimilate the participant's game-play style, which will be later re-enacted by the physical robot assembled by the participant.

As there is still a discrepancy in the relative scale of a car and a track in the virtual and physical environments, further adjustments for the toy-car control UI is required to improve the performance. As the toy car has only one camera, we decomposed the single image into three 128x128 pixel images and fed them as inputs for the network since the camera can capture images with a sufficiently high resolution. For future work, we consider using a fish-eye camera to capture a wider area or three cameras to match the input images more accurately to those in VR.

3 CONCLUSIONS AND FUTURE WORK

In this ongoing project, we tested our proposed framework for an educational module that can introduce AI and robotics together in one package in a short period of time through the example of AI driving with a toy car. The tested idea can be extensible as a framework to many other examples of robotics projects, such as controlling a soccer-playing robot, and can make ideas of AI and robotics more accessible to everyone. The current state of the art for single-board computers still has a practical limitation for a size of networks to run in real time, and further advancement will help us introduce more complex application examples. In terms of the driving model, for future work, we can compare the reactions of our AI models to those trained by professional drivers and could develop a coaching system. The physical miniature course needs to be made visually closer to VR scenes to minimize the gap for a more accurate performance of the network. Our AI model is trained to assimilate the participant's game-play style, which will be later re-enacted by the physical robot assembled by the participant. Through this approach, we intend to demonstrate the AI's ability to personalize things and hope to stimulate participants' curiosity and motivation to learn.

ACKNOWLEDGMENTS

The authors would like to thank Yuji Ito, Chikako Takei, Yoann Pencreach, and Shunta Shimizu at Forum8 Co., Ltd. for the generous support on their VR software, UC-win/Road.

REFERENCES

Forum8 Co., Ltd. 2018. UC-win/Road: Professional Software for driving simulation, http://www. forum8.co.jp.
Google LLC, 2018. TensorFlow^{TM}: An open source machine learning framework for everyone, https://www.tensorflow.org
Narahara, T., 2015. Design Exploration through interactive prototypes using sensors and microcontrollers, In Computers & Graphics: An International Journal of Systems & Applications in Computer Graphics, Elsevier Science & Technology, vol. 50 (2015), pp. 25-35.
NVIDIA Corporation, 2016. End-to-End Deep Learning for Self-Driving Cars, https://devblogs.nvidia.com/deep-learning-self-driving-cars
NVIDIA Corporation, 2018. NVIDIA Isaac SDK: Accelerate Your Creation of Autonomous Machines, https://developer.nvidia.com/isaac-sdk
Raspberry Pi, 2018. https://www.raspberrypi.org/products/camera-module-v2
Raval, S., 2018. How to Simulate a Self-Driving Car, https://github.com/llSourcell/How_to_simulate_a_self_driving_car
Shibata, N., 2018. car-behavioral-cloning, https://github.com/naokishibuya/car-behavioral-cloning
Tamiya Inc., 2018. https://www.tamiyausa.com
Udacity, 2017. The Udacity open source self-driving car project, https://github.com/udacity/self-driving-car
Unity Technologies, 2018. ML-Agents: An Open Source Unity Machine Learning Agents, https://unity3d.com/machine-learning

Professional Works

Entry Structure, Tokyo Japan

Gluckman Mayner Architects. 2000 - 2003

Professional Experience in Architecture 1997 – present

Licensed and Registered Architect in NY state and Japan

Selected Professional Works in the following pages

Gluckman Mayner Architects. New York, NY 2000-2005

Mori Art Center. Tokyo, Japan 2000-2003

Role: Project Architect under design principals and a project manager from schematic design through to the opening of the museum for the design of Entry Structure (Cable-net-shell structure pavilion), 30,000sqf. Exhibition spaces, and coordination between Japanese local architects and general construction companies (Mori Building Co. Irie-Miyake Architects, JV: Kajima & Obayasi)

MoMA Store. New York 2004

Role: One of the three project architects for interior design of the 5,700 total square-feet store and the fixture designs from schematic design to construction document phase.

Hotel Puerta de America. Madrid, Spain 2003-2004

Role: Project Architect for Interior design of hotel rooms, suites and common spaces in 14-storey-building in Madrid from schematic design to construction document phase under design principals and a project manager.

Museo Picasso Malaga, Madrid, Spain 2004

Role: Project Team member. Schematic Design and Design Development phases. Worked on a publication of a monograph of the project with 2X4 Inc. (Construction Documents were produced by local architects in Spain with GMA.)

Philadelphia Museum of Art Annex. Pennsylvania 2004

Role: Project Team. Construction Document phase.

Vassar College – Kenyon Hall Renovation. Poughkeepsie, NY 2002

Role: Project Team. Construction Document phase.

Skidmore Owings and Merrill LLP. New York, NY 1997-2000

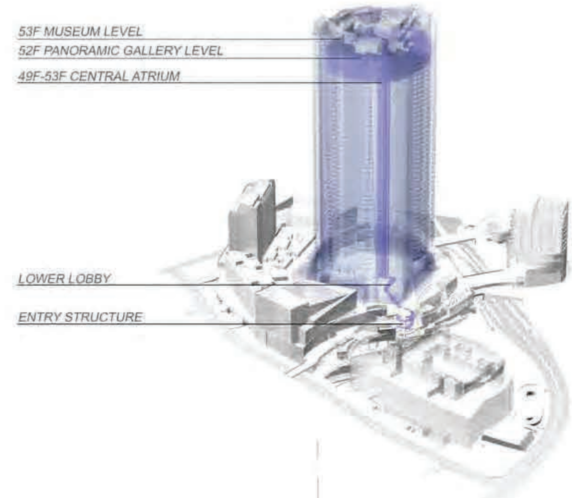
Kuwait Police Academy, Kuwait (Principal: Roger Duffy) 1997-2000

Role: Design Team member for Master plan, schematic design, and design development phases for total 4.5 million square feet campus design.

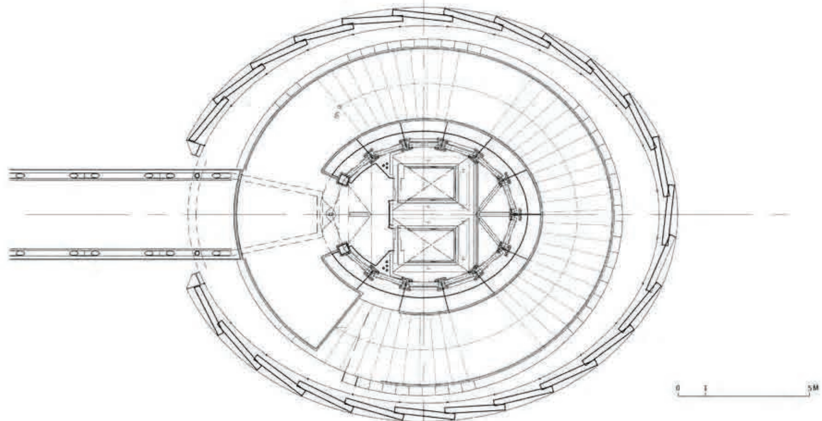
MORI ART CENTER, TOKYO
GLUCKMAN MAYNER ARCHITECTS
2000-2003

General Data:
The Museum, comprised of 100,000 total square-feet, will occupy the top two floors of a fifty-three-story office tower designed by Kohn Pederson Fox. The museum will encompass 35,000 square-feet of exhibition space, an observation gallery with panoramic views of Tokyo, a museum shop, restaurant café, administrative offices and art handling spaces. A separate structure at the base of the tower will provide a distinct and iconic entrance to the Museum.
A 100-foot-high entry pavilion has a unique structural system of cable-net-shell with enclosure of 60-foot-tall singled-glass cone. The pavilion takes visitors from the vehicular drop-off and shopping plaza levels at the base of the complex, up 3 to 5 floors to a seventy-foot-long bridge that leads into Mori office tower. Once inside the office building, visitors can get information about the museum at the lower lobby, then catch express elevators to the 52nd floor, where the museum proper begins.

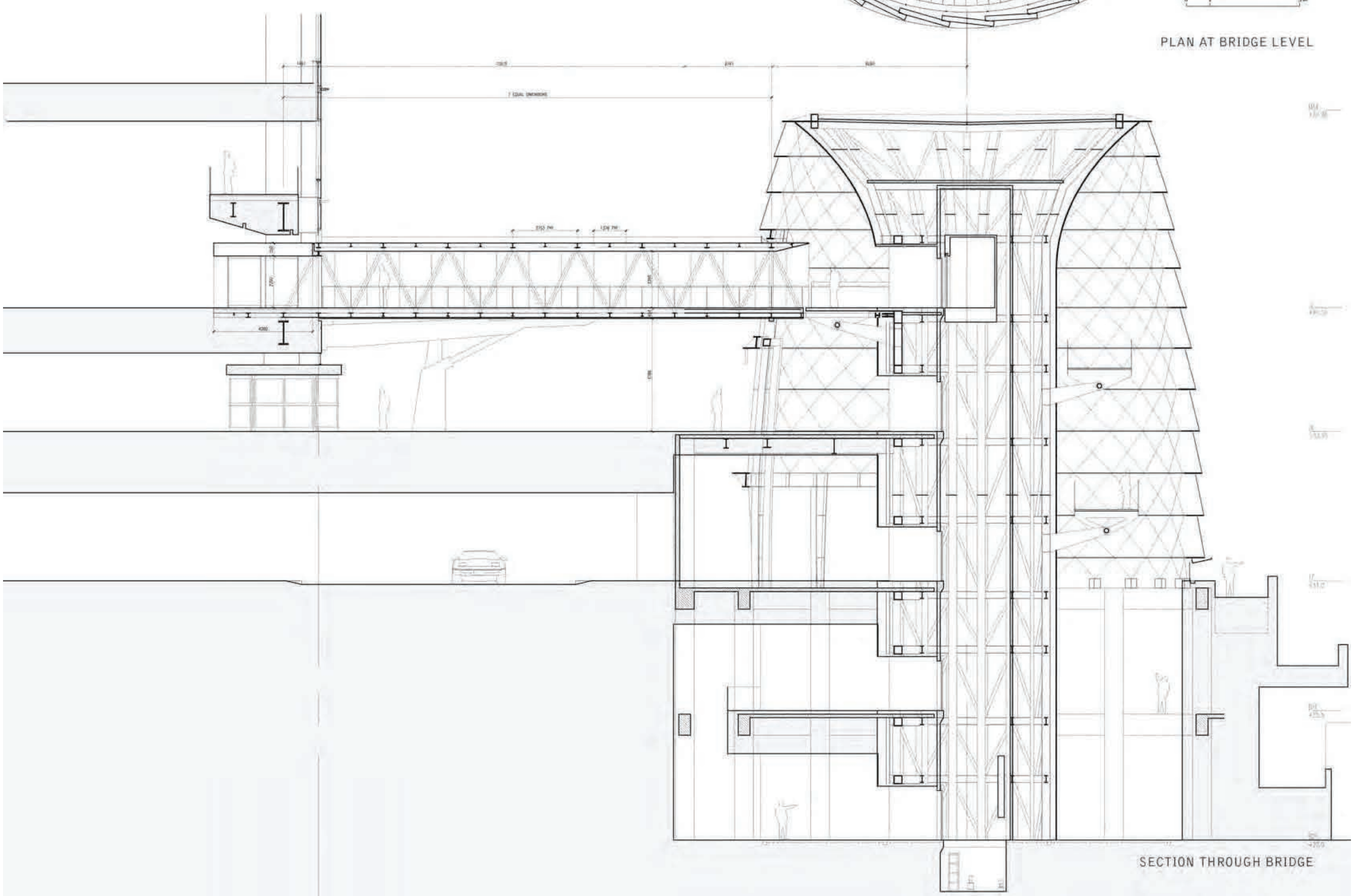
Contribution:
Project Architect from schematic design through to opening of the Museum. The work included design of Entry Structure (Cable-Net-Shell structure entry pavilion), Exhibition space, Observation gallery, Atrium space between 49th and 53rd floor and coordination between Japanese local architects & general construction companies (Mori Building Co. Irie Miyake Architects, JV:Kajima & Obayasi)



SITE OVERVIEW

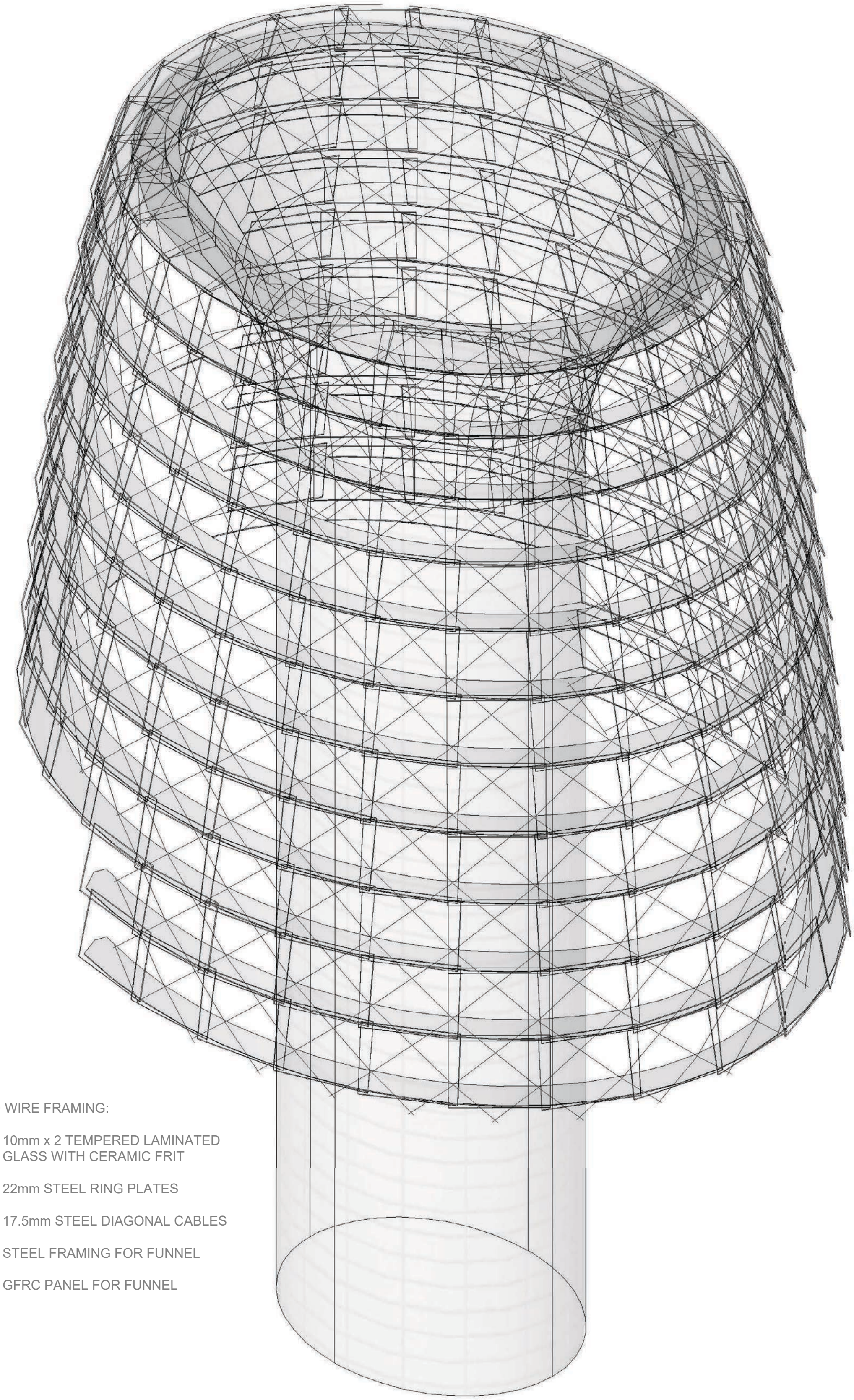


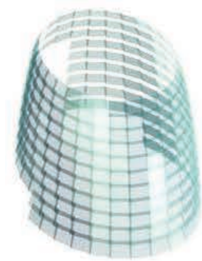
PLAN AT BRIDGE LEVEL



SECTION THROUGH BRIDGE

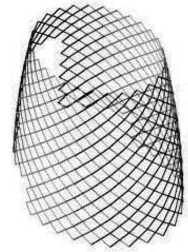


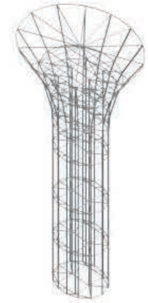


1.

2.



3.



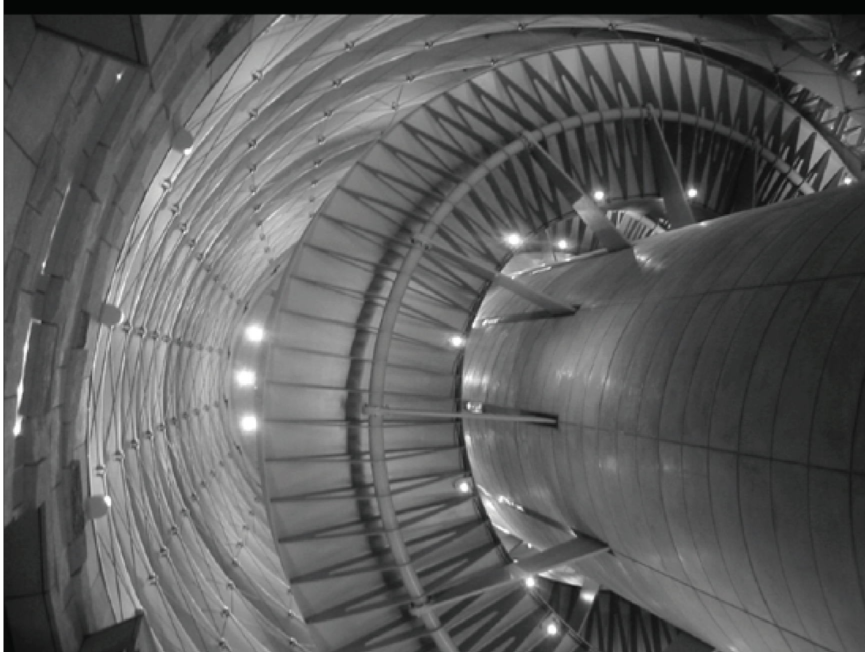
4.

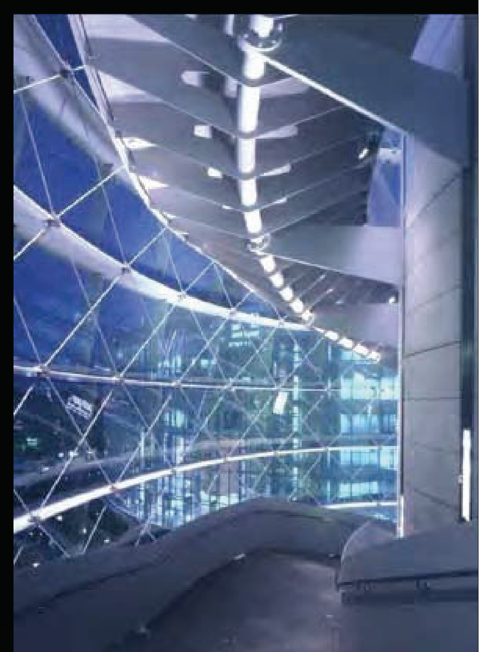
5.

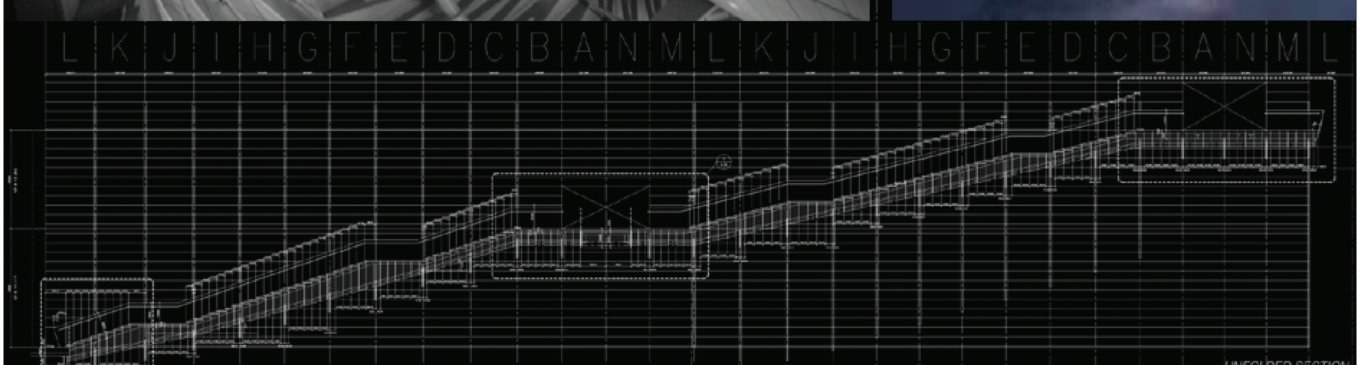
3D WIRE FRAMING:

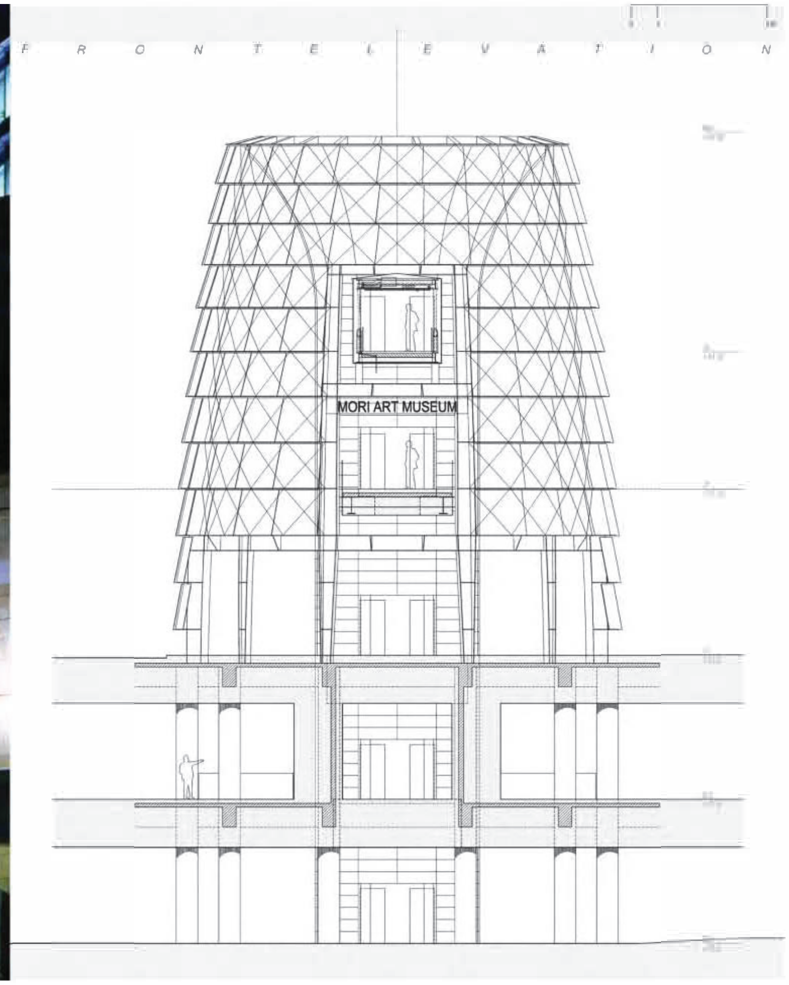
1. 10mm x 2 TEMPERED LAMINATED GLASS WITH CERAMIC FRIT
2. 22mm STEEL RING PLATES
3. 17.5mm STEEL DIAGONAL CABLES
4. STEEL FRAMING FOR FUNNEL
5. GFRC PANEL FOR FUNNEL

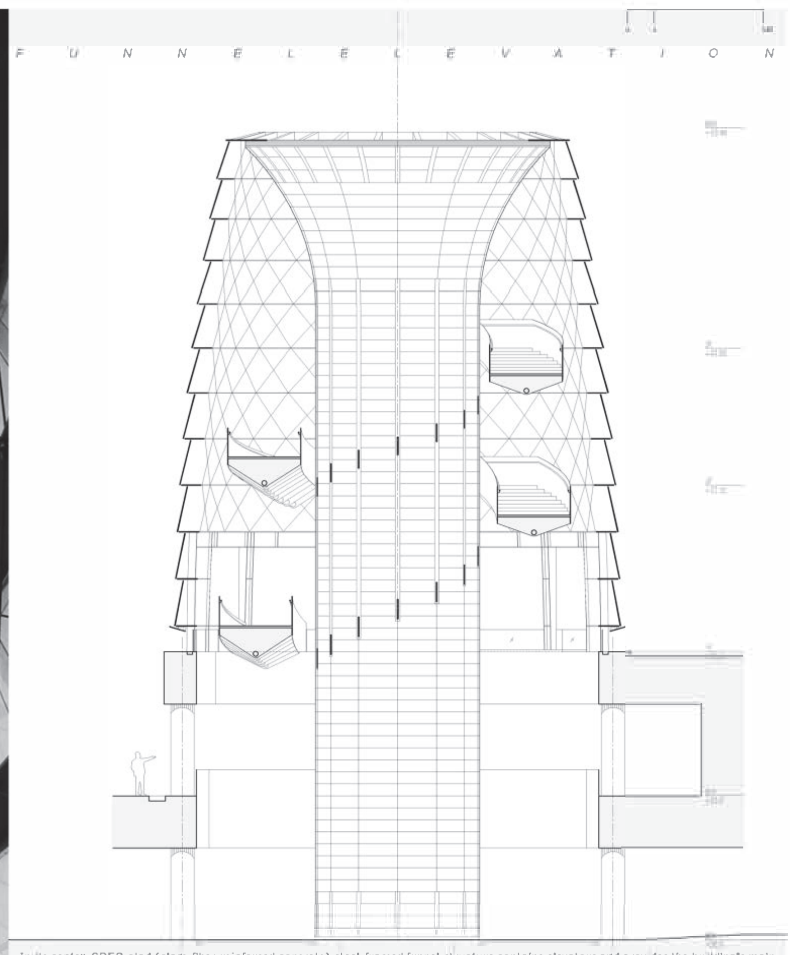
2000 – 2003
Mori Art Center
Gluckman Mayner Architects





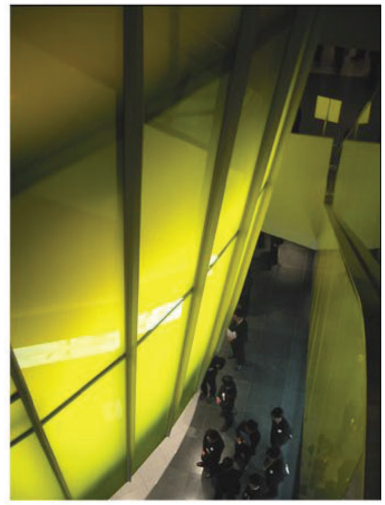






In its center, GRFC-clad (glass fiber reinforced concrete) steel-framed funnel structure contains elevators and provides the building's main vertical support. Slots at the funnel contain supply and return air grills, opening for the stair brackets, lighting and other equipments

TARO NARAHARA

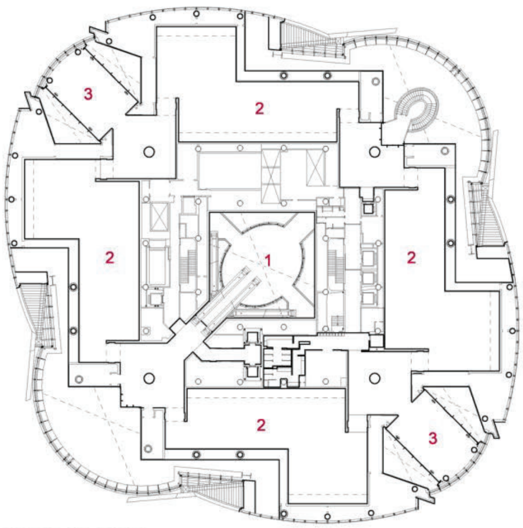


MORI ART MUSEUM 52-53FL

GLUCKMAN MAYNER ARCHITECTS
2000-2003

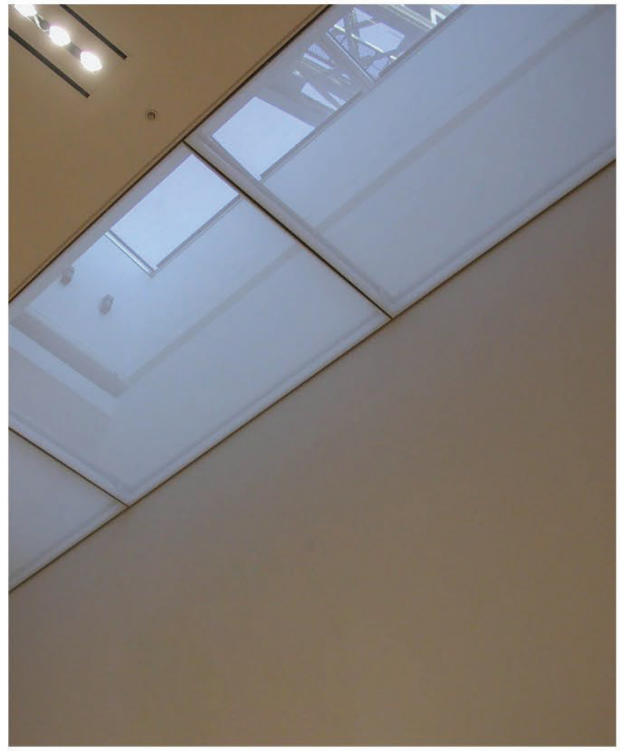
ENGINEER:
DEWHURST McFARLANE AND PARTNERS
ENGINEERING NETWORK.

GENERAL CONTRACTOR:
OBAYASHI, KAJIMA (joint venture)

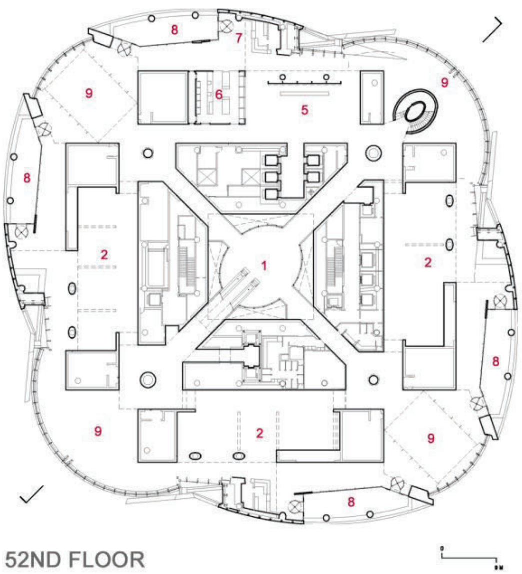


53RD FLOOR

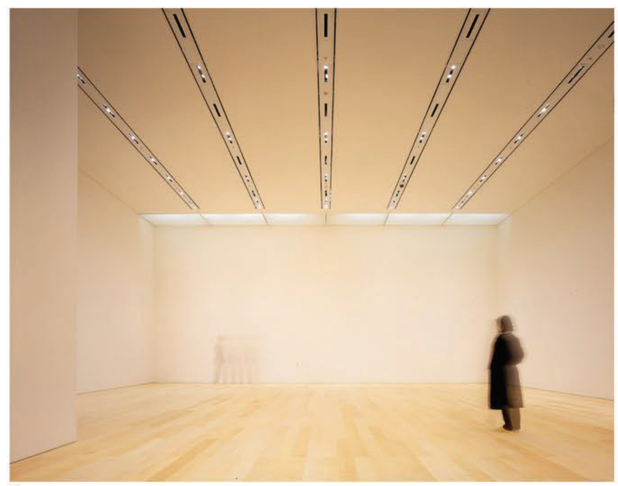
1. ATRIUM
2. GALLERY
3. ART & TECH. GALLERY
4. ELEVATOR LOBBY
5. ARRIVAL LOBBY
6. MUSEUM SHOP
7. CAFE
8. TERRACE
9. OBSERVATION



1



52ND FLOOR



2

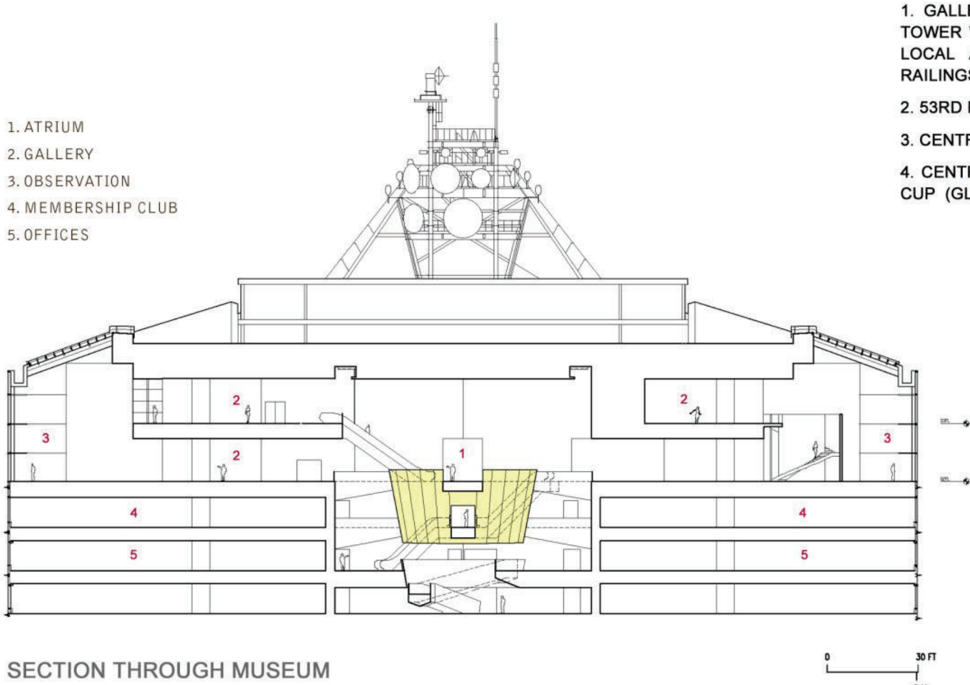
1. GALLERY SKYLIGHT: HAVING SKYLIGHT AT THE TOP OF 53-STORY-TOWER WAS A ENORMOUS CHALLENGE FOR BOTH DESIGN TEAM AND LOCAL ARCHITECTS, DUE TO THE MECHANICAL EQUIPMENT AND RAILINGS FOR THE MAINTENANCE GONDORA AT THE ROOF LEVEL.

2. 53RD FLOOR GALLERY

3. CENTRAL ATRIUM AT 52ND FLOOR

4. CENTRAL ATRIUM AT 51ST FLOOR BRIDGE: LOOKING INTO THE DIXY-CUP (GLASS-CLAD CIRCULATION CHAMBER)

1. ATRIUM
2. GALLERY
3. OBSERVATION
4. MEMBERSHIP CLUB
5. OFFICES

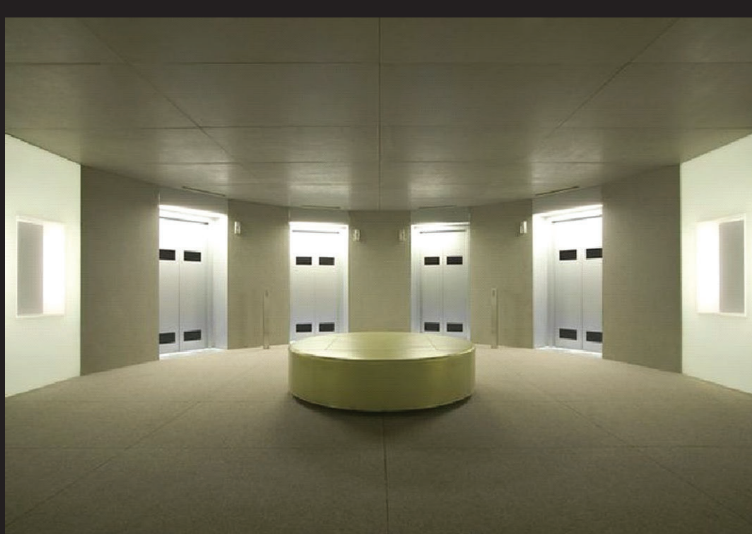


SECTION THROUGH MUSEUM

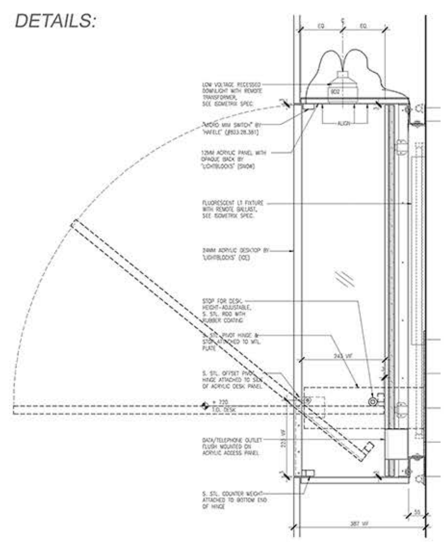
Hotel Puerta America
Spain, Madrid 2005

Gluckman Mayner was invited to design twenty-eight standard rooms and two suites in this concept hotel, which features twelve floors, each by a different designer. The initial architectural concept derived from the capsule hotel notion of a "box within a box." This concept is used to organize the approach to the different kinds of activities that take place within the small space of the hotel room. Each room features "rich" materials – such as luxurious fabrics in metallic tones – juxtaposed with "poor" materials such as cement board. In addition, natural substances like mica-flecked cement, felted wool, and Spanish granite are used alongside artificial ones like acrylic and recycled plastic.

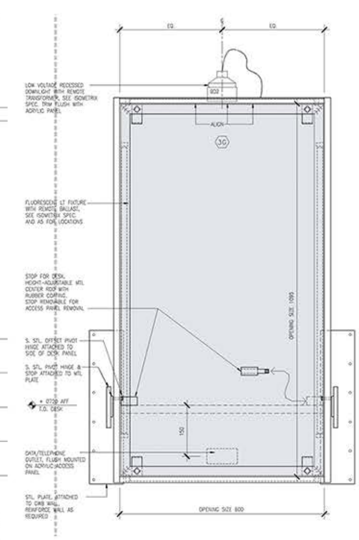
My Role:
Project Architect for the interior design of hotel rooms, suites, and common spaces in a 14-storey building in Madrid from schematic design to construction document phase.



DETAILS:

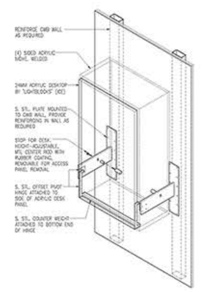


SECTION DETAIL: FOLDABLE TABLE NICHE

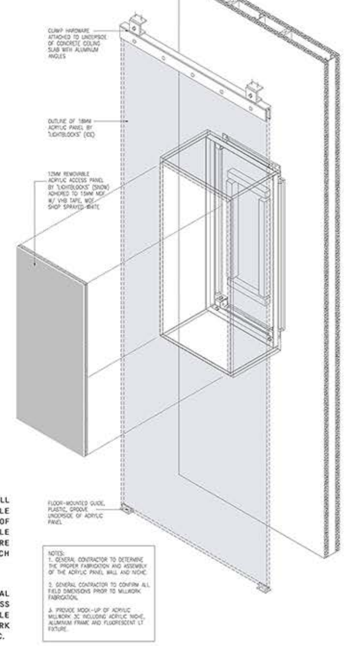


ELEVATION DETAIL: FOLDABLE TABLE NICHE

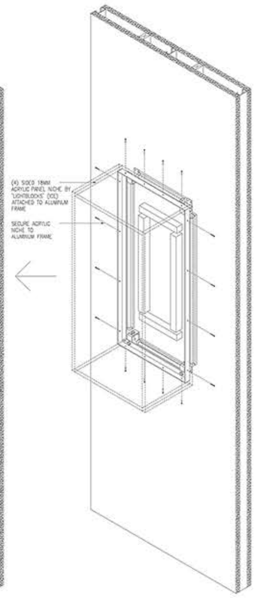
T R A N S L U C E N T W A L L



UNFOLDABLE TABLE NICHE DIAGRAM



INSTALLATION DIAGRAM 2: TRANSLUCENT WALL



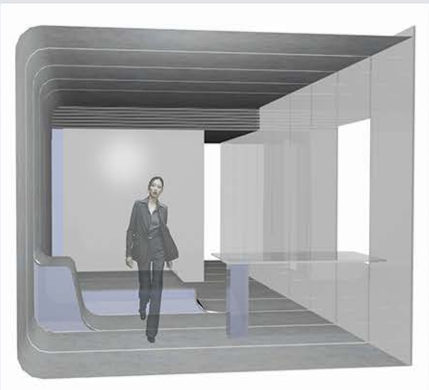
INSTALLATION DIAGRAM 1: ACRYLIC NICHE

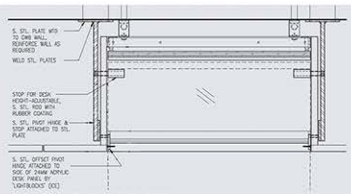
NICHE DETAIL:

INTENTION OF THE DETAIL WAS TO CONCEAL ALL THE ATTACHMENT HARDWARES FROM VISIBLE AREA TO ACHEIVE MONOLITHIC APPEARANCE OF THE TRANSLUCENT WALL AND NICHE, WHILE PROVIDING ACCESS TO THE LIGHTING FIXTURE BEHIND THE NICHE. USE MAGNETIC PUSH LATCH TO RELEASE ACCESS PANEL.

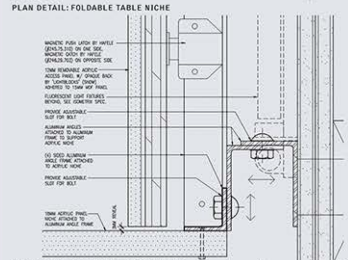
BELOW:
LIFE-SIZE MOCK-UP ON SITE. DUE TO THE LOCAL FIRE CODE, TRANSLUCENT WALL WILL BE GLASS FROM ACRYLIC. ATTEMPT TO REGISTER WHOLE WALL+NICHE CONSTRUCTION AS A MILLWORK WAS REJECTED. NICHE WILL STAY AS ACRYLIC.

RIGHT:
RENDERING VIEW OF THE HOTEL ROOM



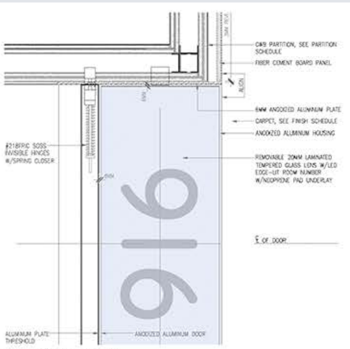


PLAN DETAIL: FOLDABLE TABLE NICHE

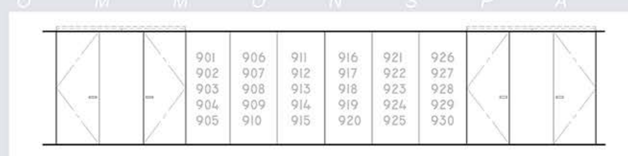


ATTACHMENT HARDWARE SECTION DETAIL

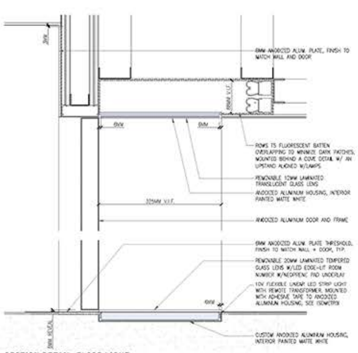
C O M M O N S P A C E



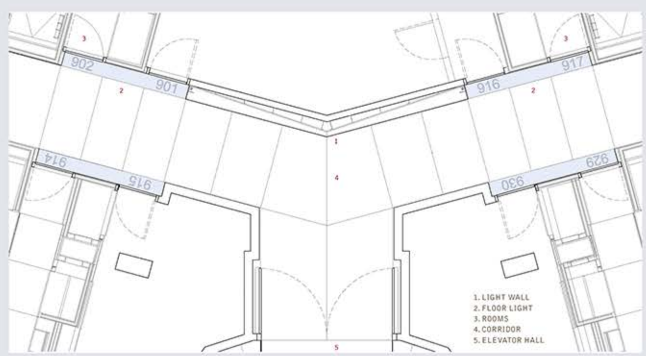
PLAN DETAIL: FLOOR LIGHT



LIGHT WALL ELEVATION



SECTION DETAIL: FLOOR LIGHT

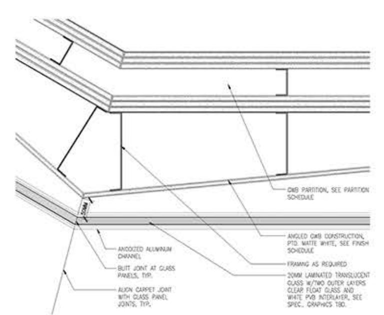


FLOOR PLAN

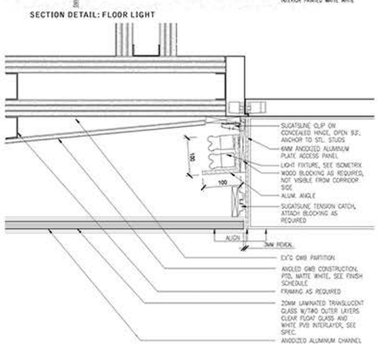
LIGHT WALL / FLOOR LIGHT DETAIL

LIGHTING AND ROOM SIGNAGES ARE INTEGLATED INTO PLANER SURFACE ON EITHER FLOOR OR WALL. COLLABORATION WITH ISOMETRIX ENABLED US TO USE EFFECTIVE LIGHTING SCHEME, SUCH AS USE OF FLEXIBLE LINEAR LED STRIPS.

LIGHTING CONSULTANT: ISOMETRIX, LONDON.



LIGHT WALL: PLAN DETAIL

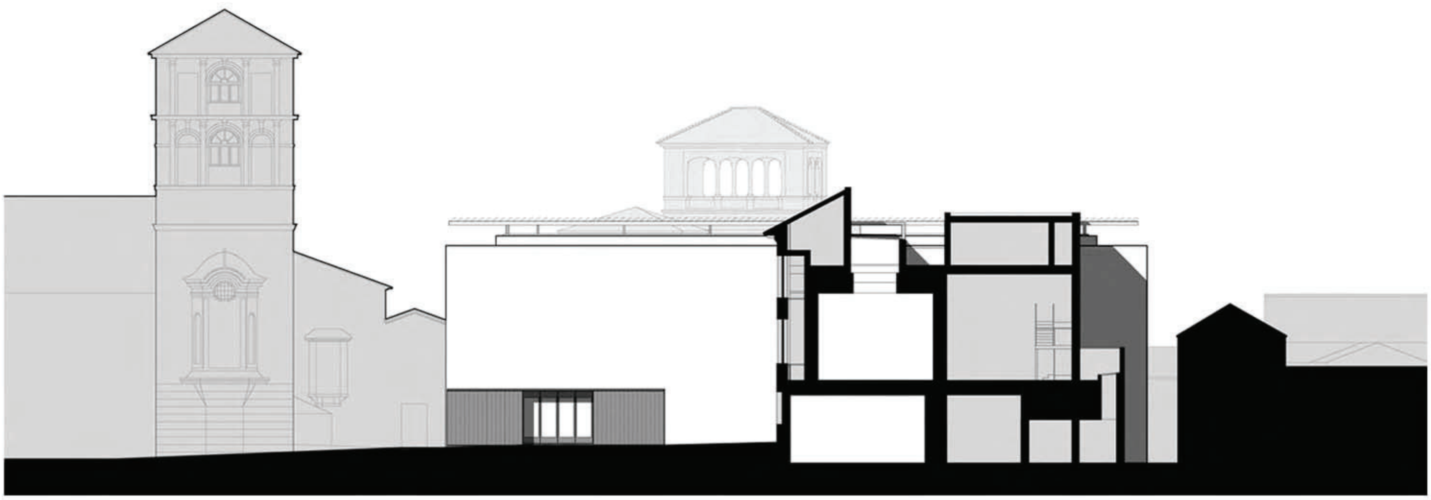


VIEW FROM THE ELEVATOR LOBBY

Museo Picasso Malaga
Malaga, Spain 2004

This museum is located in the historic city center of Malaga, birthplace of Pablo Picasso. A 16th-century palace was fully restored to house the Main Entry and Permanent Collection galleries. Six new buildings were carefully inserted into the urban fabric to house the ambitious program, totaling 90,000 square feet. A large building, containing the Special Exhibition galleries, and a series of smaller buildings, containing ancillary program, create the boundaries of a new public plaza. The new structures respect the scale, texture and articulation of the existing built context, while their simple geometric forms, rendered in white plaster, clearly announce a sympathetic modern intervention.

Office: Gluckman Mayner Architects
Role: Worked as one of the project Architects in the project team on Schematic design, design development, and the publication of a monograph with 2X4 Inc.



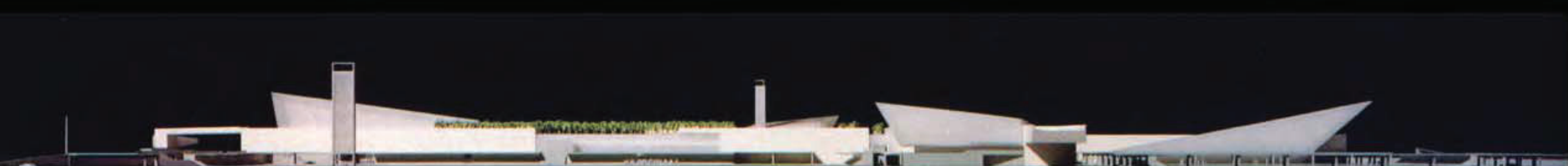
Kuwait Polica Academy, 1997-2000, Skidmore, Owings, and Merrill, LLP

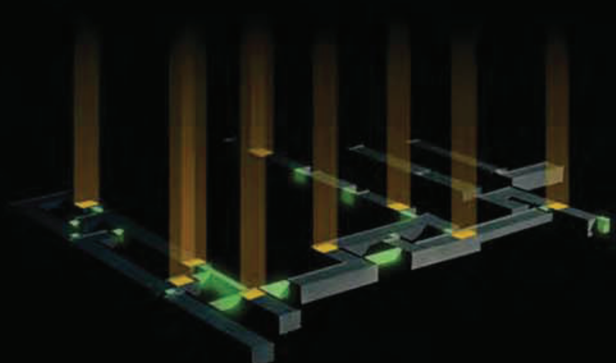
OBJECT BUILDINGS

COURTYARDS

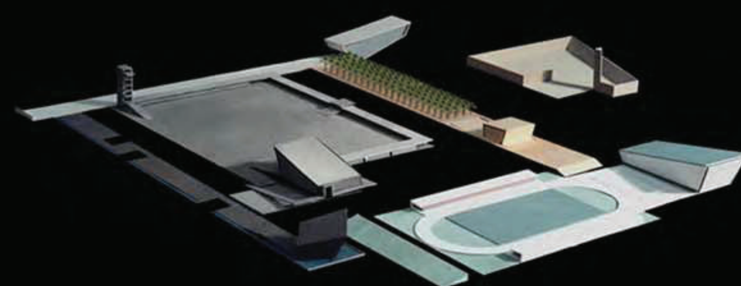
1. DORMITORIES
2. CLASS ROOMS
3. LIBRARY
4. ADMINISTRATION
5. DINNING HALL
6. NATATORIUM
7. AUDITORIUM
8. MOSQUE
9. PRACTICAL TRAINING COURT
10. ASSENBLY COURT
11. ATHLETIC COURT

ROPE BUILDINGS



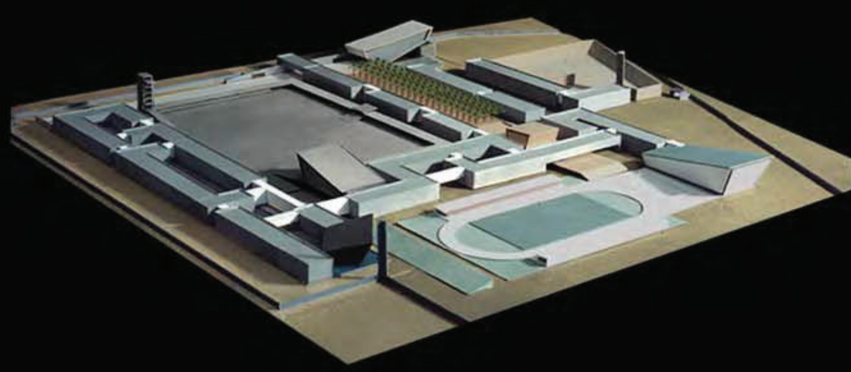


LIGHT CHAMBERS



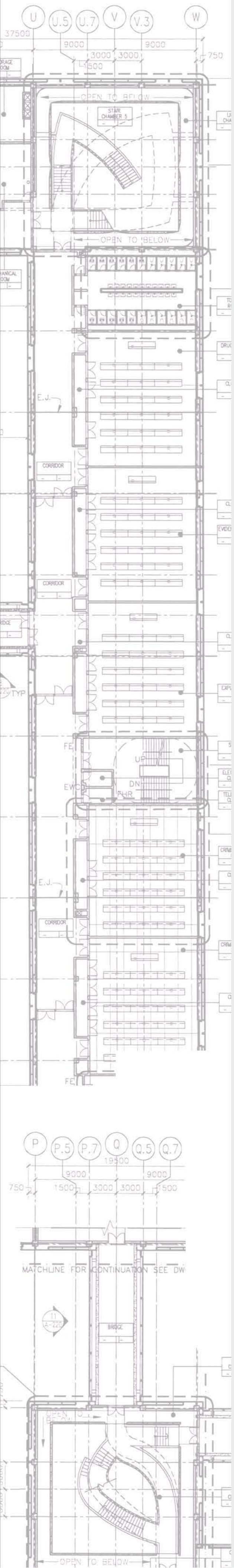
SPACES AND VESSELS

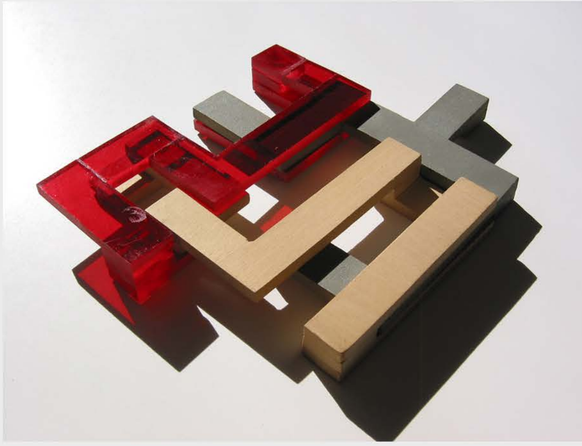
ROPE BUILDINGS



ASSEMBLAGE

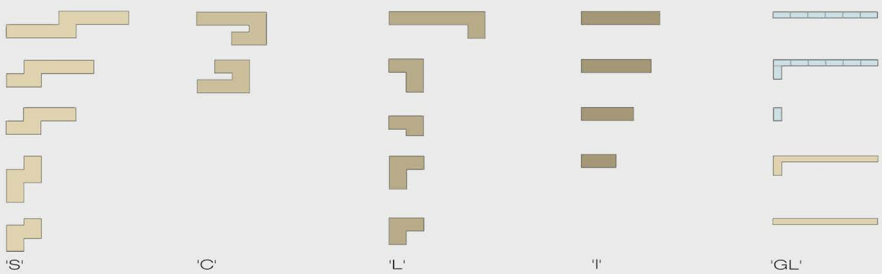
LAYERS



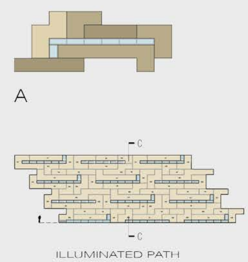


R O P E - C O N C E P T

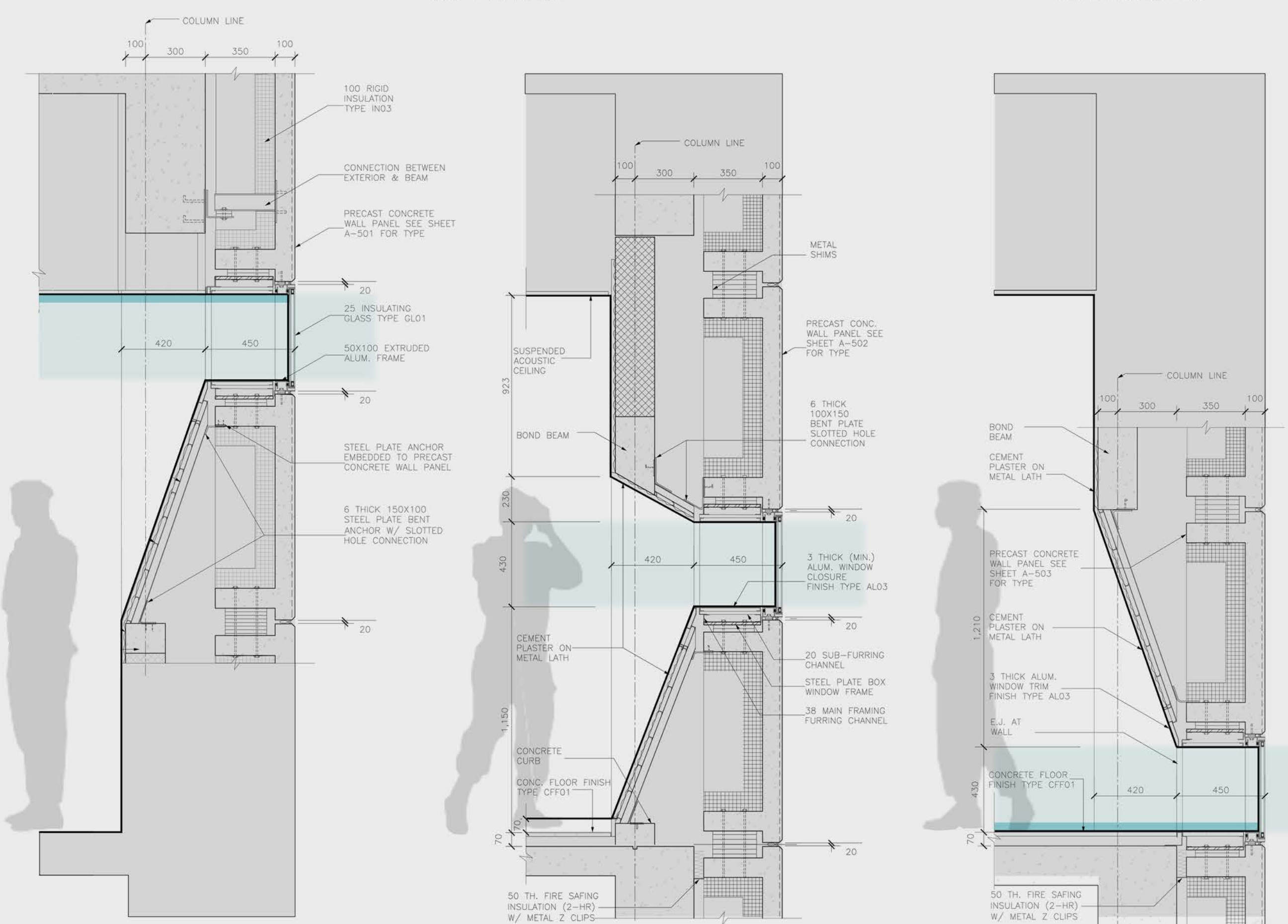
My contribution to the projects was to introduce the notion of the “ROPE” building that became major conceptual framework for the entire campus planning. 500M X 500M square site has almost no contextual element surrounding it, and the development of coherent system to design the whole campus was necessary. Entire campus planning initiated by my tiny conceptual model. Idea was to create EXPANDABLE and ADAPTABLE GENERIC STRUCTURAL PROTOTYPE that can simultaneously provides architectural program and circulation spaces, three-dimensionally throughout the site. The project in arid and hot desert climate zone required net-works of air-conditioned circulation spaces throughout the campus for the high rank officers and shaded walkways for the students marching the exterior campus, and this prototype perfectly provides the solution for the climate and the program requirement.   My idea was also interpreted as a traditional Islamic geometric motif by the principal designer and used as a form-making strategy for the project.



ROPE SHAPES



ROPE MODULES



ROPE DETAILS

MoMA Design and Book Stores
New York, NY 2004 (closed)

The retail spaces for the new Museum of Modern Art consist of two main areas: the 6,400-square-foot Design and Book Store on 53rd Street and the 1,600-square-foot bookstore on the second floor of the museum. In addition, an open retail area was created on the sixth floor adjacent to the temporary exhibition galleries. Working closely with the museum staff and retail specialists, Gluckman Mayner Architects designed each of the separate stores with a distinctive palette of materials while demonstrating extraordinary attention to the functional requirements of storage and display.

Office: Gluckman Mayner Architects
My Role: One of the project architects in the design team for the interior design of a total 5,700 ft^2 store area and display fixtures from the schematic design to construction document phase.

SIGGRAPH 2022 VANCOUVER+ 7-11 AUG

Narahara, T., “Presenting Architectural Research in VR,” ACM SIGGRAPH 2022 Educator's Forum, 2022 August. https://doi.org/10.1145/3532724.3535602 [Video].

Presenting Architectural Research in VR

Taro Narahara
narahara@njit.edu
New Jersey Institute of Technology
Newark, New Jersey, USA

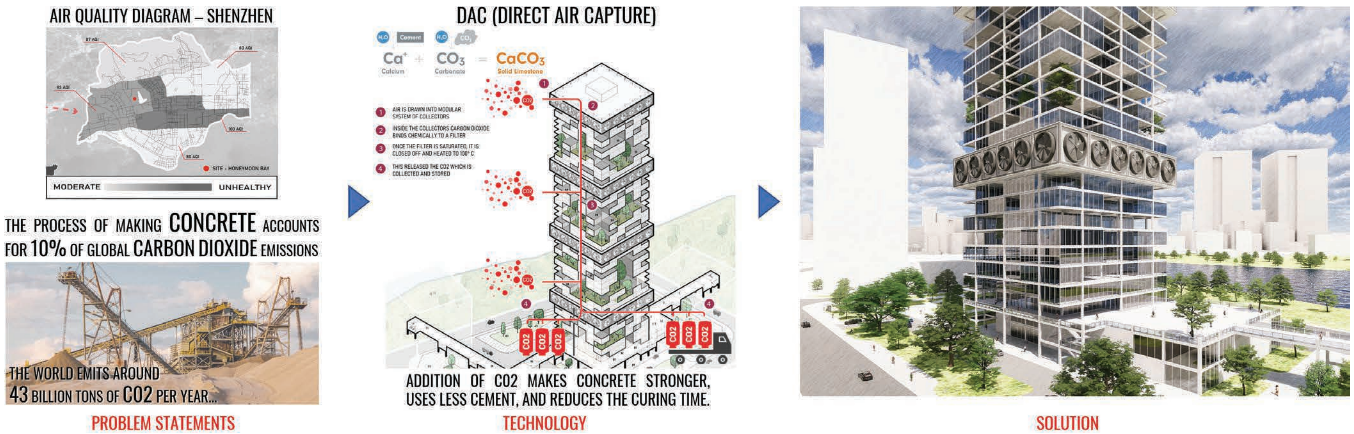


Figure 1: Carbon Tower by Janine Lojko & Mary Riccio. Incorporating direct air capture (DAC) units in the multipurpose skyscraper, as well as including a concrete and carbon facility, this design can clean the air of carbon dioxide while distributing carbon-cured concrete and stored carbon to the surrounding construction and businesses. A VR model for presentation (right).

ABSTRACT

In recent years, new standards for sustainable development and innovations in technologies require us to rethink the forms and functions of architecture that meet emerging needs for future cities. Anticipated developments in digital communication, autonomous vehicles, smart cities, and new energy harvesting technologies are expected to change our lifestyles, leading us to reconsider how we design and organize spatial conditions in our built environment. Such changes will challenge the roles of designers and how they represent and promote their design solutions using new media, including virtual reality (VR) and digital design technologies.

In this submission, I present a design assignment that challenges students to create speculative building designs by applying anticipated technologies that can make our built environment more sustainable. The process in this assignment can be applied to other design fields, including product design and urban design. Many design disciplines can benefit by exploring educational approaches that help integrate new emerging technologies for sustainable product developments and incorporating new digital means for representations for such visions.

SIGGRAPH '22 Educator's Forum , August 07-11, 2022, Vancouver, BC, Canada

ACM ISBN 978-1-4503-9366-9/22/08.
https://doi.org/10.1145/3532724.3535602

CCS CONCEPTS

• **Applied computing → Architecture (buildings)**; *Education.*

KEYWORDS

architecture, digital design, VR, technological convergence

ACM Reference Format:
Taro Narahara. 2022. Presenting Architectural Research in VR. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Educator's Forum (SIGGRAPH '22 Educator's Forum), August 07-11, 2022.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3532724.3535602

1 INTRODUCTION

In this submission for the Engaging Education Techniques and Assignments session, I introduce a design assignment that challenges students to create speculative building designs by applying anticipated technologies to make our built environment more sustainable. The assignment is project-based and requires creative thinking and multidisciplinary problem solving for speculative scenarios through the interplay of technology and design thinking. The project can be assigned individually or to student groups. The steps for the process of this assignment are summarized as follows. 1). Students are provided with existing site locations, programs for buildings, and any constraints for construction and regulatory codes. 2). Students find existing environmental problems for the proposed site for their building design. 3). Students research both existing and anticipated technology that can resolve the problems. 4). Students integrate the new technology into their architectural solutions creatively. 5). Students explore virtual reality environments to represent their visions to guest experts for assessment and discussion.

SIGGRAPH 2022 VANCOUVER+ 7-11 AUG
Narahara, T., "Presenting Architectural Research in VR," ACM SIGGRAPH 2022 Educator's Forum, 2022 August. https://doi.org/10.1145/3532724.3535602 [Video].

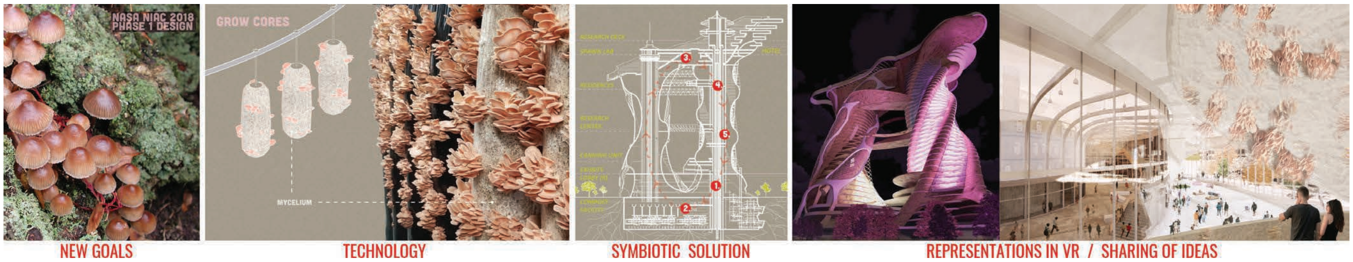


Figure 2: Mycelium Factory by Reyne Bennett (student). Studies by NASA shows applications of fungal mycelium to building constructions using its excellent mechanical properties. The project serves as a metabolic tower designed to vertically cultivate mushrooms for food and research, while facilitating a symbiotic relationship between fungi, bacteria, plants, and humans.

Table 1: Metadata in tabular format

Summary	Students target problems in future living, e.g., pollution, and create design solutions by applying anticipated technologies for sustainability.
Learning outcomes	Students learn and practice creative problem solving discover non-traditional design generation and presentation methods.
Classification	Architecture, VR, technological convergence.
Audience	Undergraduate students at all levels. Goals can be adjusted based on experience and skills.
Dependencies	Varied based on the selected project.
Prerequisites	None. Non-sequential assignment.
Strengths	Scalable in scope and complexity to some other design disciplines. Students learn a non-traditional presentation format including VR. Opportunity to present to experts. Can be assigned individually or to student groups.
Weaknesses	Require students to develop their own goals and requirements in contrast to conventional design exercises based only on goals prepared by instructors, which could be challenging.
Variants	Unlimited design options and approaches.
Assessment	Rubric can be designed around quality and effectiveness metrics. Open discussions with invited experts from industry and academia on presentations using creative communication media.

2 METHODOLOGY

The assignment is designed to motivate creative thinking by challenging students to work within the requirements from the selected site location, climate conditions, purpose, and usage types of buildings, including required areas. During the introduction stage, students study the local conditions, including its culture, climate, industry, and demographics, and target specific existing problem areas for the current environment, such as air quality, the life cycle of building materials currently used, and increasing density. Then, they will seek improvement for future living conditions for the site.

In the next stage of research, students search for technology and innovative strategies to remedy the problems found from the current site conditions in the previous step. Students learn from research papers and articles that explain scientific mechanisms to improve the underlining problems in the cities and existing application examples of the technology from scientific databases (ACM, IEEE) using search engines (Google Scholar, etc.).

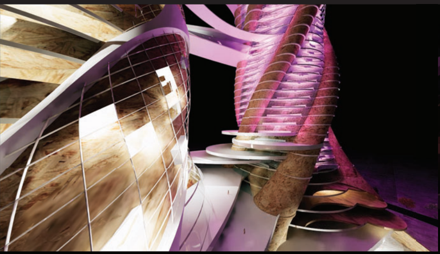
After the research stage, students find creative solutions to incorporate the selected technology into their architectural designs. Starting with diagrammatic floor plans representing adjacency among different rooms for commonly used building types, such as office or residential towers, students fit the proposed technology and their architectural strategies into the preliminary layouts incorporating emerging new programs associated with the technology. For example, some groups introduced the visitor center and demonstration facilities for the adopted new technology.

After students' designs get matured, they find the best way to represent their architectural visions to hypothetical stakeholders, clients, investors, and collaborators such as engineers who could technically support their projects. Tutorials on presentations using immersive VR in a computer game development environment (Twinmotion, Unity3D, Unreal, etc.) were introduced by the instructor, which helped students develop their own creative approaches to articulate each project's strength (see Table 1 for details).

Upon completion, media can be reviewed by invited experts from industry and academia in class. Reviewers rated students' VR, drawings, and slide presentations based on concept, aesthetics, and technical accomplishments. Contents in immersive VR are used not only for representations but also for evaluations of spaces yet to be built inside the simulated environments through the lens of virtual. Comments from reviewers promote discussions as to how students should further proceed with their explorations.

3 CONCLUSION

To become the next leader in the field of design, students need to have the habit of developing one's creative solutions to meet the new standards. Simply following templates and receipts from existing design solutions might no longer be sufficient to provide innovative solutions that can respond to ever-changing dynamic demands from society under the era of technological convergence. This project aims to prepare students for such future scenarios as they venture into the professional world.

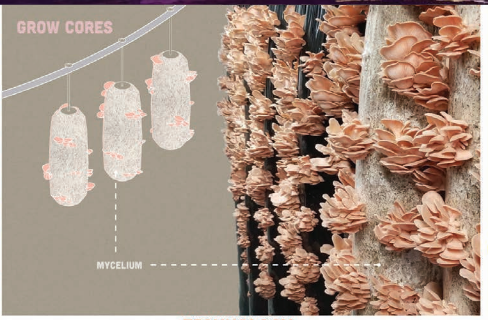


MYCO TOWER

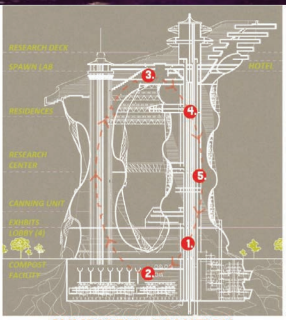
Mycelium Factory by Reyne Bennett (student). Studies by NASA shows applications of fungal mycelium to building constructions using its excellent mechanical properties. The project serves as a metabolic tower designed to vertically cultivate mushrooms for food and research, while facilitating a symbiotic relationship between fungi, bacteria, plants, and humans.

Reyne Bennett

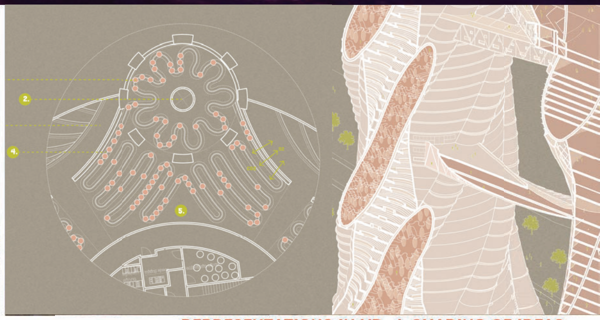
NEW GOALS



TECHNOLOGY

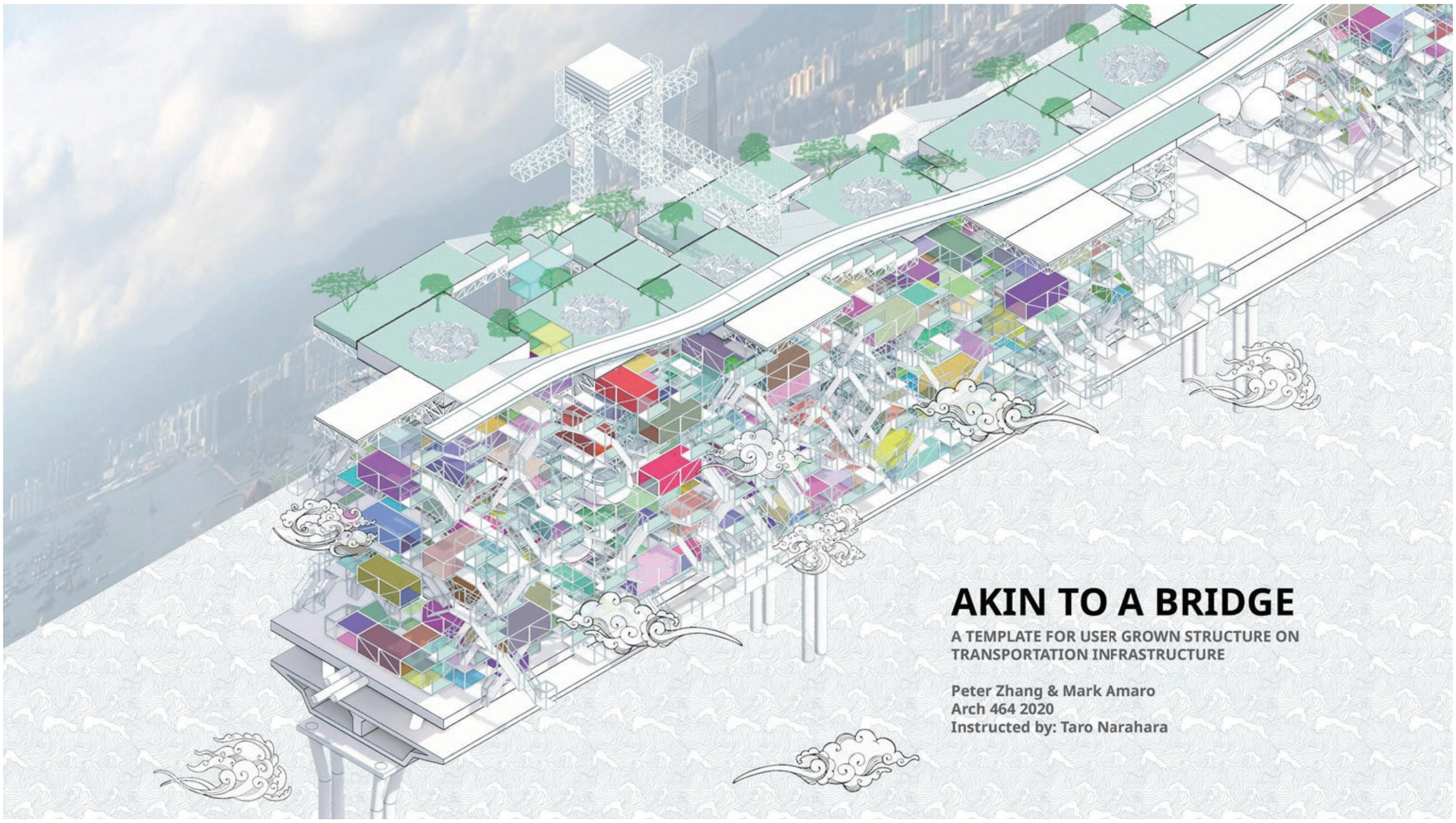


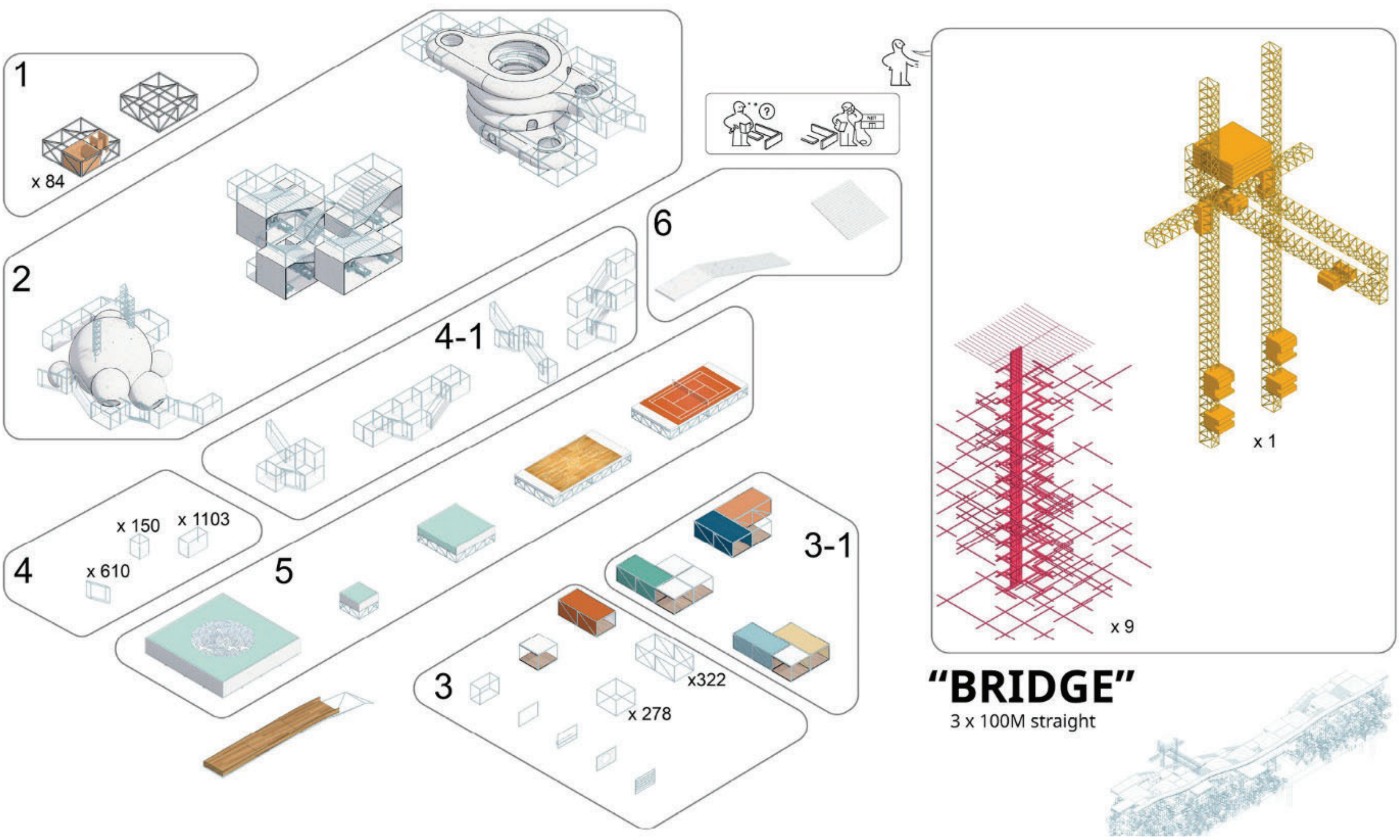
SYMBIOTIC SOLUTION



REPRESENTATIONS IN VR / SHARING OF IDEAS

Student Projects:

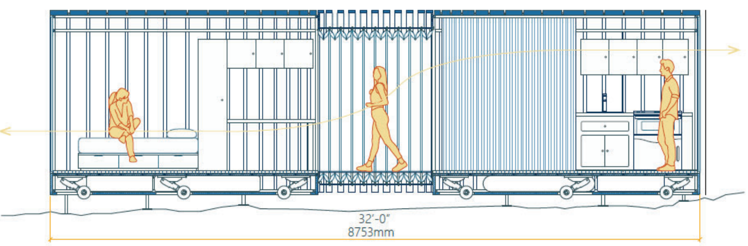


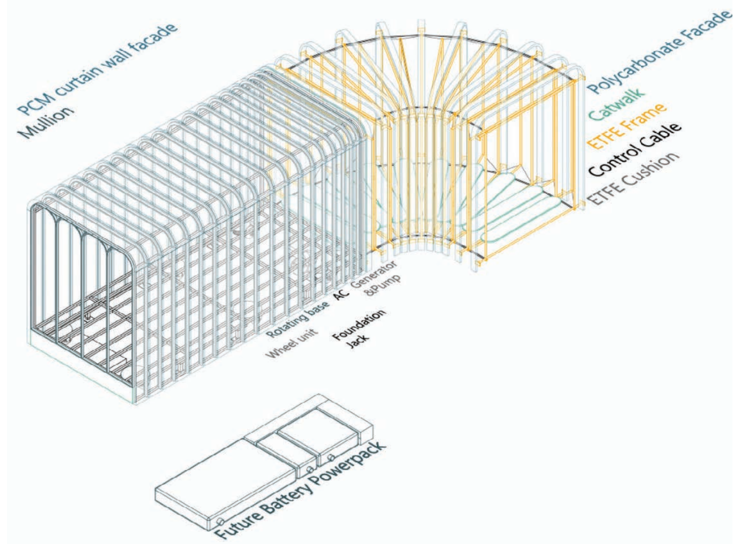


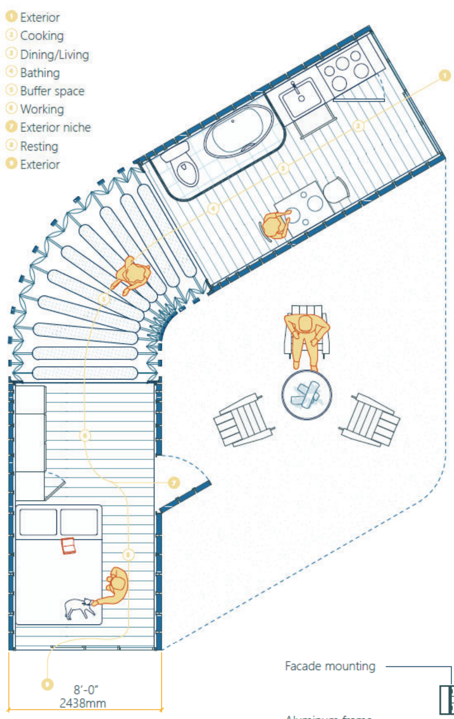
Using the density and specific relationship to the transportation infrastructure of Hong Kong as a context, this project explores the possibility of integrating the flexibility of pedestrian footbridges on a macro scale that connects city districts. The system allows for the generation of modular structures containing multiple attributes acting as BIM that can integrate information for mechanical and environmental analysis and can be oriented to pursue different design goals, such as a general orientation of residential units towards the valuable view or the connection of smaller modules in forming community space. Like the living city, the bridge grows to adapt to its user and the surrounding.

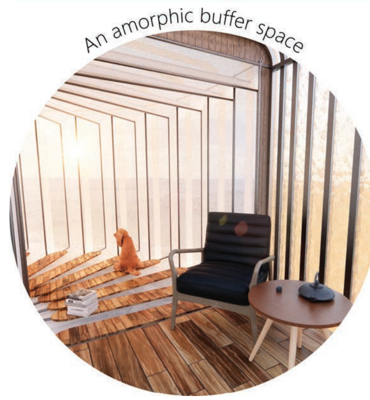
Peter Zhang and Mark Amaro

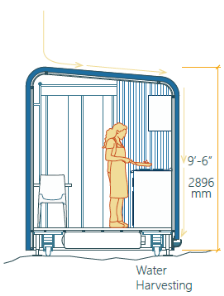
Student Projects:

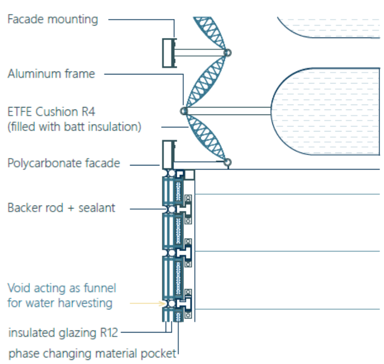


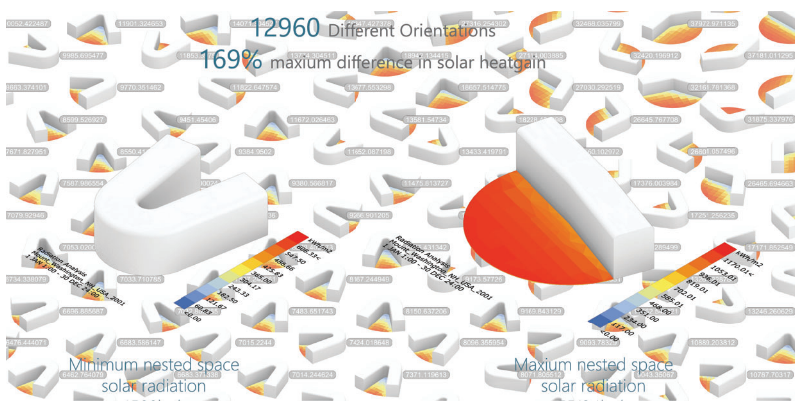






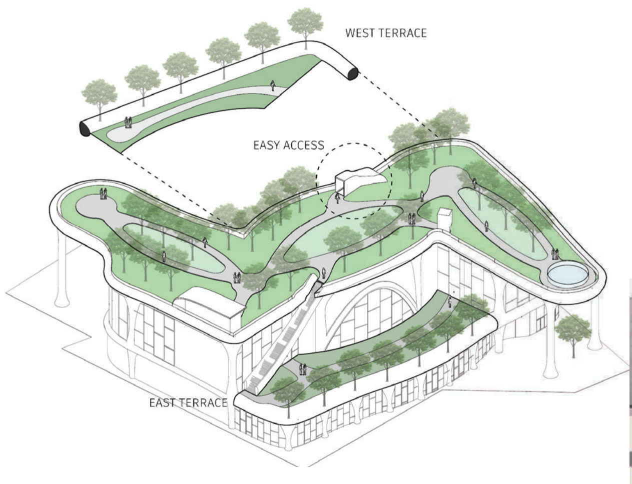






This project aims to provide people the chance to explore new destinations with enhanced control and comfort. "Any place" is a flexible modular system that consists of two customizable units and a collapsible joint space. The joint space provides a subtle separation between spaces and the ability to fold, orienting each unit at will to create nested outdoor spaces and to optimize solar gain. The facade of "any place" works as a comfortable greenhouse thanks to the usage of phase changing material (PCM) curtain, "Any place" allows occupants to discover new destinations and engage intimately with its environment.

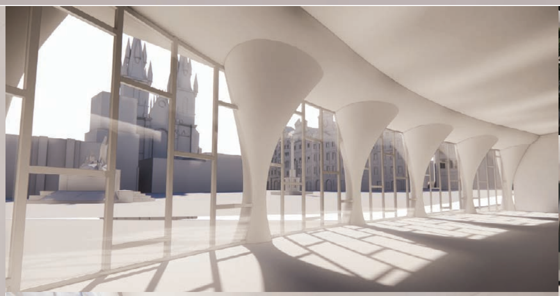
Student Projects:

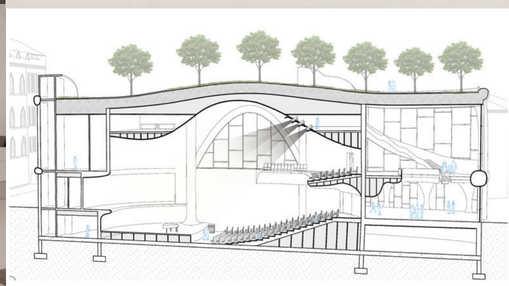


FLOW

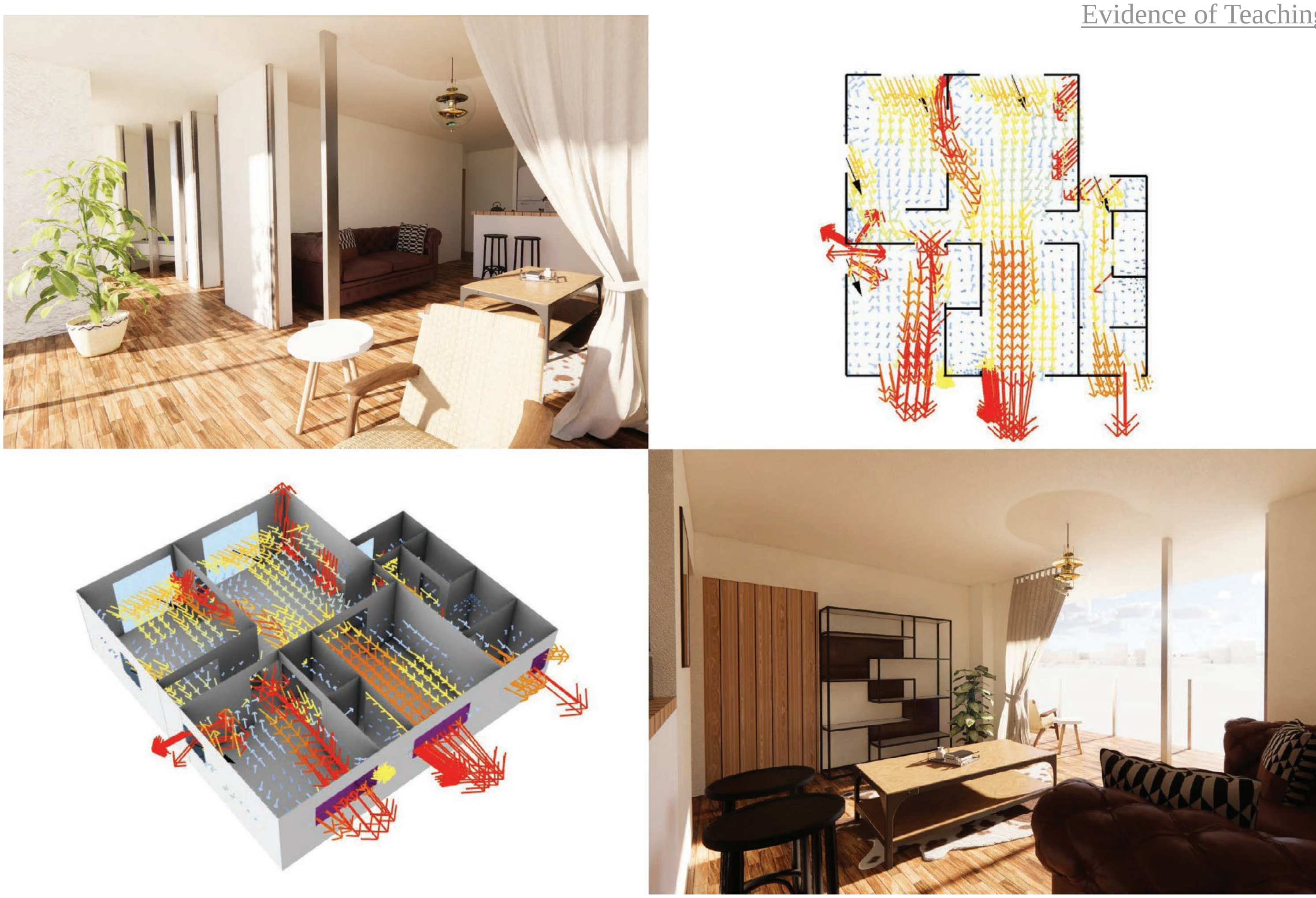
Anthony Volpe | Taro Narahara | 2022.

Located in the historic site in Prague, this concert hall is a part of the urban fabric that gives visitors to the old town square access to a park that is integrated into the building, allowing visitors to interact with the historic area in novel ways. Restoring greenery to cities is essential for lowering CO2 emissions from densely populated areas, protecting biodiversity, and improving sustainability. The project places a strong emphasis on the idea of "Flows": the flow between users and the town square, between people and nature, and between each program of the building.

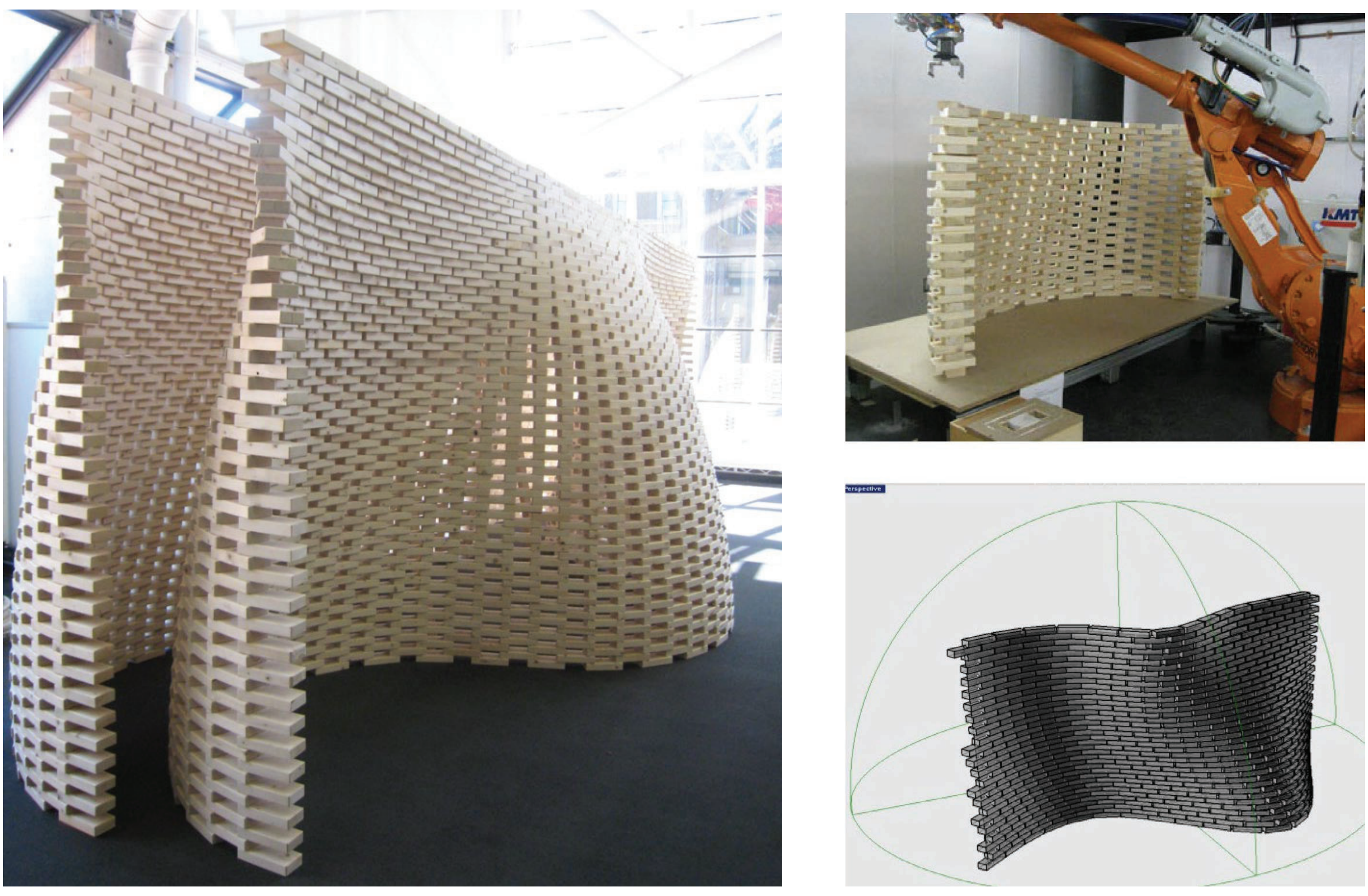




Anthony Volpe

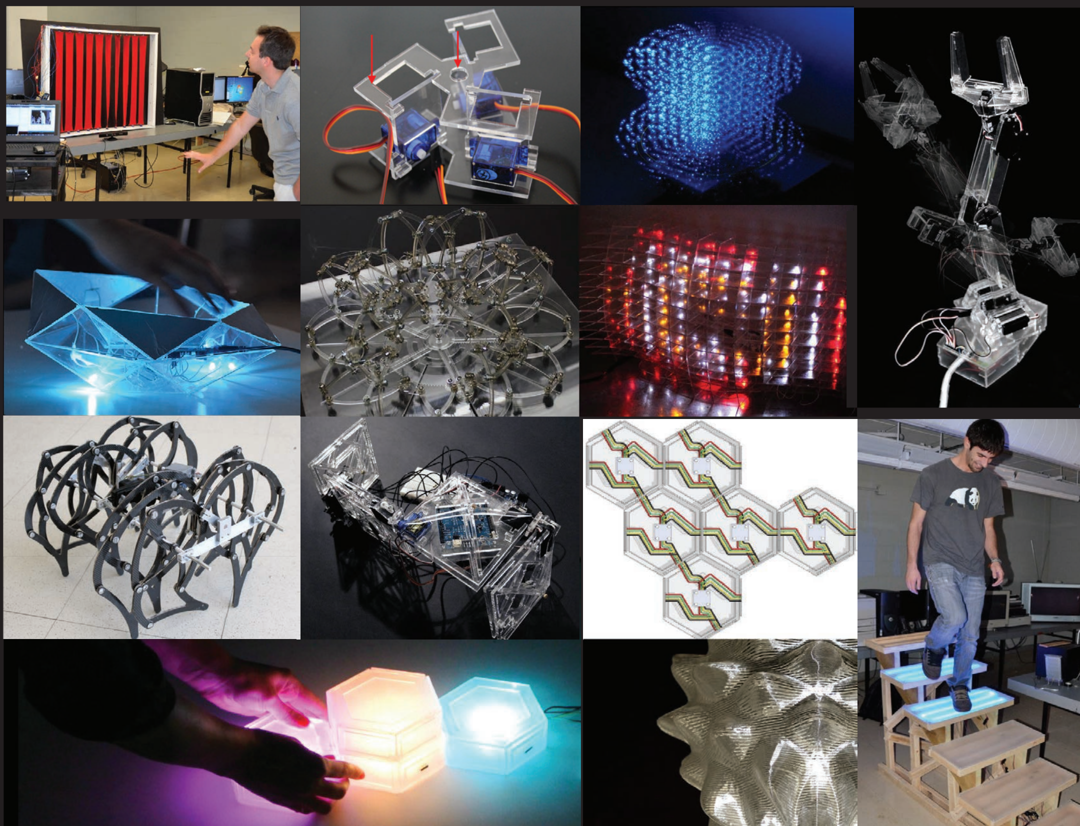


Taught environmental analysis and optimization, covering wind, energy use, CO2 emissions, and interactive VR environment creation.



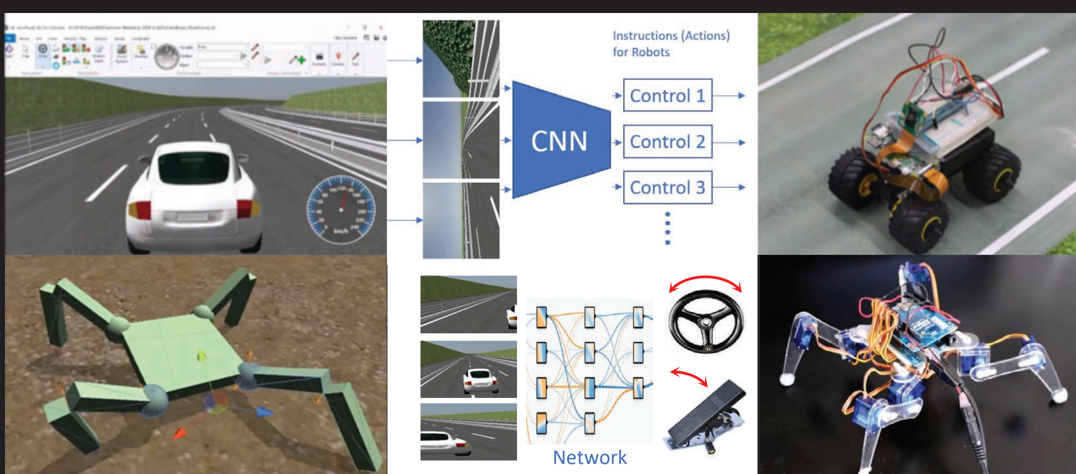
Robotic Digital Fabrication Group Project at Harvard GSD, 2010 – participated as RA/TA, providing code and assisting fellow students.

Teaching Interactivity 2010 - 2018:

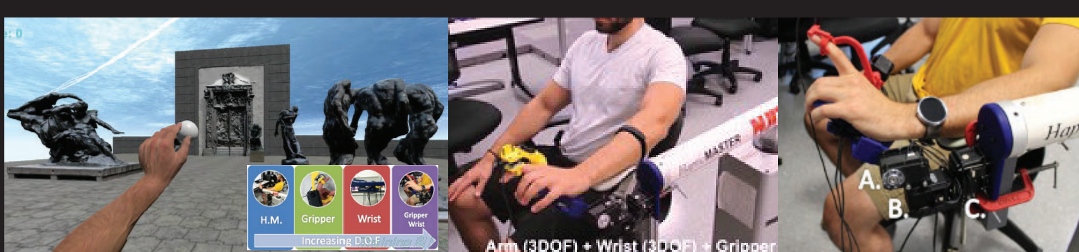


Education Papers:

Design Exploration through interactive prototypes using sensors and microcontrollers, Computers & Graphics: An International Journal of Systems & Applications in Computer Graphics, Elsevier Science & Technology, vol. 50, pp. 25-35, 2015.
[PDF , Video]
https://doi.org/10.1016/j.cag.2015.04.008



Personalizing homemade bots with plug & play AI for STEAM education
SIGGRAPH Asia 2018 Technical Brief
Tokyo, Japan, 2018.
[PDF, Video]
https://doi.org/10.1145/3283254.3283270

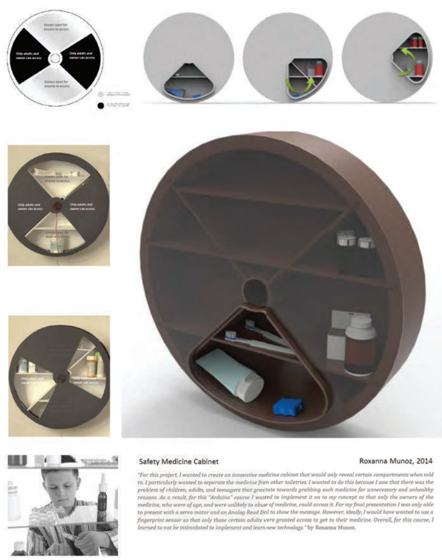


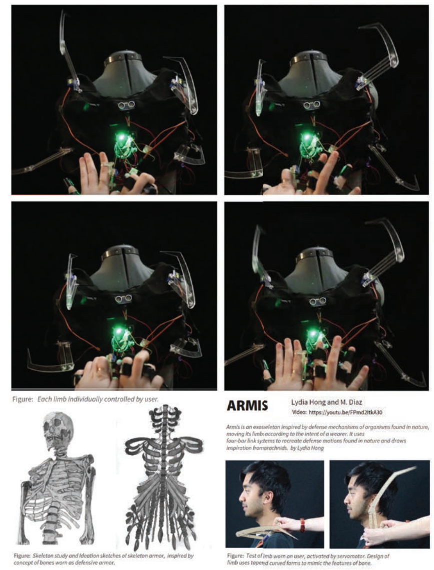
Haptic Collaboration: Biomedical Engineering Meets Digital Design,
SIGGRAPH 2015 Talks, Los Angeles CA, 2015.
[PDF, Video]
https://doi.org/10.1145/2785585.2792520

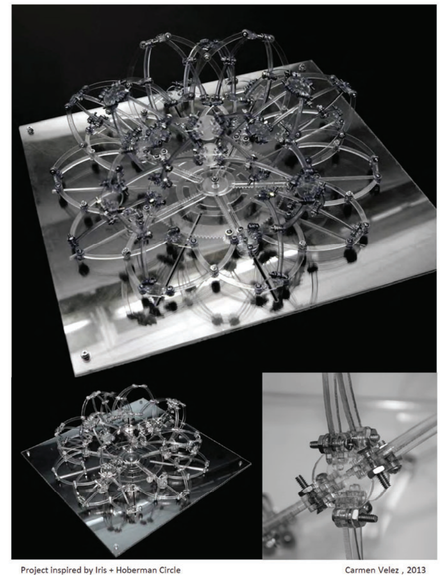
NSF: Major Research Instrumentation Program (MRI), Co-PI, “MRI-Development of an Open Architecture and Scalable Exoskeleton for Research on the Restoration of Ambulation of Persons with Disabilities” PI: Foulds, R., Co-PIs: Adamovich, S., Narahara ($225,500), 2016-18.

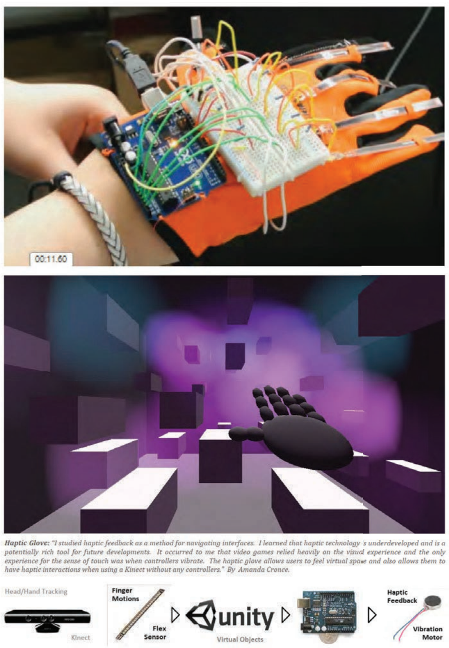
Student Projects: Smart Products for Design Students

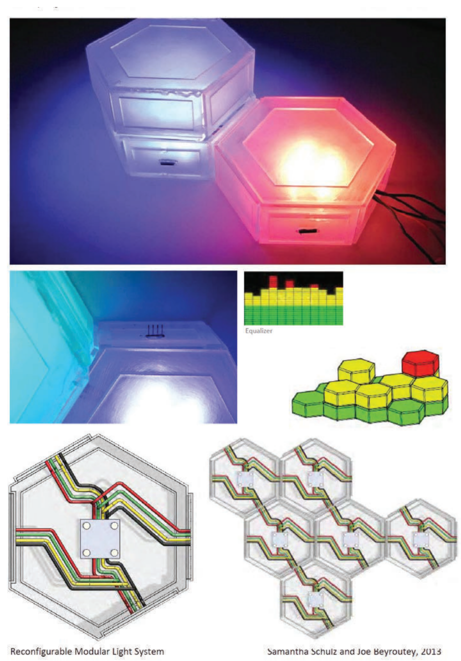
Sample Videos

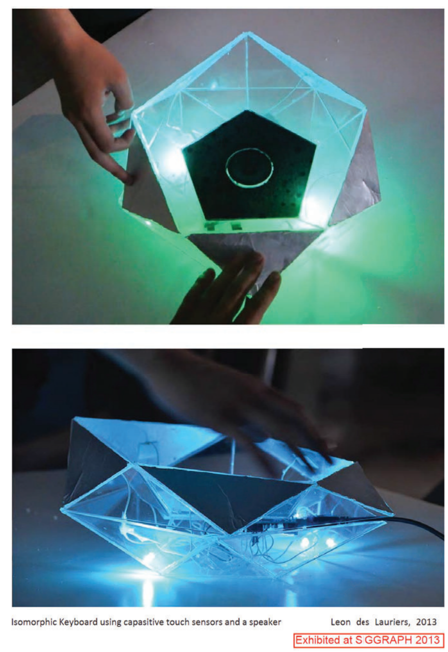


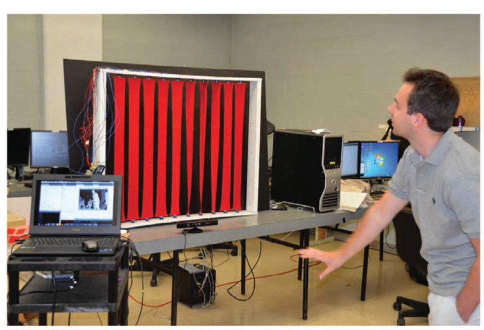












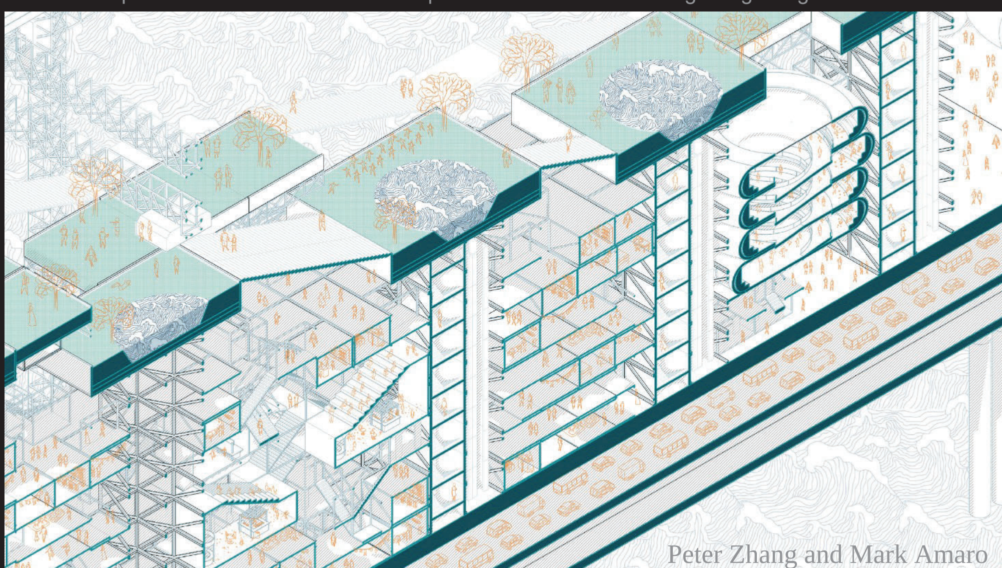
Over the years, I have taught elective courses at NJIT on smart products, adaptive designs, and intelligent spaces. In Computers & Graphics: An International Journal of Systems & Applications in Computer Graphics, Elsevier Science & Technology (2015) [**PDF, VIDEO**], I presented an educational case study and its pedagogical lessons from a project-based course for beginning design students to produce interactive prototypes using sensors, actuators, and microcontrollers. In the first half, a series of short project-based modules utilizing scaffolding of code templates in conjunction with toolkits for physical prototypes were introduced, followed by a more open-ended investigation of project-based individual creative final projects. Each module with instructions on prototyping and programming in pairs helped students to build and see abstract logic in programming through the physical behaviors of prototypes without overpowering student creativity and motivation. This strategy allowed students to acquire extensible knowledge that does not rely on higher-level software functions or specialized but inflexible plug-ins. This paper was an extended and revised version of a paper presented at the EUROGRAPHICS 2014 conference.

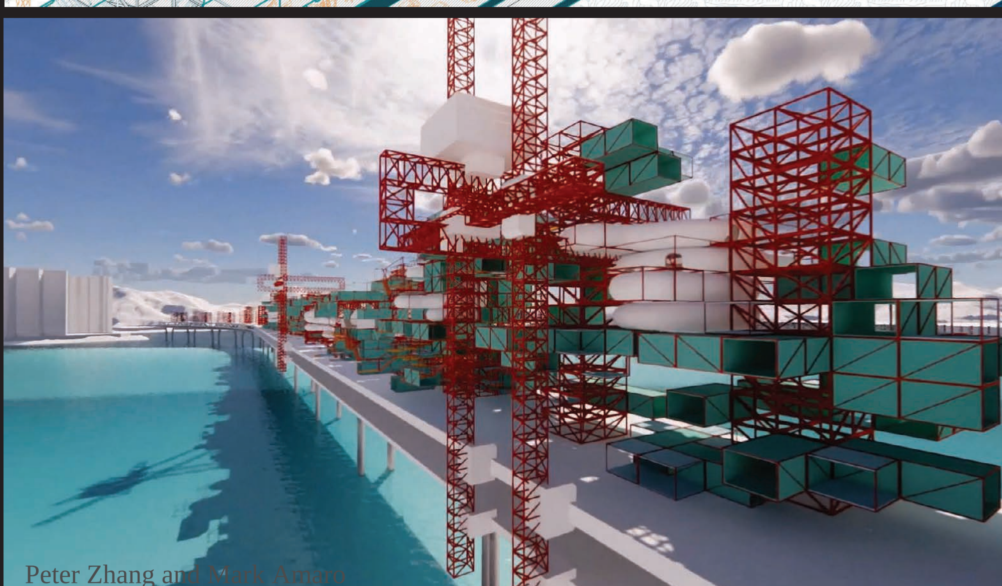
Examples of VR projects by students

Student Project using Game interaction technology

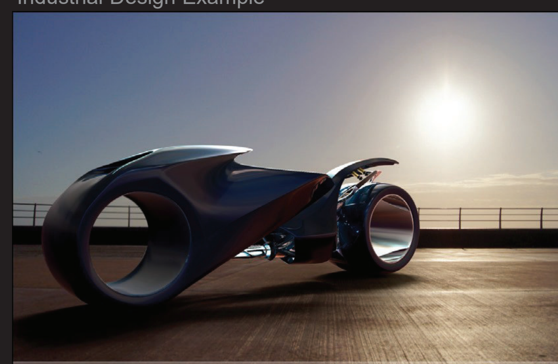
CG background and character creation

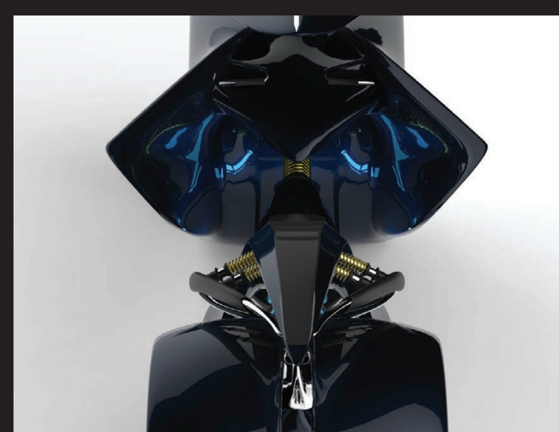
Student Proposal for Virtual Architecture: Expandable Structure on Hong Kong Bridge

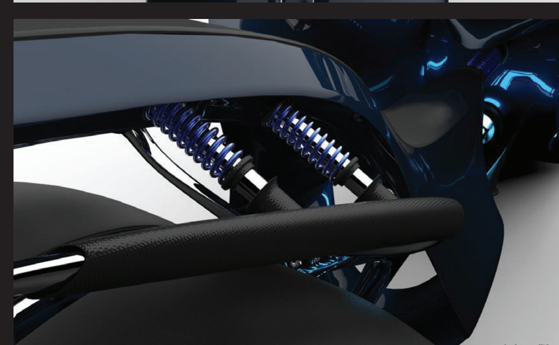




Industrial Design Example

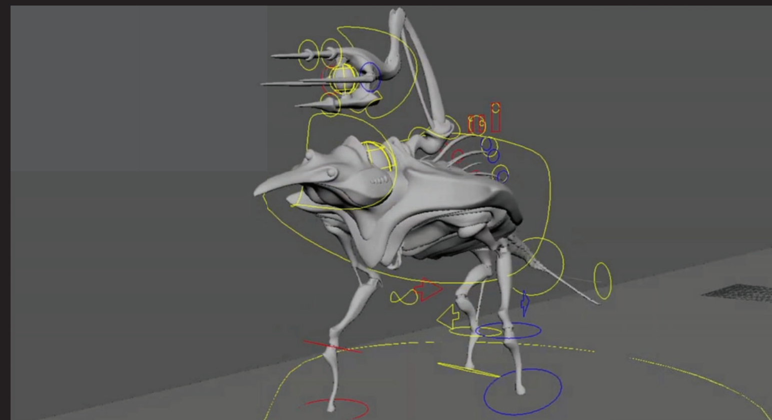






Student Projects: 3-D Animations / Game Developments

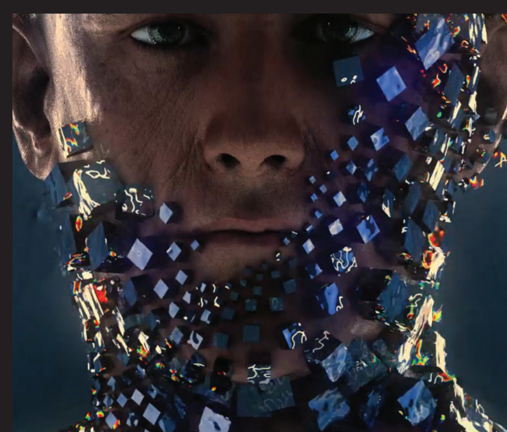
Sample Videos



Nicholas Seccondro, 2020

Trysten Davis, 2018

Brandon Kong and Dongsheng Han, 2021



Alexander Schuppel and Kathleen Nguyendon, 2018.

Ian Yunis, 2014 (Exhibited at SIGGRAPH 2014)

After the studio review with undergraduate digital design students

NOVEL IMAGE DESCRIPTORS AND LEARNING METHODS FOR IMAGE CLASSIFICATION APPLICATIONS
by Ajit Puthenputhussery (Ph.D. Computer Science: 2017-), Dr. Chengjun Liu (Dissertation Advisor), Dr.Taro Narahara (Committee Member).

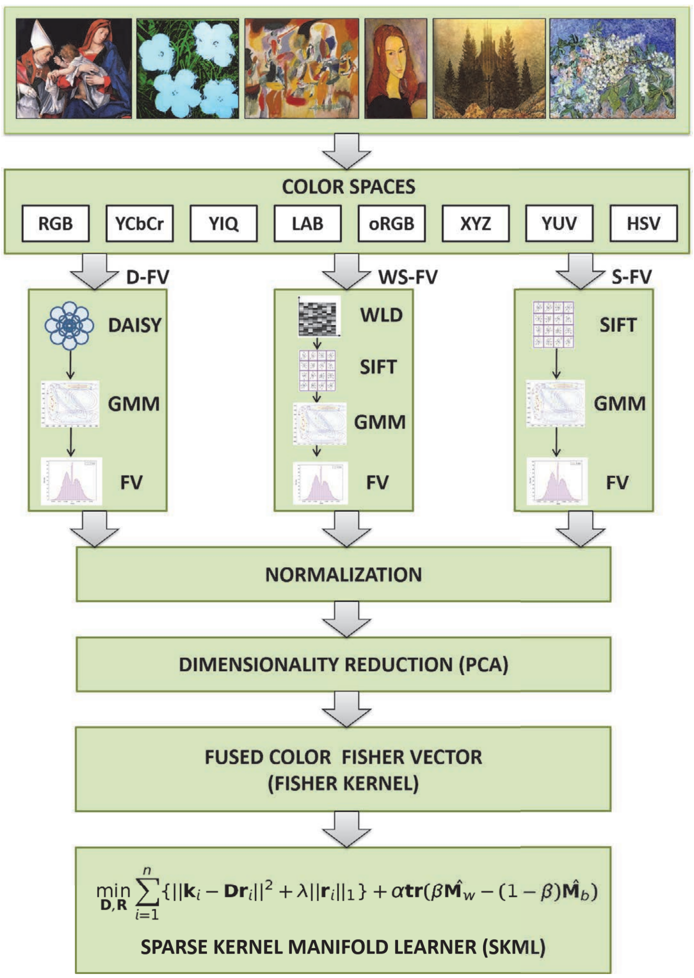
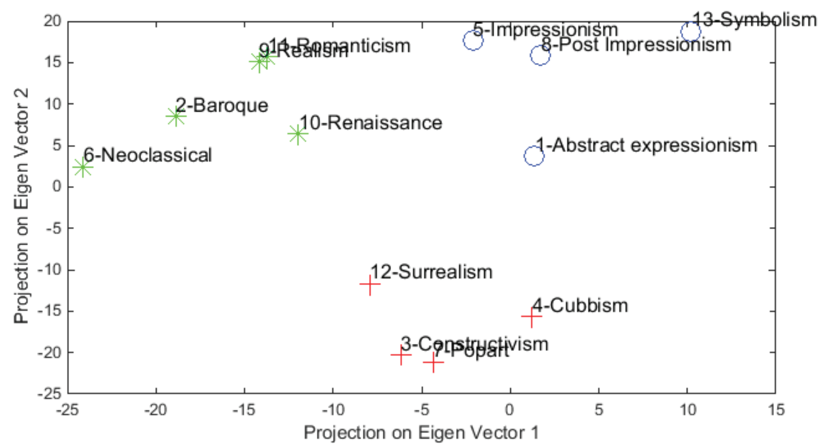


Figure. The framework of our proposed SKML method.



Due to its wide applications, image classification is a hot topic in computer vision and machine learning. Big data has increased the need for robust image descriptors and learning methods to process many images for various visual applications. This proposal explores new image descriptors and learning methods by incorporating important visual aspects and improving feature representation in the discriminative space for image classification.

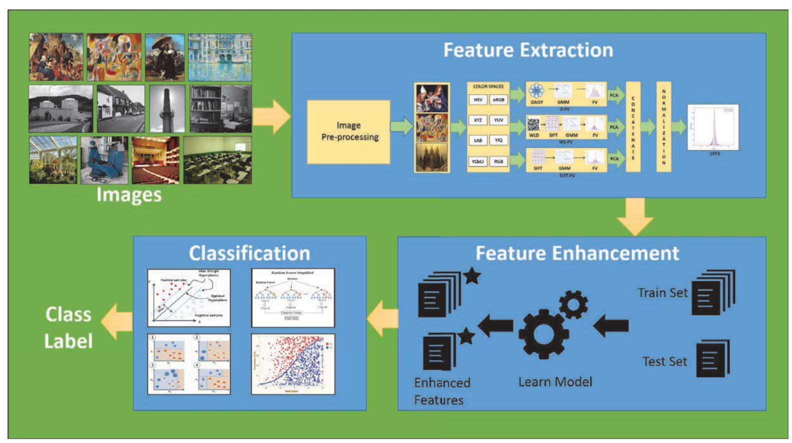
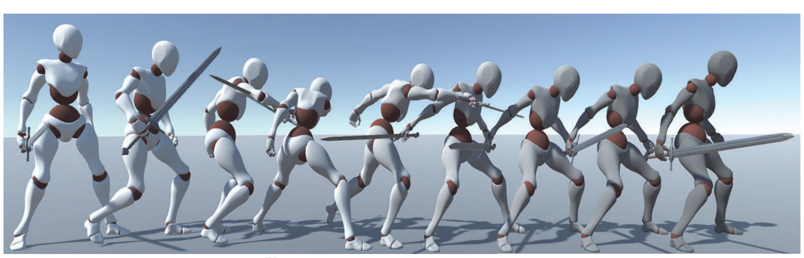


Figure. The style influence cluster graph for the Painting-91 dataset.

SLASH: DATA-DRIVEN PERCEPTUAL ANALYSIS OF SWORD FIGHTING MOTION by Yunhao Zhang (Ph.D. in Informatics, 2022-), Dr. Tomer Weiss (Dissertation Adviser), Dr. Taro Narahara (Committee member).

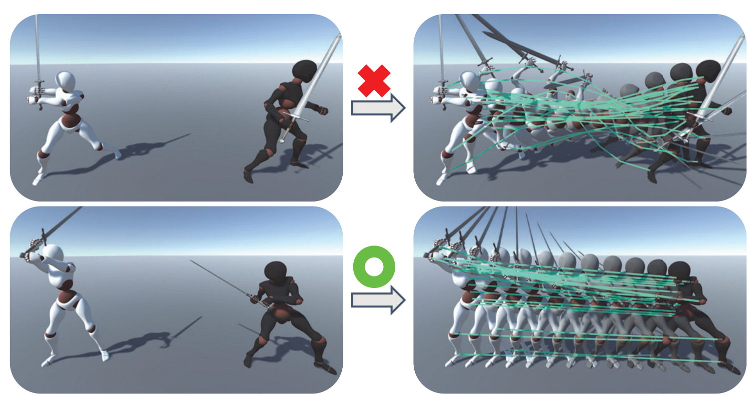


$$f(\mathbf{I}_{base}, \mathbf{I}_{compared}) \Rightarrow D_{jr}(f_1, f_2) = \sum_{i,j=1}^{K} \left(1 - (q_i \cdot q_j)^2\right) [i, j \notin I]$$

Distance of two frames(Joints Rotation)

- $D_{jr}(f_1, f_2)$ represents the distance function for joint rotation
- f_1 and f_2 are compared frames
- q represents the quaternion of joints $q \in R^4$
- i, j are the indice of joints in f_1, f_2 respectively
- I is a collection which contains the joints should be ignored (fingers etc)
- K is total number of joints

Motion Retrieval - Joints Rotation Distance



Application Pipeline - Transition Construction (right)

Realistic character animations are still mostly manually produced. Fighting animations, which are popular in animation venues, are difficult to create due to their high dynamic and detail content. We present an annotated sword combat motion dataset. Our dataset includes 15,927 data points from 203 motions depicting various fighting styles, including pose coordinates, text annotations, video, and other modalities. We infer novel perceptual motion attributes using crowdsourcing, allowing users to explore datasets using artistic features that are hard to define. A suggestion engine, sequence creation, and user-empowered exploration for content analysis and creation are among our interactive media applications.

Ph.D. Dissertations: I served as a dissertation committee member to advise several graduate students from CS & Engineering

Assessment of a Hand Exoskeleton on Proximal and Distal Training in Virtual Environments for Robot Mediated Upper Extremity Rehabilitation by Kevin Abbruzzese (Ph.D. Biomedical Engineering: 2016)
Dr. Richard Foulds (Dissertation Advisor), Dr.Taro Narahara (Committee Member).

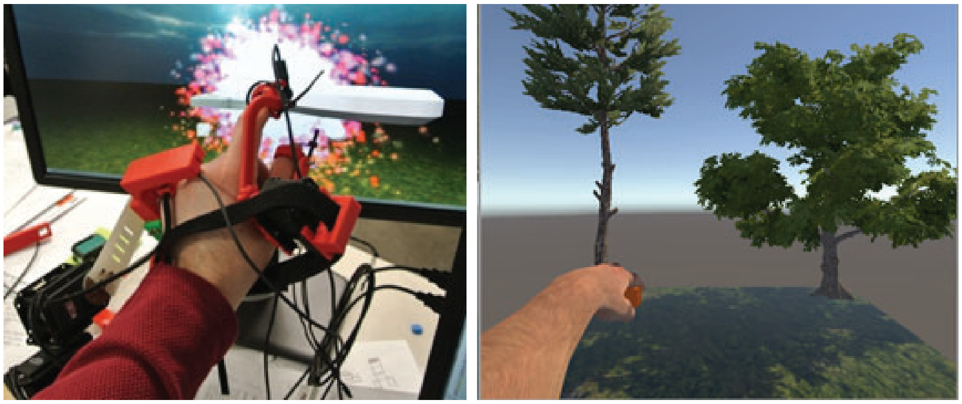


Figure. Left: Wrist exoskeleton with gripper interfaced with VE. Right: VE simulation in Unity of pinch task

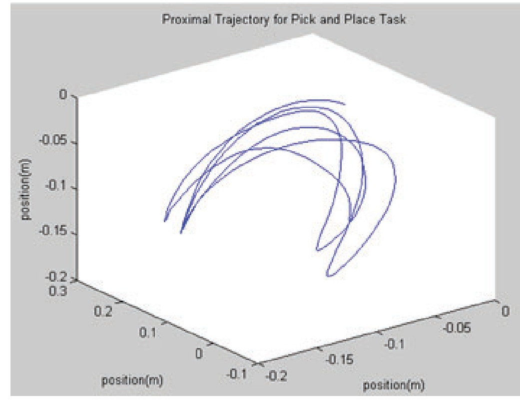


Figure. Proximal arm trajectory evaluated with the Haptic Master

Figure. Experimental groups for Pick and Place Task in order of increasing DOF. All groups use Haptic Master(HM) with the respective DOF.

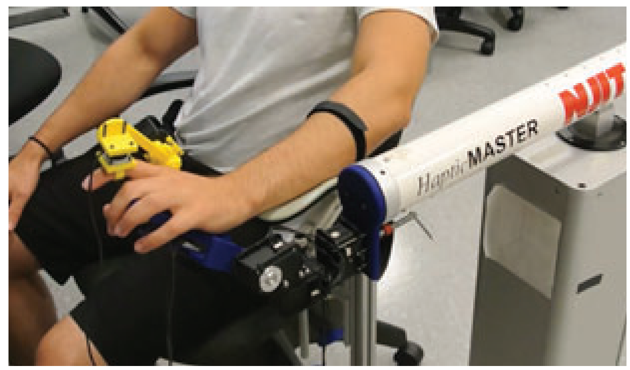


Figure. NJIT HandsOn System complete with 3 proximal DOF at the shoulder, 3 DOF at the wrist, and 1 DOF for pinch.

Stroke causes 800,000 US disabilities each year (Center of Disease Control 2015). Robotic rehabilitation may improve training and motor plasticity in stroke patients. To test the efficacy of increased DOF during proximal and distal training in Upper Extremity (UE) rehabilitation, a motorized hand exoskeleton that supports wrist flexion/extension, abduction/adduction, pronation/supination, and finger pinch will be integrated with a 3-DOF haptic robot (Haptic Master, FCS Moog). To assess the significance of increased DOF during virtual training, subjects were randomly divided into four groups: Haptic Master control group (HM), Haptic Master with Gripper (HMG), Haptic Master with Wrist (HMW), and Haptic Master with Gripper and Wrist (HMGW). Each group was given 120 seconds to transport as many cubes to a target in a virtual Pick and Place Task. Since assisted daily living tasks require coordinated arm and hand movement, those with both proximal and complete distal hand control (HMGW) may benefit more from the Pick and Place Task. These features form a comprehensive system to retrain lost function and improve motor control in motor-impaired populations.

INVESTIGATION OF NEW LEARNING METHODS FOR VISUAL RECOGNITION
by Qingfeng Liu (Ph.D. Computer Science: - 2017), Dr. Chengjun Liu (Dissertation Advisor), Dr.Taro Narahara (Committee Member).

Visual recognition is one of the most difficult and prevailing problems in computer vision and pattern recognition due to the challenges in understanding the semantics and contents of digital images. Two major components of a visual recognition system are discriminatory feature representation and efficient and accurate pattern classification. This dissertation therefore focuses on developing new learning methods for visual recognition.

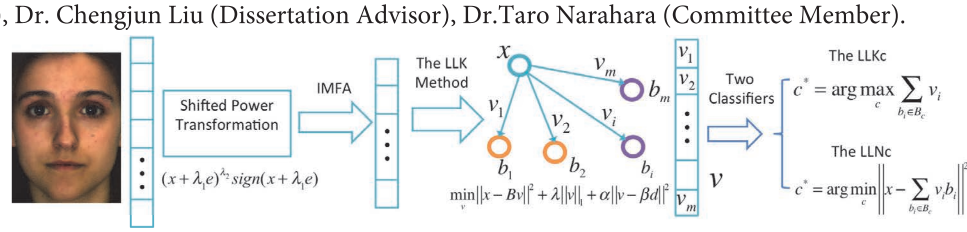


Figure. The pipeline of the proposed LLK method. The input pattern vector is first processed by the shifted power transformation. The IMFA method then extracts discriminative features. The proposed LLK method further derives a new representation v. The LLKc and the LLNc are finally applied for robust visual recognition.

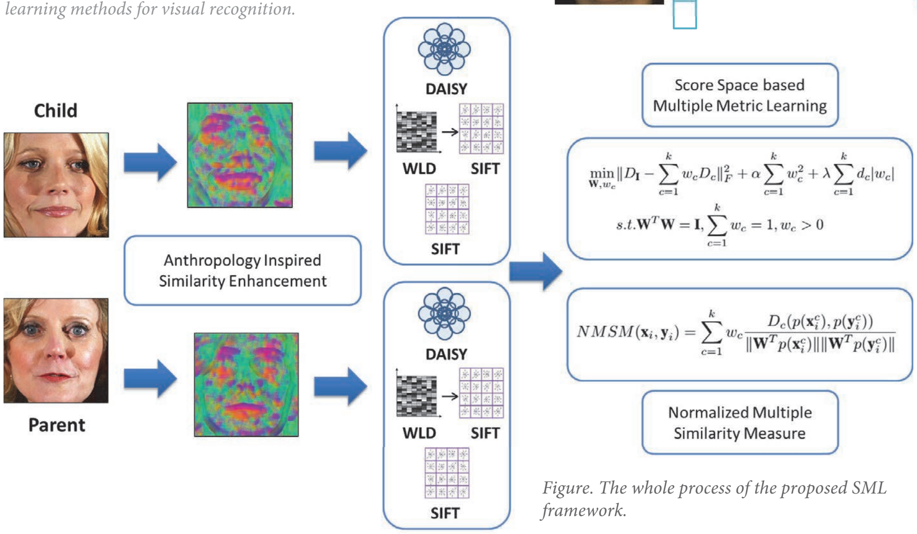


Figure. The whole process of the proposed SML framework.

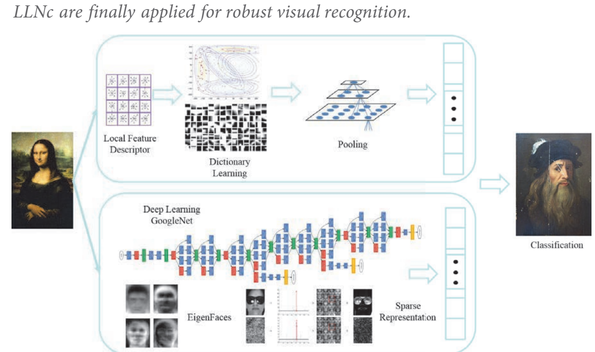
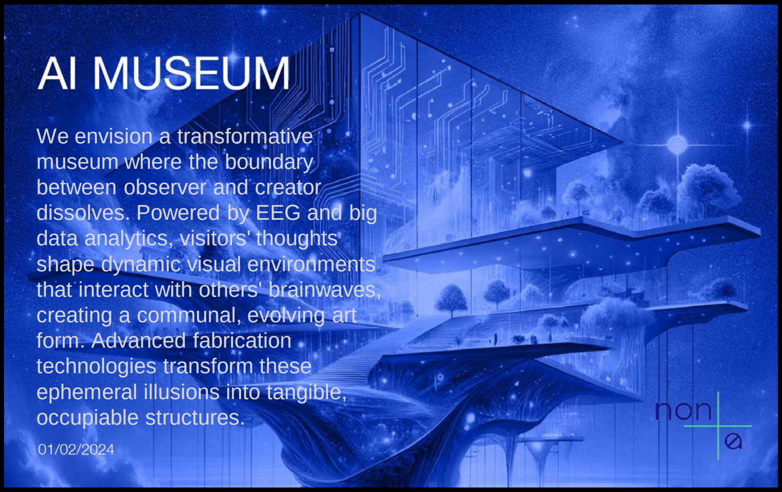


Figure. Two visual recognition frameworks (the k-means and the deep convolutional neural networks)

Honorable Mention, **AI Museum**, International Design Competition hosted by NON-ARCHITECTURE SRL (2024)
Collaboration with Oyke Alcin (M.S. in Arch. student at NJIT).

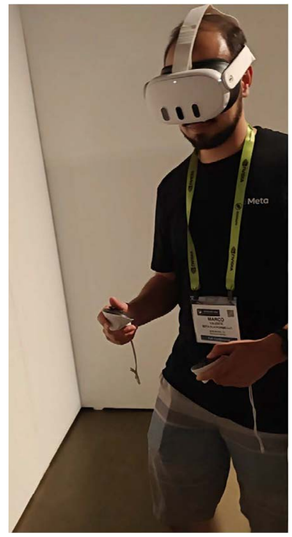


SIGGRAPH '23 Student Work Exhibit

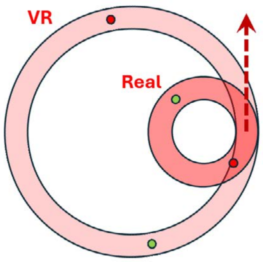
My courses incorporate VR and AI-driven generative animation techniques to advance design development.

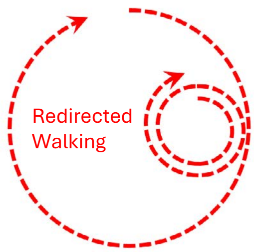
Click for video example

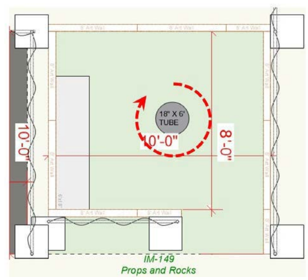
Vincenty, M., Grebler, J., Piza, C., Dalgo, O., & Narahara, T. (2024). Props and Rocks: Passive Haptic Mixed Reality for Navigating Far-off Worlds. In *ACM* ***SIGGRAPH 2024 Immersive Pavilion*** [Link].

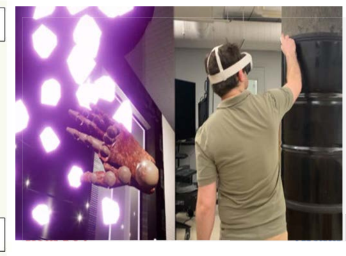


Our project immerses participants in a world blending virtual and physical realities. Users navigate a divided society, solving spatial puzzles and interacting with props via haptic feedback. Employing innovative techniques, including **stencil buffer** and **redirected walking**, our project seamlessly transitions between virtual worlds, advancing VR integration and offering educational opportunities









AI Tools to Synthesize Characteristics of Public Spaces: A case study for Jersey City. Riether, G., & Narahara, T. (2023). *Proceedings of the 41st Association for Education and Research in Computer-Aided Architectural Design in Europe (**eCAADe**) Conference*. [PDF]. (My role: Develop and share the concept with the co-author, produced all generative AI images, and wrote half the paper.)

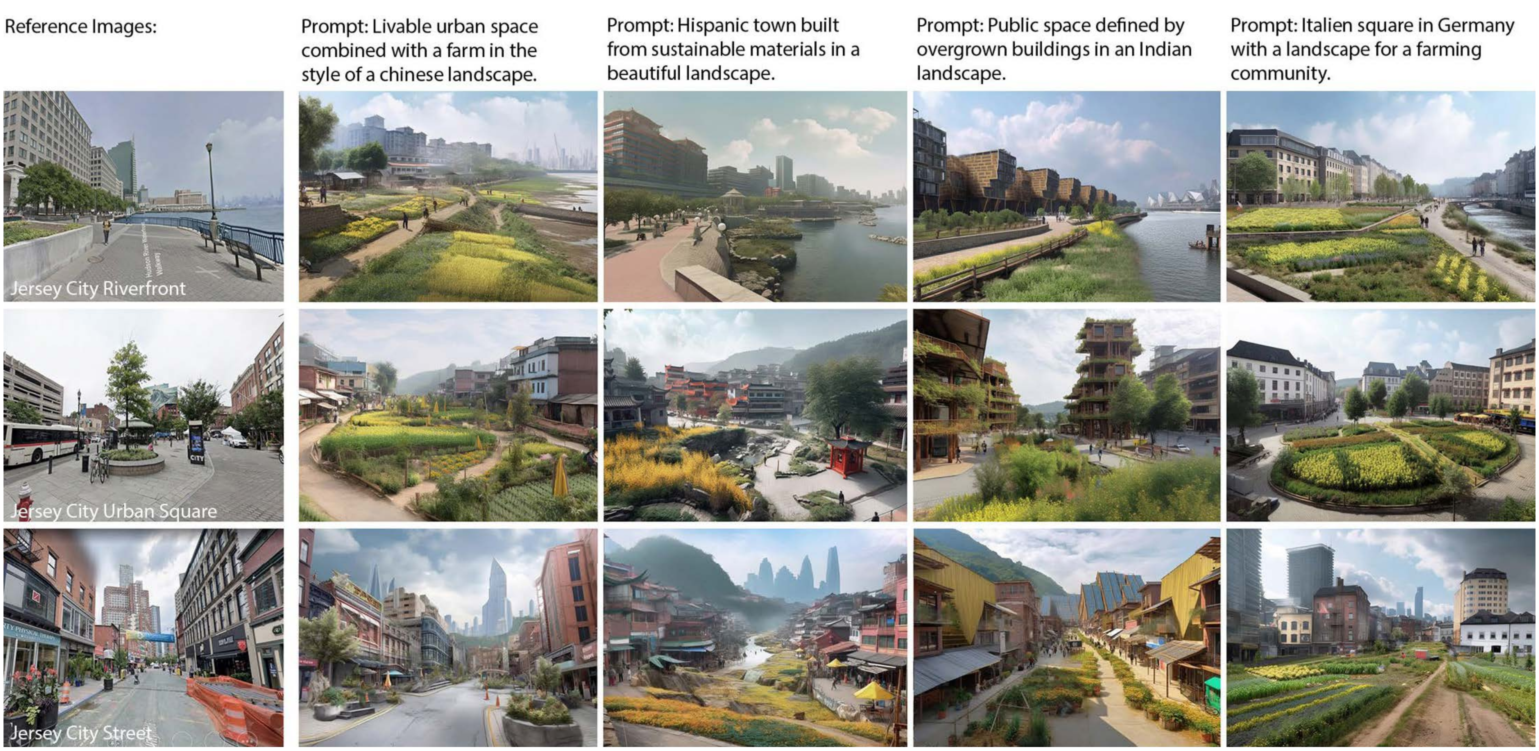
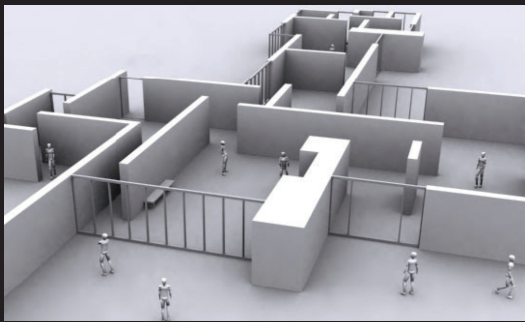
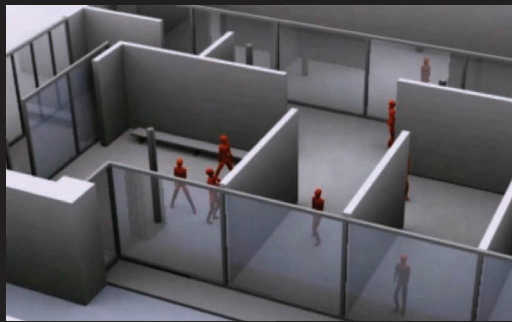


Figure. AI renderings using Google Street View images of urban samples, quality prompts, and cultural references.

This paper investigates how artificial intelligence (AI) tools can be used in the early design phase of public space to synthesize qualitative information such as cultural differences and individual perception into a common understanding of qualities of a public space. In that way AI tools can be used to synthesize emotional responses citizens may have to urban spaces as a form of feedback during the design process. To do that the investigation leverages AI tools' text-guided image-to-image translation techniques, their capacity to assess the association between images and texts and the premise for synthesizing a common understanding of characteristics and qualities of public spaces.

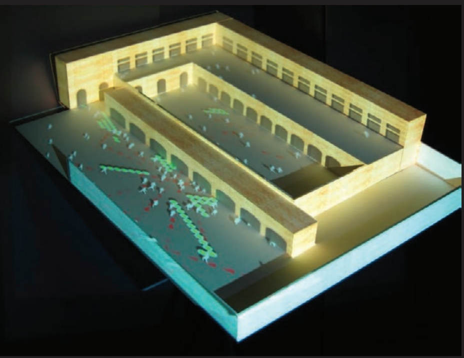
Other Selected Research Projects 2005 - 2015:

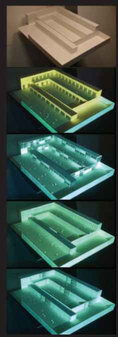


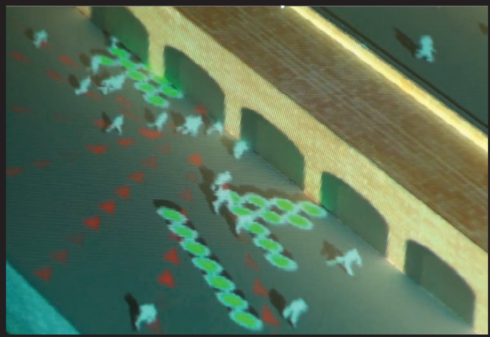


The Space ReActor: Walking a Synthetic Man through Architectural Space 2007.
[PDF, Thesis, Website]

The proposed method introduces a walking scale figure in a geometric model. Through agent-based computation, it moves inside the model and displays various behaviors in reaction to spatial characteristics such as transparent surface, opaque surface, perforation, and furniture. The figure is assigned a psychological profile with a different degree of sociability and reacts to proximity and visibility of others in the same model. (Adviser: T. Nagakura)

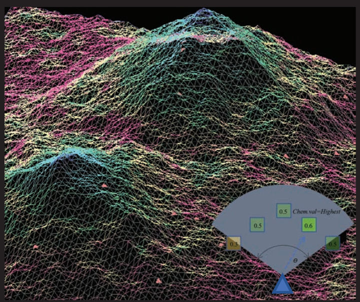




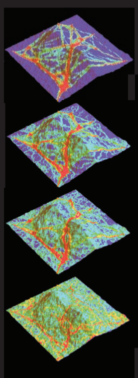


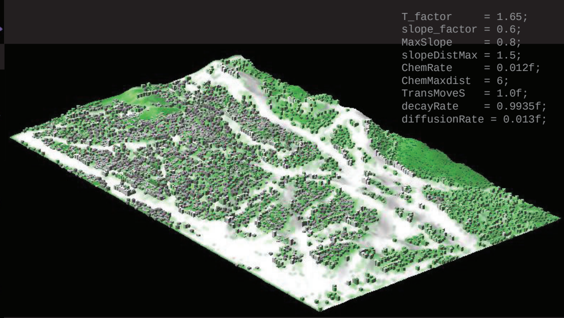
Crowd Mapper: Projection-based Interactive Pedestrian Agents for Collective Design, 2015.
[PDF, Website]

Crowd Mapper is a projection-based, spatial design tool that can visualize possible pedestrian behaviors associated with spatial configurations. The method allows real-time editing of building designs and agents' objectives by users. The projection mapping allows architects to project multiple design schemes with 3-D depth. While the method conforms to the traditional production standard using massing models, it stimulate architects' creativity by its real-time editing capabilities.



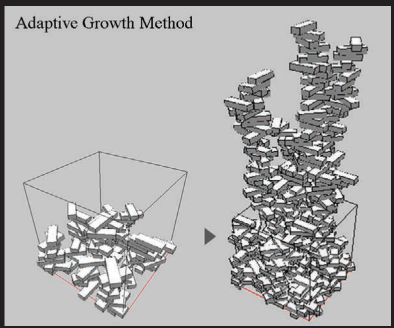
Agent's Behaviors

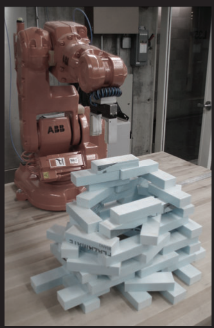


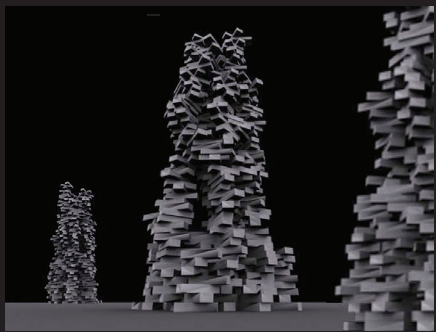


Self-organizing Computation: A Program for Simulating the Origins of Urban Form, 2010.
[PDF, Thesis, Website]

The project explores a generative aspect of agent-based models on an urban scale and proposes a method that simulates growth processes of settlement patterns. The method uses interactions and feedback among agents and environment and produces new instances of a spatial layout with paths and buildings from primary inputs of a given landform and environmental conditions. The implementation is based on the formation of trail systems by certain ant species using chemotaxis.

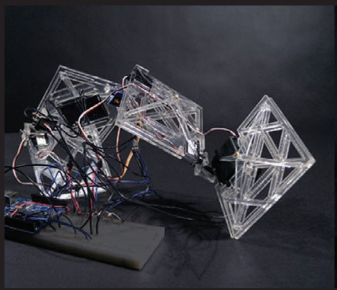


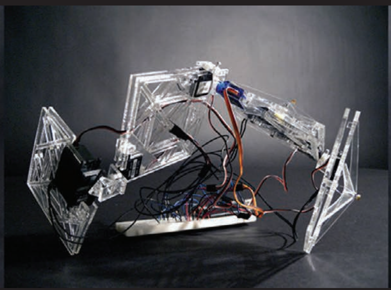


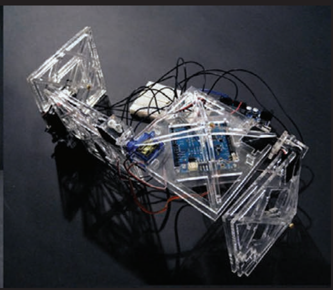


A Generative Approach to Robotic Fabrication: Adaptive Growth vs Plan Execution 2013.
[PDF, Website]

The proposed experiment uses an industrial robot arm to produce structures by stacking unit bricks without comprehensive hard-coded instructions for construction ("blueprints") from the outset. The program can produce structures that satisfy a certain characteristic – a simple rule of physics and a footprint of the site – while maintaining some level of morphological variations using a stochastic selection process. The system can proceed without having a fixed blueprint or a specific position in every step.







Designing for Constant Change: 2010. An Adaptable Growth Model for Architecture
[PDF, Website]

Design of universal components that can tolerate environmental and circumstantial changes over time is a challenge for an architect. I propose a scaled prototype of architectural components that can reconfigure into globally functional configurations based on feedback from locally distributed sensors embedded inside the component. The project aims at demonstrating a design system that can respond to dynamically changing environment over time.

Credit:

Dr, Taro Narahara
Associate Professor

New Jersey Institute of Technology
University Heights
Newark, NJ 07102-1982
Email: taronarahara@gmail.com