Automatic Rendering of Building Floor Plan Images from Textual Descriptions in English

Mahak Jain, Anurag Sanyal, Shreya Goyal, Chiranjoy Chattopadhyay, Gaurav Bhatnagar

Abstract

Human beings understand natural language description and could able to imagine a corresponding visual for the same. For example, given a description of the interior of a house, we could imagine its structure and arrangements of furniture. Automatic synthesis of real-world images from text descriptions has been explored in the computer vision community. However, there is no such attempt in the area of document images, like floor plans. Floor plan synthesis from sketches, as well as data-driven models, were proposed earlier. This is the first attempt to automatically render building floor plan images from textual description. Here, the input is a natural language description of the internal structure and furniture arrangements within a house, and the output is the 2D floor plan image of the same. We have experimented on publicly available benchmark floor plan datasets. We were able to render realistic synthesized floor plan images from the description written in English.

I. INTRODUCTION

Drawing floor plans for a building is an essential an inseparable part of early design phase of every construction project. However, this is a highly technical and sophisticated process, hence requires experience, skill and time. Rendering of floor plan images automatically from the requirements given by the clients is a hard problem. One motivation to generate floor plans, for a predetermined building exterior, is that procedural building generators often create only a building facade without the interior. However, converting the client’s requirements, stated in English language, requires Natural Language processing.

A typical system with natural language processing capability may indulge into analysis and/or generation of sentences written in a particular language. Analysis is a crucial task as it recognizes several paraphrases for the same command or information. Synthesizing images from textual description has gained a lot of interest after the introduction of models like Word2Vec [1], Adversarial Models [19]. Although, researchers have made significant progress in synthesizing natural images, document image is yet to be explored to that extent. In the field of computer graphics, synthesis of images has always been the focus. There has been efforts to vectorize scanned documents, or converting sketches into printed graphics. However, translating natural language description into graphical documents has not been explored. To the best of our knowledge, ours is the first attempt in synthesizing floor plan images from textual descriptions in English.

Figure 1 depicts a typical scenario, where a textual description is converted into a 2D floor plan image. In this paper, we concentrate on the internal organization of spaces within the building, and not on the external appearance. Our proposed framework is designed for computer graphics applications, like computer games, virtual reality as well as business applications such as real-estate rent/sale, building information modeling. Hence, instead of quantitative evaluations, we have presented the qualitative results. In practice, it is very difficult for a common user to articulate the requirement of a house clearly to an architect. Synthesis of a prototype model and an iterative refinement of the same will be immense help to both. Under this premise we decided to propose a learning framework to automatically render building floor plan images which looks visually similar to actual floor plans from textual description written in English.

In this paper, we focus on building an artificially intelligent framework for providing support to the architects or designers in the early-stages of the architectural design. Any building design process begins with an idea and with series of transformation (both cognitive and practice) results in to the end product, i.e. the final physical infrastructure. However, it is not easy (if at all possible) to translate the idea into something tangible that will help in early design phase. Internationally, standard for semantic building models such as the Industry Foundation Classes (IFC) become more widely adopted. However, the adaptation of the same is far from reality. Currently, there are no adequate means of using such building information models and the information they contain for supporting the early stages of the architectural design process.

II. RELATED WORK

Automatically designing layout helps the architects in the early design stages. Readers are requested to refer to the detailed review paper presented in [2]. Graph theoretic approaches has also been popular among the researchers for various application areas like GIS, computational geometry, and most importantly architectural design. In [3], an area-optimization problem was proposed and it was shown that it is NP-hard to find a minimum-area rectangular layout of a given contact graph.

In [11], a new graph-theory based model was proposed for representing and solving the problem of automatic layout design. The model is embodied in a detailed algorithm for generating floor plans that meet specified constraints. However, the model is not fully automatic as the complete specification of rooms and adjacencies, as well as the initial layout, are left to the architect. In [2], a review of the history of automated facility layout, focusing particularly on a set of techniques which optimize a single
As we enter the house, there is a lobby. It is square shaped with 25 feet. There is door on the second wall. This door leads us into hall. Hall has 5 walls with dimensions - 23, 24, 25, 36, and 24 feet. It has doors on third and fourth walls. The first door leads to bedroom. It is 4 sided with dimensions 23, 24, 23 and 24 feet. It has a door on the first wall. This door leads to bathroom. It is rectangular with dimensions 20x30 feet. There is no door in here. Exit bathroom. Exit bedroom. The second door leads to kitchen. It is 4 sided with dimensions 23, 24, 23 and 24 feet. It has a door on the first wall. This door leads to dining area. There is no door in here. Exit dining area. Exit kitchen. Exit hall. Exit lobby.

Fig. 1. Illustration of floor plan generation from textual description.

III. METHODOLOGY

Figure 2 depicts the proposed framework. As shown, there are various intermediate stages, which are discussed in details in the following sub-sections.

A. User Requirement Understanding

To understand the requirements specified by the user we first pre-process the description to extract useful information from it. This requires splitting the large raw text (requirements) into sentences to obtain more meaningful information. Then we extract the room related information and extract their positions in floor plan. Next, we perform the sentence Tagging, i.e. labelling each sentence based on their semantics. This tagging enables extraction of adjacency information among different rooms and entities. Finally we determine the appropriate placement model to keep the furniture as per the users requirement.
1) Text Summarization: There are two approaches to summarize text, extractive and abstractive. Extractive summarization uses statistical approach for selecting important sentences or keywords from document and concatenate them to form a summary. While, the abstractive approach generates a summary that keeps original intent and more like how humans summarize. In this work, we have followed a extractive model. The basic steps of the proposed model are; (i) pre-processing of sentences, (ii) construction of an undirected graph to calculate similarity between sentences, (iii) labelling of each word with a word class, (iv) ranking of the sentences based on the word class similarity, and (v) finally, inclusion of the top ranked sentences in the summary. Let, there are two sentences $D, Q$, which are comprised of a set of $n$ words. Also the the sentence $D$ is represented as: $D = d_k, k = 1, 2, ..., n$. Here, $d_k$ denotes a word. The sentences are considered as the vertices of the un-directed graph, while edges between a pair of node is labelled with the similarity score between the pair of sentences represented by those two vertices. The similarity score [4] followed in our work is represented as:

$$
S(D, Q) = \sum_{i} \left( \frac{\mathcal{F}(q_i, D) (\alpha + 1) \mathcal{I}(q_i)}{\mathcal{F}(q_i, D) + \alpha (1 - \beta + \beta \frac{|D|}{L}) + \gamma} \right)
$$

(1)

where, $\mathcal{F}(q_i, D)$ how many times the word $q_i$ occurs in the description $D$, $|.|$ is the length function, and $L$ is the average length of a sentence in the corpus. The function $\mathcal{I}(q_i)$ is defined as:

$$
\mathcal{I}(q_i) = \log \frac{|D| - \mathcal{D}(q_i) + 0.5}{\mathcal{D}(q_i) + 0.5}
$$

(2)

where, $D$ is the total number of sample description we have collected from the volunteers, and $\mathcal{D}(q_i)$ is the number of documents in which $q_i$ is present. The hyper parameters $\alpha, \beta$, and $\gamma$ are assigned with values 1.2, 0.5, and 1, respectively. The proposed method is very similar to Page Ranking Algorithm. We take reduction ratio as input in this step. The advantages of such approach are: (i) easier to work with and does not require learning, (ii) easy to compute because it does not deal with the semantics. However, the limitations are: (i) lack of flexibility and cannot paraphrase. Example - If there is no sentence that explicitly describes the location of a particular room, we will not be able to extract its location, (ii) results in lengthy summary, and (iii) suffers from inconsistencies when presented with conflicting data. Despite its limitations, the rationale behind adopting extractive summarization is that an abstractive approach: (i) has extremely high hardware demands, (ii) unavailability of substantial amount of data to train such a model, and (iii) high computational complexity of the abstractive techniques.

B. Sentence Classification and Clustering

The summary generated in previous step needs to be rearranged. Thus, we convert words into vectors using word2vec. A Binary CNN (Convolutional Neural Network) classifier is trained for each room to label each sentence with a room which most
likely it describes. Then, we cluster sentences room-wise. The data for each room used to train the classifier is scraped from internet as well as from the descriptions collected through survey. The rooms for which we have trained Binary CNN are: (i) bedroom, (ii) kitchen, (iii) bathroom, (iv) hall, (v) dining area. There is an additional CNN model trained to label relational sentences. The sentences which describes adjacency or door connectivity relationship between multiple rooms are defined as relation sentences. The CNN network contained 3 convolutional layers of filter sizes 3, 4 and 5 with ReLU activation function, where output from each layer is max-pooled using pooling layer. A drop-out layer is added to regularize the CNN network and Softmax function is used to normalize the output probabilities.

For example, consider the sentence "The bedroom leads to bathroom." The reason we used binary classifier is that there can be sentences which describe multiple rooms. For example, "The bedroom is adjacent to hall and there is a bed in the centre." This sentence describes bedroom and as well as define an adjacency relation between hall and bedroom. Figure 3 shows the a typical depiction of the CNN layers. As a result of the parsing we are to extract the following entities from the description:

1) Rooms and their Labels
2) Various furniture present in the rooms
3) Relative position among the rooms
4) Relative position of the furniture within a room
5) Position of doors and windows on the walls of the floor plan

1) Information extraction: From the previous step, we get sentences labelled with room tag or relation tag. We treat sentences with relation tag separately to extract relational information. As part of relational information, we extract only door connectivity relation between rooms. From sentences labelled with room tag, we extract information like shape, dimensions, architectural objects inside room, door connectivity and wall sharing for each tag using techniques like tokenization, regular expression matching. For extracting rooms and architectural objects, we have built our own custom dictionaries. From the extracted information, we are generating a JSON object for each room which contains the following details:

Room {
    type : hall
    shape : rectangle
    sides : 4
    dimensions : [250, 200, 250, 200]
    door_placement : [(2:1), (3:1)]
    furnitures : [(sofa:1), (chair:1)]
}

The attribute values are stored in such a data structure for the ease of development and future scope. For example, the value [(2 : 1)] under the door placement denotes that there is one door at the second wall. The walls are numbered in the clockwise manner. From sentences labelled with relation tag, we make a graph with rooms as nodes and an edge is drawn between two rooms if the rooms are connected by a door. This graph is termed as Door Connectivity Graph. Figure 4 depicts a DFS tree for the floor plan image shown in Fig. 1. Thus, this step gives two outputs: (i) List of JSON objects, and (ii) Door Connectivity Graph.

2) Floor Plan Rendering: Using the above extracted information, we generate the output image of the floor plan. We traverse the door connectivity graph in Depth First Search fashion. Using the extracted information of each room and keeping track of previously visited room in recursive manner, we keep generating the coordinates of each room and architectural objects. There
are twelve different decor symbols defined in the SESYD [5] dataset. In this paper we have used those symbols to represent the architectural objects. A space of floor plans is parameterized by the horizontal and vertical coordinates of the rectilinear segments that form the shape of the plan. There are two coordinate systems, namely global (for the floor plan) and local (for the rooms). We first synthesize the rooms, in isolation, individually as per the local coordinate system. After that, we align the entry room to the global coordinate system, and then position the other rooms with respect to that. The final arrangement is obtained after placing all the rooms to the global coordinate system.

IV. RESULTS

In the following subsections we are presenting the results of the individual processing units of the proposed framework. Due to lack of space we are showing all the intermediate steps of rendering a single floor plan from a given textual description.

A. Text Summarization

As discussed in Sec. III-A1 the input to this stage could single or multiple descriptions of a floor plan. As an output, this stage generates a summary by extractive technique. Figure 5 depicts the result of summarization on one such description. It can be observed from the result that the proposed method is able to extract the necessary quantitative information about all the rooms. Moreover, it also extracts the adjacency and connectivity information from the given raw text. Thus, the output has the necessary information which can be used for sentence tagging and clustering (as discussed in Sec. III-B).

B. Information Extraction

It takes tagged sentences as input and generate the JSON object, which has the necessary information discussed in Sec. III-B1. Figure 6 depicts the result of summarization on one such description. Figure 6(a) denotes the transformation of a given description about a room to its corresponding JSON object. Figure 6(a)-(d) shows the JSON objects for the other rooms.

C. Door Connectivity Graph

Another key component of the floor plan generation algorithm is the generation of the Door Connectivity Graph (DCG) as discussed in Sec. III-B1. Figure 7(a) depicts a summerized description, while Fig. 7(b) depicts the DCG of the same. As shown in Fig. 7(b), the nodes depict the rooms (indicated by the first letter of every room type), while the edges denote that a pair of rooms are connected to each other via a door.
The floor plan has 5 rooms - kitchen, bathroom, hall, bedroom and dining. Lobby leads to hall. Hall and kitchen are adjacent and connected. Hall has a sofa and a chair. Kitchen leads to dining hall from its right door. Dining area has a table in corner. Dining and bedroom are adjacent. Bathroom and bedroom are adjacent and connected. There is a big bed in bedroom. Bathroom has a tub in corner. There is only a sink in kitchen. Lobby has 2 chairs. As we enter the house, there is a lobby and it is square shaped with 150 feet. It is at entrance and is a small room. Two chair are there in lobby and there is door on the second wall. The door of lobby leads to hall. Hall is 4 sided with dimensions 250x200 feet and has doors on second and third walls. It has a chair and sofa. First door of hall leads to bedroom. Bedroom is 4 sided with dimensions 200x100 feet and has a door on first wall. It has a bed in it. Door of bedroom leads to bathroom. Bathroom is rectangular with dimensions 100x100 feet. It has a tub inside it. Second door of hall leads to kitchen. Kitchen is 4 sided with dimensions 100x200 feet with a door on second wall. There is a sink and basin inside kitchen. Door of kitchen leads to dining area. Dining area has 4 walls with dimensions 150x100 feet. There is a dining table placed in center.

The door of lobby leads to hall. Hall is 4 sided with dimensions 250x200 feet and has doors on second and third walls. First door of hall leads to bedroom. Bedroom is 4 sided with dimensions 200x100 feet and has a door on first wall. Door of bedroom leads to bathroom. Second door of hall leads to kitchen. Kitchen is 4 sided with dimensions 100x200 feet with a door on second wall. Door of kitchen leads to dining area. Dining area has 4 walls with dimensions 150x100 feet. The floor plan has 5 rooms - kitchen, bathroom, hall, bedroom and dining. Hall and kitchen are adjacent and connected. Kitchen leads to dining hall from its right door.

D. Image rendering

The coordinates of room are generated in recursion and random corner is picked to place the architectural objects inside room. It is generated using the extracted information stored in room objects. Figure 8 depicts result of rendering a single room from the textual description. As shown in the Fig. 8(b) the room labeling, shape, positioning of the door is as per the JSON object shown in Fig. 8(a). The furniture (sofa, and chair) are placed along the wall. While placing the furniture, following points are kept in mind that they don’t block the door (bounding boxes are not overlapping).

Figure 9 depicts the rendering of an entire floor plan from a given description. It can be observed that the proposed framework is able to generate a floor plan as per the specification given in the textual description. Fine grain details like wall width (thick, thin), type of materials to be used for walls (brick, concrete, glass etc), positioning of the furniture have not been taken into consideration (neither in the description, nor in the synthesis), at present.
The door of lobby leads to hall. Hall is 4 sided with dimensions 250x200 feet and has doors on second and third walls. First door of hall leads to bedroom. Bedroom is 4 sided with dimensions 200x100 feet and has a door on first wall. Door of bedroom leads to bathroom. Second door of hall leads to kitchen. Kitchen is 4 sided with dimensions 100x200 feet with a door on second wall. Door of kitchen leads to dining area. Dining area has 4 walls with dimensions 150x100 feet. The floor plan has 5 rooms - kitchen, bathroom, hall, bedroom and dining. Hall and kitchen are adjacent and connected. Kitchen leads to dining hall from its right door.

(a) Summarized description

(b) Door Connectivity Graph

Fig. 7. An example of Door Connectivity graph constructed from a given description.

| Type  | Hall      |
|-------|-----------|
| Shape | Rectangle |
| Sides | 4         |
| Dimensions | [250, 200, 250, 200] |
| Door Placement | [(2:1), (3:1)] |
| Furnitures | [“sofa”:1], [“chair”:1] |

Fig. 8. An example of rendering a single room from the information extracted and JSON objects.

V. CONCLUSION AND FUTURE WORK

In this paper, we have presented a framework that takes the client’s requirement written in English language, and convert the same into a 2 dimensional floor plan. The proposed method is a combination of natural language processing, machine learning and computer graphics, along with the domain knowledge of architectural design. The floor plan images thus generated are to a great extent designed for ideal condition. In real-world, an architect designs a floor plan by keeping into mind several qualitative factors Some of them are: (i) the weather condition, (ii) the purpose of the building, e.g. if the premise is to be used as classroom, then the orientation, aspect ratio, placement of the furniture will be crucial, (iii) aesthetics of the rooms, (iv) accessibility between a pair of rooms, (iv) non-rectilinear structure of the walls, etc. We believe that introduction of such factors into the framework will make it more robust, contemporary, and usable. The present work is probably the beginning of a new research field which will benefit the real-estate industry in the early design phase.

REFERENCES

[1] Mikolov, Tomas and Sutskever, Ilya and Chen, Kai and Corrado, Greg S and Dean, Jeff Distributed Representations of Words and Phrases and their Compositionality, Curran Associates, Inc., p.p. 3111–3119, (2013)
[2] March, Lionel and Steadman, Philip The geometry of environment : an introduction to spatial organization in design, London : RIBA Publications, 1971
[3] Buchsbaum, Adam L. and Gansner, Emden R. and Procopiuc, Cecilia M. and Venkatasubramanian, Suresh Rectangular Layouts and Contact Graphs, ACM Trans. Algorithms, 4(1), pp. 8:1–8:28, (2008).
[4] Lv, Yuanhua and Zhai, ChengXiang Lower-bounding Term Frequency Normalization, Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 978-1-4503-0717-8, 2011
The door of lobby leads to hall. Hall is 4 sided with dimensions 250x200 feet and has doors on second and third walls. First door of hall leads to bedroom. Bedroom is 4 sided with dimensions 200x100 feet and has a door on first wall. Door of bedroom leads to bathroom. Second door of hall leads to kitchen. Kitchen is 4 sided with dimensions 100x200 feet with a door on second wall. Door of kitchen leads to dining area. Dining area has 4 walls with dimensions 150x100 feet. The floor plan has 5 rooms - kitchen, bathroom, hall, bedroom and dining. Hall and kitchen are adjacent and connected. Kitchen leads to dining hall from its right door.

(a) Textual description  
(b) Rendered floor plan

Fig. 9. An example of rendering an entire floor plan image from textual description written in English.

[5] Delalandre, Mathieu and Valveny, Ernest and Pridmore, Tony and Karatzas, Dimosthenis Generation of synthetic documents for performance evaluation of symbol recognition & spotting systems, IJDAR 13(3) pp. 187-207, (2010).
[6] Legakis, Justin and Dorsey, Julie and Gortler, Steven Feature-based Cellular Texturing for Architectural Models, Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, 2001.
[7] Robin S Liggett Automated facilities layout: past, present and future, Automation in Construction, 2000
[8] Vincenzo Donato Towards design process validation integrating graph theory into BIM, Architectural Engineering and Design Management, 13(1), 22-38, 2017.
[9] Jerzy Grobelny and Rafa Michalski A novel version of simulated annealing based on linguistic patterns for solving facility layout problems, DAS, 2010Knowledge-Based Systems, 124, p.p.55-69, (2017).
[10] Abbas Ahmadi, Mir Saman Pishvaee and Mohammad Reza Akbari Jokar A survey on multi-floor facility layout problems, Computers & Industrial Engineering, 107, 158-179, 2017
[11] A. Schwarz and D.M. Berry and E. Shavit Representing and solving the automated building design problem, 26(9), 689-698, 1994.
[12] Pottmann, Helmut and Schiftner, Alexander and Bo, Pengbo and Schmiedhofer, Heinz and Wang, Wenping and Baldassini, Niccolo and Wallner, Johannes, Sketching Reality: Realistic Interpretation of Architectural Designs, ACM Trans. Graph. 29(1), 27(3), p.p. 0730-0301, (2008)
[13] Xuejin Chen and Kang, Sing Bing and Xu, Ying-Qing and Julie Dorsey and Heung-Yeung Shum Text information extraction in images and video: a survey, ACM, 27(2), 2008.
[14] Müller, Pascal and Zeng, Gang and Wonka, Peter and Van Gool, Luc, Image-based Procedural Modeling of Facades, ACM Trans. Graph, 26(3), 2007.
[15] Lefebvre, Sylvain and Hornus, Samuel and Lasram, Anass By-example Synthesis of Architectural Textures, ACM Trans. Graph, 29(4), 2010
[16] Whiting, Emily and Ochsendorf, John and Durand, Frédéric, Procedural Modeling of Structurally-sound Masonry Buildings, ACM Trans. Graph., 28(5), 2009.
[17] Müller, Pascal and Wonka, Peter and Haegler, Simon and Ulmer, Andreas and Van Gool, Luc, Procedural Modeling of Buildings, ACM Trans. Graph., 25(3), p.p. 614-623, 25(3), (2006).
[18] Harada, Mikako and Witkin, Andrew and Baraff, David Interactive Physically-based Manipulation of Discrete/Continuous Models, Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques, p.p.199-208, (1995)
[19] Moustafa Elzantot and Yash Sharma and Ahmed Elgohary and Bo-Jhang Ho and Mani B. Srivastava and Kai-Wei Chang Generating Natural Language Adversarial Examples, 2018.
[20] Marcus, Mitchell P and Marcinkiewicz, Mary Ann and Santorini, Beatrice Computer-generated Residential Building Layouts, ACM Trans. Graph, 29(6), (2010).
[21] Hahn, Evan and Bose, Prosenjit and Whitehead, Anthony Persistent Realtime Building Interior Generation, Proceedings of the 2006 ACM SIGGRAPH Symposium on Videogames, 2006
[22] Pottmann, Helmut and Schiftner, Alexander and Bo, Pengbo and Schmiedhofer, Heinz and Wang, Wenping and Baldassini, Niccolo and Wallner, Johannes Freeform Surfaces from Single Curved Panels, ACM Trans. Graph., 27(3), (2008)
[23] Jess Martin Procedural house generation: A method for dynamically generating floor plans, Symposium on Interactive 3D Graphics and Games, 2006.