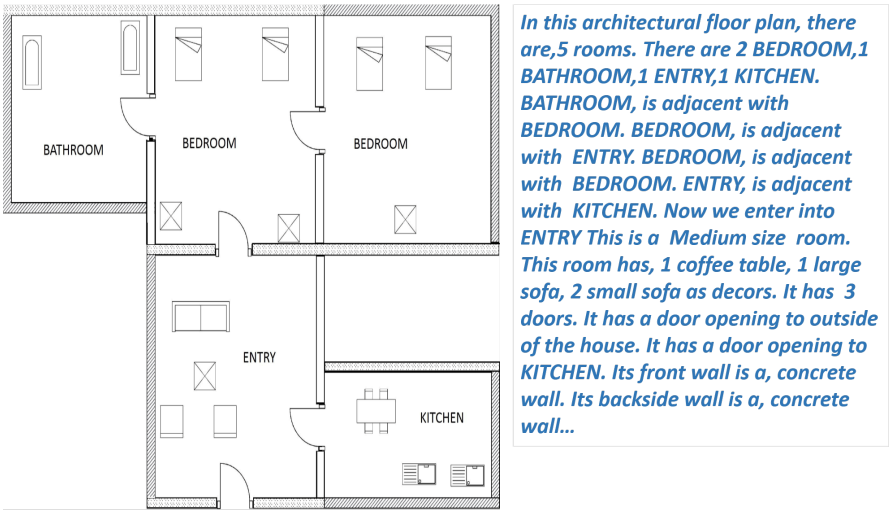
4

Textual interpretation generation of indoor scenes using annotated floor plans

In Chapter 3, the construction of the dataset BRIDGE was described along with the targeted experiments. The dataset is constructed to perform decor classification, caption, and paragraph generation and evaluation. Understanding a floor plan requires detection and recognition of its several components, such as decors, walls, rooms, and text (if annotated). The annotations proposed in the BRIDGE dataset are used to recognize these components and their evaluation in general floor plan images. Hence, before the construction of BRIDGE, we explored floor plan understanding using the pre-annotated floor plans. A variety of floor plans are annotated with room names and dimensions. This information can be used in their knowledge and interpretation. For accurate interpretation, efficient OCR techniques need to be used. The information contained in an annotated floor plan image is extracted using OCR techniques and compiled in a database structure. Other information that could not be extracted using OCR, such as wall characterization, doors, decor, and entry identification, are extracted using suitable feature matching techniques. All extracted information is compiled into a database structure and rooms inside the floor plans are parsed in a depth-first search manner to generate a consolidated description for the entire floor plan. The generated description for the indoor scene has wide applications such as indoor navigation. The automatic navigation system also facilitates visually impaired people by (i) helping them to localize themselves inside the house or (ii) giving continuous feedback about the surroundings so that users could realize whether they have reached their destination or not. However, all these require extensive computation and use of various sensors and actuators. This chapter generates a description of an indoor scene by taking cues from a building floor plan. The narration is synthesized in such a way as if some person is navigating through the house with a wearable camera. This narration can cater to (i) potential buyers of the home who are unaware of the technical details of the floor plans, (ii) visually impaired people (using some text reader software) can imagine the interior of the house, (iii) textual descriptions, with the help of natural language processing can be used for similar floor plan retrieval. Figure 4.1 exemplifies the problem and the potential solution for real-world floor plan images. The key characteristics that make this work unique are: (i) proposing a unified framework for egocentric vision-based narration synthesis from floor plan images; (ii) an efficient technique to classify several construction materials used for building the house; and (iii) semantic classification of various categories of rooms in the floor plan.

The rest of the chapter is organized as follows. Section 4.1 gives a brief overview of the proposed system, Sec. 4.2 describe the various steps taken to understand each component of the floor plan image and extract information, Sec. 4.3 describe the method proposed for generating description from the data extracted in the previous sections, Sec. 4.4 describes the experiments performed on various metrics to validate the proposed method and Sec. 4.5 summarize the proposed method in this chapter, also describing the potential future steps.



In this architectural floor plan, there are,5 rooms. There are 2 BEDROOM,1 BATHROOM,1 ENTRY,1 KITCHEN. BATHROOM, is adjacent with BEDROOM. BEDROOM, is adjacent with ENTRY. BEDROOM, is adjacent with BEDROOM. ENTRY, is adjacent with KITCHEN. Now we enter into ENTRY This is a Medium size room. This room has, 1 coffee table, 1 large sofa, 2 small sofa as decors. It has 3 doors. It has a door opening to outside of the house. It has a door opening to KITCHEN. Its front wall is a, concrete wall. Its backside wall is a, concrete wall...

(a) Input Image (b) Ego centric Description

Figure 4.1 : An illustration of the problem of egocentric vision based narration synthesis from a given input floor plan image.

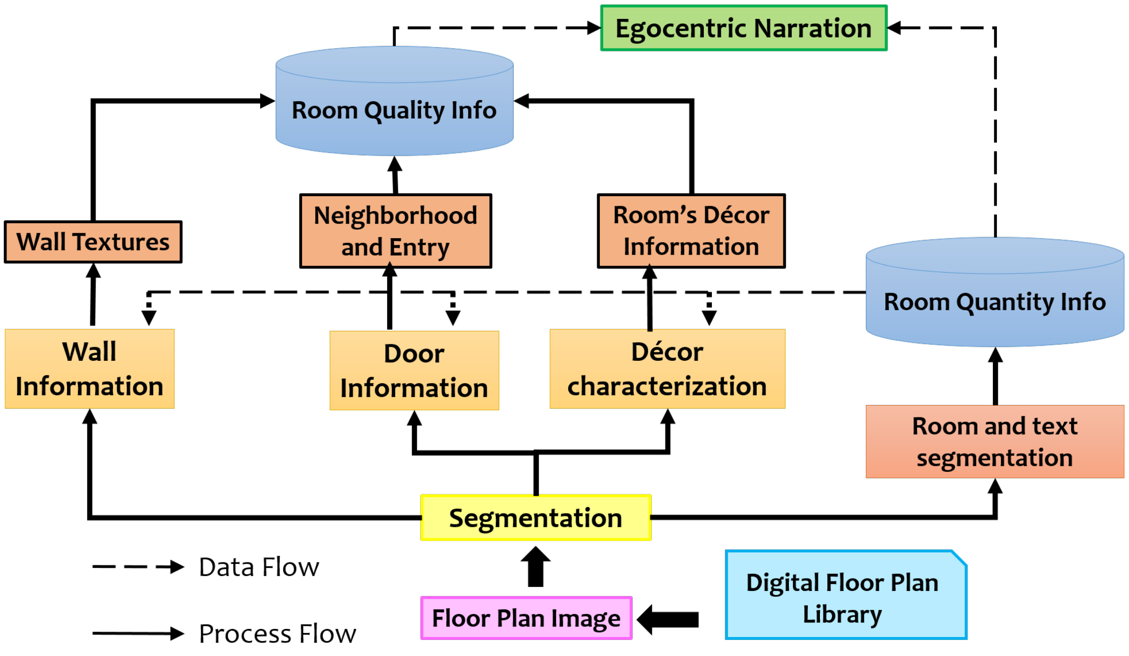


Figure 4.2 : Block diagram of our proposed framework.

Table 4.1 : Room Quantity Info schema for the Example in Fig. 4.4. Room Names: Bed Room (BD), Bath Room (BT), Entry (EN), Kitchen (KN).

Pivot	Name	CordX	CordY	Area	Label
803, 386	BD	596, 596, 1246, 1246	21, 738, 21, 738	467279	2
1479, 382	BD	1271, 1271, 1948, 1948	21, 740, 21, 740	487180	4
262, 439	BT	21, 21, 573, 573	21, 608, 21, 608	324150	1
1033, 1078	EN	596, 596, 1250, 1250	763, 1528, 763, 1528	498879	3
1722, 1232	KN	1274, 1274, 1951, 1951	1126, 1524, 1126, 1524	269710	5

4.1 BRIEF OVERVIEW

In this chapter, a scheme for understanding and interpretation of annotated floor plans is proposed. For this purpose, we annotated the existing dataset ROBIN (Annotated-ROBIN) as discussed in chapter 3, is used. Figure 4.2 represents the overall framework of the system. The whole process is divided into two main stages, segmentation and processing its results for information extraction and narration synthesis. The first stage performs room segmentation and text only image processing in parallel. Room segmentation gives information such as walls, doors, windows, decors, area of the rooms, and text-only images. Text only image is processed and to generate a logical room name out of it. This information is stored collectively in a relation (schema) "Room Quantity Info" (see: Tab. 4.1), which is used for further processing of walls material, neighborhood and entry detection, decor characterization, and room's size identification. Finally, the extracted qualitative information is stored in a relation "Room Quality Info" (see: Sec. 4.3), which is further used for narration synthesis. For both schemas, room ***Label*** is the primary key and used to map all information.

4.2 INTERIOR PROCESSING

4.2.1 Semantic Segmentation

We have adopted the technique proposed in D. Sharma, C. Chattopadhyay and G. Harit [2016] for the identification of rooms. Walls are detected by performing morphological closing on the input floor plan image I. To delineate room boundaries, we detect doors using scale-invariant features and close the wall image gaps corresponding to the door locations. We identified the connected components in the wall image by applying the flood fill technique to obtain the rooms. The obtained connected components are the required rooms, and their locations are obtained. Moreover, we calculate the areas of the respective rooms and store all the information obtained in the "Room Quantity Info". Finally, labels are assigned to each room after connected component labeling. Table 4.1 stores the final information of this processing step and the column `CordX, CordY, Area, Label` stores its results. Rooms with an area of more than 750000 pixels are classified as large (L), while areas below 450000 pixels are considered small (S), which is done empirically. All other room areas are considered as medium (M).

Algorithm 1 Room Caption Detection and Recognition

```
function TEXTPROCESS($I$)                                  ▷ $I$=Input Image
    $CC \leftarrow$conCmp($I$)                             ▷ Connected Components of $I$
    while $\zeta \in CC$ do                                ▷ $\zeta$ is individual component
        if $6 \leq |\zeta| \leq 50$ then                   ▷ $|.|$ = Major axis length
            $X_{max} \leftarrow max(X_\zeta)$              ▷ $X_\zeta$ set of $x$ co-ord $\in \zeta$
            $X_{min} \leftarrow min(X_\zeta)$
            $Y_{max} \leftarrow max(Y_\zeta)$              ▷ $Y_\zeta$ set of $y$ co-ord $\in \zeta$
            $Y_{min} \leftarrow min(Y_\zeta)$
            $Pv_x \leftarrow X_{r_{min}} + round((X_{r_{max}} - X_{r_{min}})/2)$
            $Pv_y \leftarrow Y_{r_{min}} + round((Y_{r_{max}} - Y_{r_{min}})/2)$
            $\zeta_\alpha \leftarrow recText(\zeta)$       ▷ Recognize the alphabet
            $\mathbb{T} \leftarrow \{\zeta_{id}, \zeta_\alpha, Pv_x, Pv_y\}$
        end if
    end while
    while $\tau \in \mathbb{T}$ do                         ▷ For all the alphabets
        d $\leftarrow ED(\tau.\zeta_{id}, \tau.\zeta_{id+1})$   ▷ Euclidean Distance
        if d $< 100$ then $\mathbb{W} \leftarrow \tau.\zeta_\alpha$
        end if
        if d $\geq 100$ then e $\leftarrow editDist(\mathbb{W}, \mathbb{V})$
            if e $\leq 2$ then
                asgnRoomCap($\mathbb{W}$)
            end if
        end if
    end while
end function
```

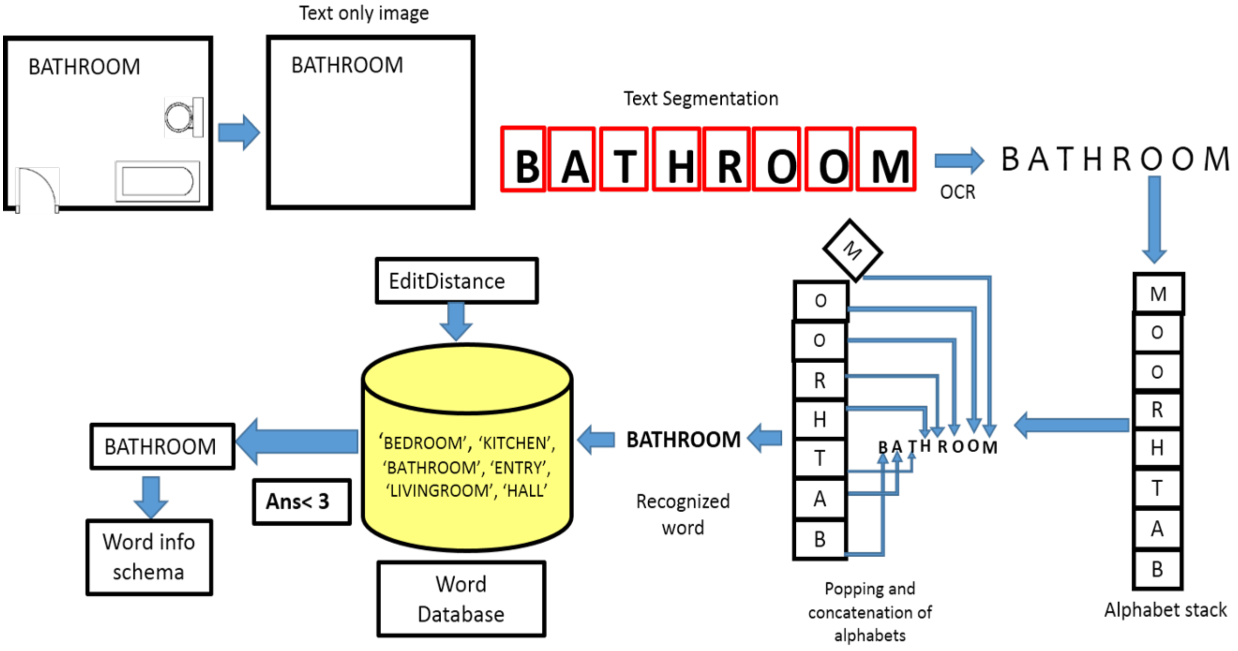
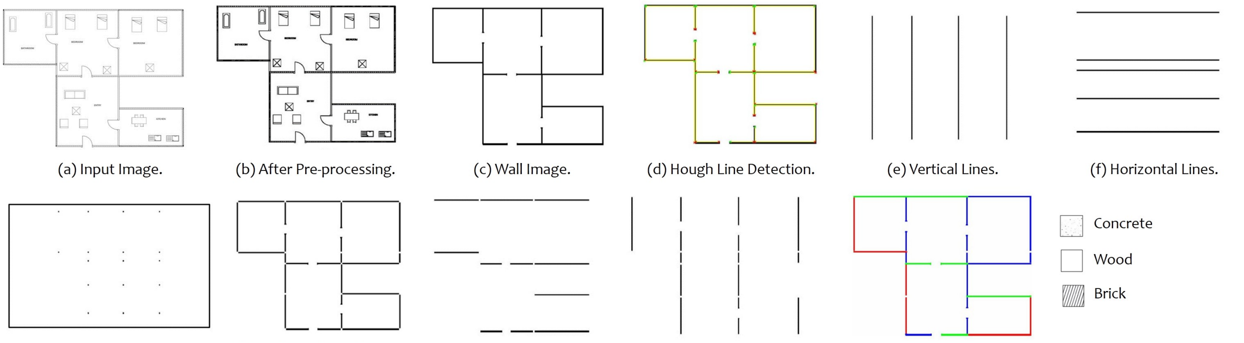


Figure 4.3 : Illustration of the text segmentation and recognition.



(a) Input Image. (b) After Pre-processing. (c) Wall Image. (d) Hough Line Detection. (e) Vertical Lines. (f) Horizontal Lines.

(g) Junction Points. (h) Walls without the Junction Points. (i) Horizontal Wall Segments. (j) Vertical Wall Segments. (k) Walls with various material highlighted with colors. (l) Legend of materials used.

Figure 4.4 : Intermediate results generated while textured wall segmentation. In part (k), ‘red’ denotes Brick, ’green’ denotes concrete, and ‘blue’ denotes wood. The corresponding textures are shown in part (l), which are adopted from Kilmer and Kilmer [2009].

4.2.2 Text only processing

Algorithm 1 detect and recognizes room captions in a floor plan. In line number 12, ζ_{id} refers identifier for a given alphabet. In line number 9, we calculate the words’ pivot points (P_v). Line number 16 evaluates the Euclidean distance between the current alphabet and the next in sequence to group them in a logical string. All the formed words here are validated by comparing it with a set of predefined words (Standard room names). Between a formed word $\mathbb{W}$ and a predefined word $\mathbb{V}$, of length $|\mathbb{W}|$ and $|\mathbb{V}|$, respectively, we calculate the Edit Distance (line no. 19) F.J. Damerau [1964]. If $\mathsf{e} \leq 2$, then it is considered as a valid room name, otherwise not. To find the belongingness of room name to any room, an inside-outside test for each room, formed by the room coordinates (“Room Quantity Info”), is performed (line no. 21 with the P_v of the room names (“Word info schema”). On success to the inside-outside test, a room name is assigned to a particular room. Quantitative information of a room is stored in “Room Quantity Info” schema (see Tab. 4.1). Column `Pivot, Name` of Tab 4.1, stores the combination of results of text-only image processing and semantic segmentation. Figure. 4.3 depicts the process of caption detection and text segmentation in floor plan images.

4.2.3 Wall Material classification

Wall image, W, as shown in Fig. 4.4(c) is obtained as an output of Sec. 4.2.1, after merging the texture with surrounding walls as shown in Fig. 4.4(b). W is given as input to Canny edge detector to get edge image $\mathbb{E}$. Hough Transform is applied on $\mathbb{E}$ (with θ= 0 and 90°) to detect horizontal and vertical lines (see Fig. 4.4(d)). Then parameter Space Matrix is generated whose peak values give potential lines. Let, n horizontal and m vertical lines are there. Consecutive horizontal lines are paired and region between these pairs is filled using flood-fill algorithm. Let $line_k$ be horizontal lines where $k = 1, 2, \ldots, n$. The pairs would be $\{(line_1, line_2) \ldots (line_{n-1}, line_n)\}$. The resulting image W_H as shown in Fig. 4.4(f) contains horizontal lines. Similarly vertical regions are filled, to obtain the image W_V as shown in the Fig. 4.4 (e) which contains vertical lines. Regions of intersection $\mathbb{C}$, junction point (see Fig.4.4(g)), between W_H and W_V is removed from the W (see Fig. 4.4(c)) by $\mathbb{C} = W \cap (W_H \cap W_V)$ where $\cap$ is the intersection operator. Removing $\mathbb{C}$ from W yields wall segment, ω as shown in Fig. 4.4(h). The longest segment would be a side of a room, which is assumed to be made up of same material.

Material Classification

Each wall segment is classified into one of the 3 classes i.e. brick (Br), concrete (Cnc) and wood (Wd) as shown in Fig. 4.4(l), and stored in F_1. For each of the given canonical texture of material, a mask is calculated and features are extracted using Local Binary Patterns (LBP)(A detailed introduction of LBP can be found in Ojala *et al.* [2002]), and a normalized histogram $(\mathbb{H}_B, \mathbb{H}_C, \mathbb{H}_W)$ is generated. Similarly, for each given wall segment s $\in \mathbb{S}$ where $\mathbb{S}$ is the set of wall segments obtained, and $|\mathbb{S}| = \mathbb{K}$, features are extracted and normalized histograms $\mathbb{H}_S$, where s $= 1,2,\ldots,\mathbb{K}$ are generated. Distance between $\mathbb{H}_S$ and histograms of a Canonical texture (H) is:

$$\mathbb{D}_{\mathrm{k}} = \sum_{i=1}^{n} \frac{(\beta_i^{\mathrm{s}} - \xi_i^{\mathrm{k}})^2}{(\beta_i^{\mathrm{s}} + \xi_i^{\mathrm{k}})} \tag{4.1}$$

where, k $\in \{Br, Wd, Cnc\}$, and $\mathbb{D}_{\mathrm{k}}$ is the distance Satorra and Bentler [2001] between a pair of histograms of material k, β and ξ are the bins corresponding to histograms of current segment s and material k respectively, and n is the number of bins. Minimum of these distances is calculated $\mathbb{D}_{min} = min(\mathbb{D}_{Br}, \mathbb{D}_{Cnc}, \mathbb{D}_{Wd})$, and corresponding texture is assigned to the segment s. All the classified segments are then represented using different colors according to the texture which they represent in image $\mathbb{F}_1$ as shown in Fig.4.4(k).

Walls to Room Mapping

Room label information is taken from room segmentation (Sec.4.2.1) where number of rooms is $\mathbb{R}$. A data structure “Wall Info Schema” is created with attributes $\{Label, Front, Backside, Left, Right\}$ (see Tab.4.2) to store the wall texture information corresponding to each room. Centroids of all the wall segments in $\mathbb{F}_1$ are calculated. For a segment s$\varepsilon\mathbb{S}$, where $\mathbb{S}$ number of segments, $(\mathrm{Cx}_s, \mathrm{Cy}_s)$ denotes its centroid. For any segment s, if the segment is horizontal, as shown in Fig. 4.4(i), then if there exists a room above it, it becomes the lower wall for that room and similarly upper wall if there is a room below it. If there exist a room at the location $(\mathrm{Cx}_s, \mathrm{Cy}_s - 30)$ in the floor plan image, then lower wall of the room at $(\mathrm{Cx}_s, \mathrm{Cy}_s - 30)$ is assigned the texture of segment s. Similarly the texture of segment s is assigned to the upper wall of any room present at location $(\mathrm{Cx}_s, \mathrm{Cy}_s + 30)$. For a segment s, if the segment is vertical as shown in Fig. 4.4(j) and if there exists a room on its Left side, it becomes the right side wall for that room and similarly Left side wall if there is a room on its right side. If there exist a room at the location $(\mathrm{Cx}_s - 30, \mathrm{Cy}_s)$ in the floor plan image, then right side wall of the room at $(\mathrm{Cx}_s - 30, \mathrm{Cy}_s)$ is assigned the texture of segment s. Similarly the segment s is checked for the left side wall of any room present at location $(\mathrm{Cx}_s, \mathrm{Cy}_s + 30)$, and its left side wall is assigned the texture of segment s. The “Wall Info Schema” (see Tab. 4.2) is updated accordingly.

4.2.4 Entry Detection

Figure 4.5 depicts the algorithm describing the entry door detection in floor plan image. Here, D_p is the door pixel, with $(\mathrm{D}_{p.X}, \mathrm{D}_{p.Y})$, corresponding x and y coordinates. The diagram shows the steps if the entry door is in a horizontal direction. The process for vertical doors goes same as horizontal doors with a difference of moving left and right instead of up and down. Moreover, the formulae given in the predicate (decision box) marked as $*$ and $**$ will be $\mathrm{D}_{p.X} - 1 \in W_f$ and $\mathrm{D}_{p.X} + 1 \in W_f$, respectively.

Table 4.2 : Wall info schema for the Example in Fig. 4.4. Material details: Concrete, Wood, and Brick.

Room Label	Front	Back	Left	Right
1	Concrete	Brick	Brick	Wood
2	Concrete	Concrete	Wood	Wood
3	Concrete	Concrete	Brick	Wood
4	Concrete	Wood	Brick	Wood
5	Wood	Wood	Wood	Wood

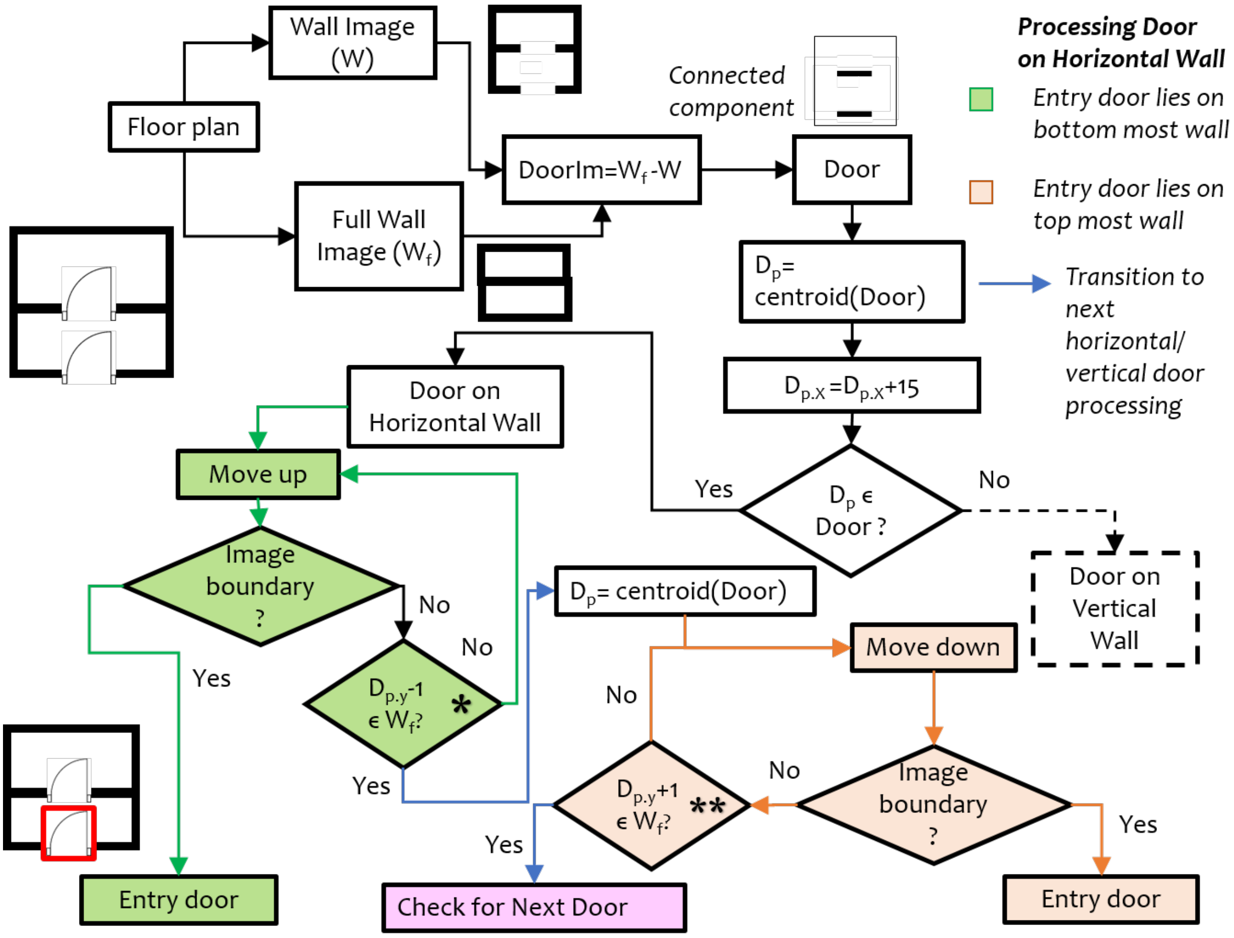
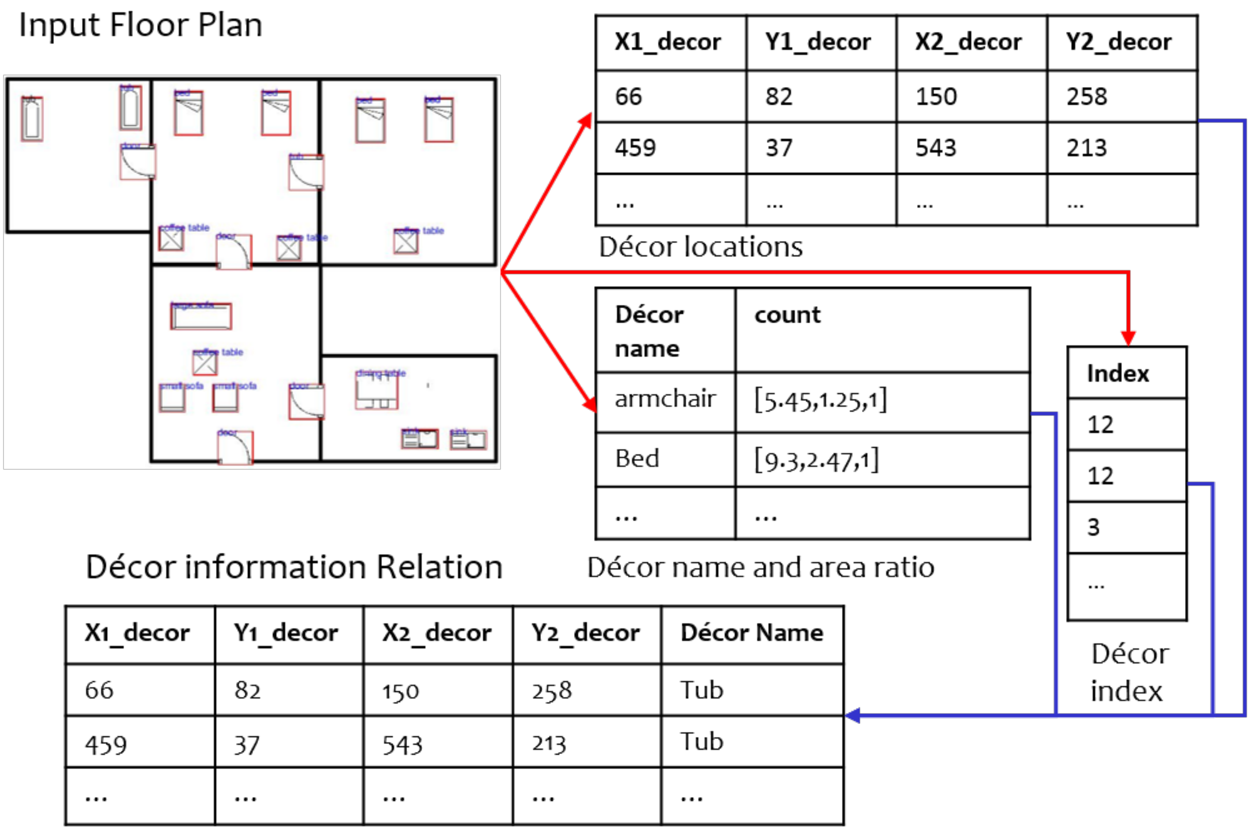


Figure 4.5 : Flow chart for entry door detection in a floor plan image.



X1_decor	Y1_decor	X2_decor	Y2_decor
66	82	150	258
459	37	543	213
...	...	...	...

Décor locations

Décor name	count
armchair	[5.45,1.25,1]
Bed	[9.3,2.47,1]
...	...

Décor name and area ratio

Index
12
12
3
...

Décor index

Décor information Relation

X1_decor	Y1_decor	X2_decor	Y2_decor	Décor Name
66	82	150	258	Tub
459	37	543	213	Tub
...	...	...	...	...

Figure 4.6 : A schematic representation of room decor characterization and final schema generation.

4.2.5 Neighborhood detection

The door coordinates obtained in the room segmentation stage are in the form of diagonally opposite vertices of their bounding boxes $<X^1_{door}, Y^1_{door}>, <X^2_{door}, Y^2_{door}>$ and are stored in a separate data structure. These coordinates are used for inside-outside test with the room coordinates stored in “Room Quantity Info” schema (see Tab.4.1) of floor plan. A door is found to be shared between two rooms if the corner points are in two different rooms. The door between a a pair of shared room is also labelled with the room name. If at least one of the set of door coordinate do not belong to any of the rooms, then that door exist on the contour of the floor plan. Its opposite coordinate is tested for its belongingness to some room and its name is mapped accordingly. It can be an entry door or any other door opening to the outside of the house. All the shared doors, with their room name are stored in a schema named $doorAdjacency$ with attributes $< Door_{From}, Door_{To} >$. If a door is opening to outside then null is assigned with its shared room. Here null represent the outside area, which do not belong to the floor plan. If a pair of rooms shares a door, then they are considered to be neighbors. The neighborhood information is stored in ”Room Qualitative Info” schema (Tab. 4.3). A door contained between two rooms is said to be opening from the room in $Door_{From}$ to the room in $Door_{to}$.

Table 4.3 : Room Quality Info schema for the Example in Fig. 4.4. Decor details: Coffee table (CT), Bed (B), Small Sofa (s), Large Sofa (S),Dining Table (D), Sink (K), Tub (T).

Name	Label	Size	Decors	Front	Back	Left	Right	Neighbors
BT	1	S	T, T	C	B	B	W	BD
BD	2	M	CT, B, B, CT	C	C	W	W	EN, BD, BT
EN	3	M	s, S, CT, s	C	C	B	W	***Null***, BD, KN
BD	4	M	B, CT, B	C	B	W	B	BD
KN	5	S	D, K, K	C	W	W	W	EN

4.2.6 Room Decor Characterization

For room decor classification, each decor image is cropped from the input floor plan image. Connected component analysis is performed over it and area of its each component is evaluated. Top 3 areas are picked from a list of non-increasingly sorted list of areas of component. The decor signature is obtained by dividing all the values with the smallest among three to obtain an area ratio. All the decors with their signatures are compared with the values of signatures stored previously in a database. The one with the smallest distance is assigned to the current decor. We store all the information as shown in the Fig. 4.6 and map them with room information present.

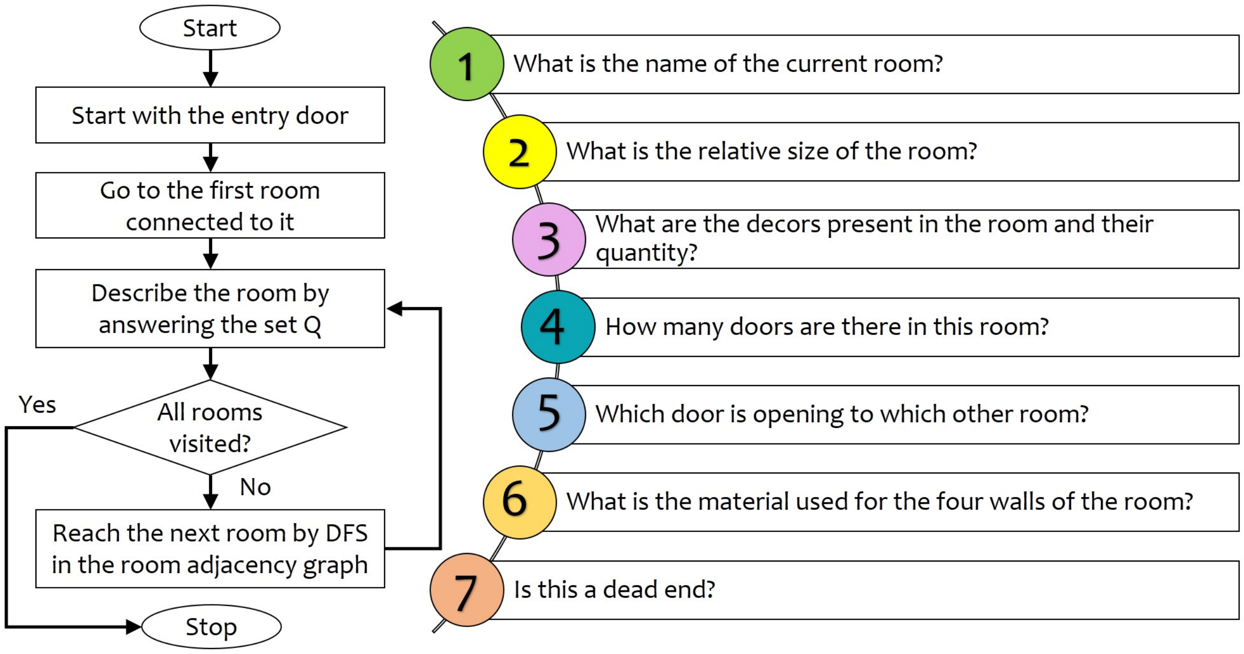
The final "Room Quality Info" schema (Tab. 4.3) stores all the necessary information needed for generating egocentric vision based narration. At each step, it keeps updating and records all the qualitative information of the floor plan image generated at every intermediate step. In this schema, ***Label*** acts as primary key and used to map information with other schemas generated.

4.3 DESCRIPTION SYNTHESIS

The description is synthesized according to the flowchart given in Fig. 4.7(a). An adjacency graph for the floor plan image is generated having each room as a node, labelled as ***Label***. To generate narration, we traverse every room in the order/path generated by Depth first search (DFS) of the adjacency graph, taking ENTRY node as root and ask a set of questions ***Q*** of the room adjacency graph, given in Fig. 4.7(b). Upon arriving in each room we use "Room Quality Info" schema (Tab.4.3) to answer the questions.

4.4 EXPERIMENTS AND RESULTS

We performed our experiment on the ROBIN dataset proposed in Sharma *et al.* [2017] to show our proposed method's effectiveness. The dataset has 500 floor plan samples, spreading over 3 classes. We have augmented the dataset by manually superimposing textures, adopted from Kilmer and Kilmer [2009], on the walls to simulate various construction materials, such as wood, brick, and concrete. Results are shown in Fig. 4.8, where our proposed method generates an egocentric narration of three different floor plan images. The narration contains all the information of the house w.r.to the eye of a person. Features like wall materials, room decors, relative size, and dead-end were detected accurately as they view it when s/he enters into the room. In Fig. 4.8 (a),

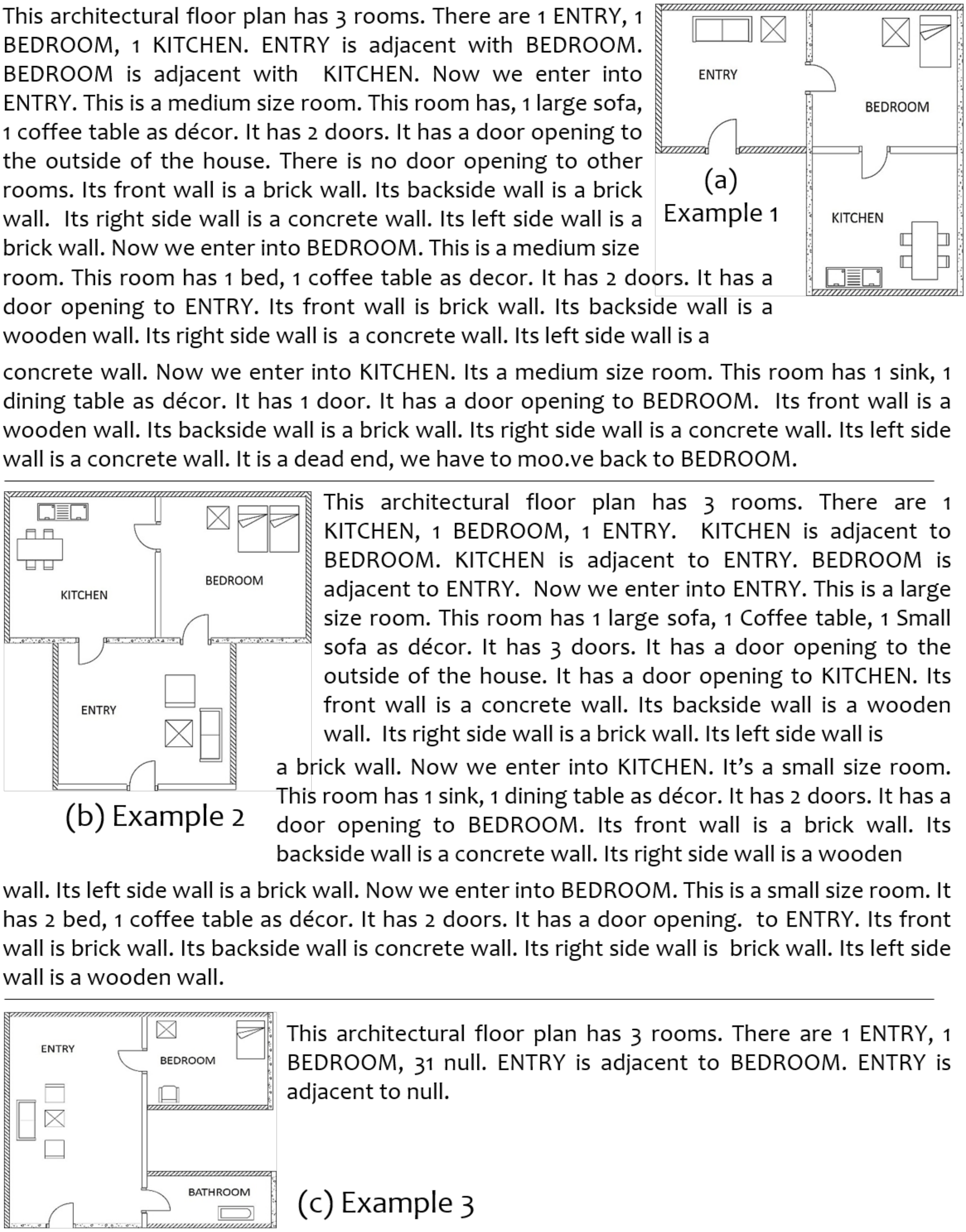

Start
Start with the entry door
Go to the first room connected to it
Describe the room by answering the set Q
All rooms visited?
Yes
No
Reach the next room by DFS in the room adjacency graph
Stop
1 What is the name of the current room?
2 What is the relative size of the room?
3 What are the decors present in the room and their quantity?
4 How many doors are there in this room?
5 Which door is opening to which other room?
6 What is the material used for the four walls of the room?
7 Is this a dead end?

(a) Flow chart for Narration Synthesis

(b) Question set Q to be answered for each room

Figure 4.7 : A flow chart for egocentric narration synthesis and set of questions to extract salient features of a room.

(b), the narration is generated correctly. It is giving correct answer of all the questions (mentioned in Fig. 4.7) asked while traversing each room. It is able to detect the material used for all the walls correctly, for both example 1 and 2 in Fig. 4.8. Also, it can classify decor, identify neighbors, and generate descriptions accordingly. However, in Fig. 4.8 (c), the generated narration is not complete since the system failed to recognize and classify the texture of the walls belonging to each room correctly and could not proceed further. The word "null" in Fig. 4.8 (c) refers to failure of text recognition. Parameters and thresholds, which are used in Algorithm 1 and in entry detection are all empirically obtained through sensitivity analysis.



This architectural floor plan has 3 rooms. There are 1 ENTRY, 1 BEDROOM, 1 KITCHEN. ENTRY is adjacent with BEDROOM. BEDROOM is adjacent with KITCHEN. Now we enter into ENTRY. This is a medium size room. This room has, 1 large sofa, 1 coffee table as décor. It has 2 doors. It has a door opening to the outside of the house. There is no door opening to other rooms. Its front wall is a brick wall. Its backside wall is a brick wall. Its right side wall is a concrete wall. Its left side wall is a brick wall. Now we enter into BEDROOM. This is a medium size room. This room has 1 bed, 1 coffee table as decor. It has 2 doors. It has a door opening to ENTRY. Its front wall is brick wall. Its backside wall is a wooden wall. Its right side wall is a concrete wall. Its left side wall is a concrete wall. Now we enter into KITCHEN. Its a medium size room. This room has 1 sink, 1 dining table as décor. It has 1 door. It has a door opening to BEDROOM. Its front wall is a wooden wall. Its backside wall is a brick wall. Its right side wall is a concrete wall. Its left side wall is a concrete wall. It is a dead end, we have to moo.ve back to BEDROOM.

(a) Example 1

This architectural floor plan has 3 rooms. There are 1 KITCHEN, 1 BEDROOM, 1 ENTRY. KITCHEN is adjacent to BEDROOM. KITCHEN is adjacent to ENTRY. BEDROOM is adjacent to ENTRY. Now we enter into ENTRY. This is a large size room. This room has 1 large sofa, 1 Coffee table, 1 Small sofa as décor. It has 3 doors. It has a door opening to the outside of the house. It has a door opening to KITCHEN. Its front wall is a concrete wall. Its backside wall is a wooden wall. Its right side wall is a brick wall. Its left side wall is a brick wall. Now we enter into KITCHEN. It's a small size room. This room has 1 sink, 1 dining table as décor. It has 2 doors. It has a door opening to BEDROOM. Its front wall is a brick wall. Its backside wall is a concrete wall. Its right side wall is a wooden wall. Its left side wall is a brick wall. Now we enter into BEDROOM. This is a small size room. It has 2 bed, 1 coffee table as décor. It has 2 doors. It has a door opening. to ENTRY. Its front wall is brick wall. Its backside wall is concrete wall. Its right side wall is brick wall. Its left side wall is a wooden wall.

(b) Example 2

This architectural floor plan has 3 rooms. There are 1 ENTRY, 1 BEDROOM, 31 null. ENTRY is adjacent to BEDROOM. ENTRY is adjacent to null.

(c) Example 3

Figure 4.8 : An illustration of the results of egocentric narration synthesis based on the proposed model.

Table 4.4 : Quantitative comparison of wall material characterization.

Method	Precision	Recall	Accuracy
SLIC	0.72	0.75	97.90
Ours	**0.90**	**0.93**	**99.50**

Table 4.5 : Quantititative results of quality of text synthesis.

Method	Precision	Recall	F-score
ROUGE-1	0.40	0.66	0.50
ROUGE-2	0.18	0.31	0.23
ROUGE-3	0.10	0.48	0.12

Among the 500 floor plans, we achieved correct results on 383 images, with an accuracy of 76.6%. Our algorithm was able to answer the 7 questions correctly for these 383 images. For 53 images our algorithm could answer only 6 questions, for 12 images, only 5 questions were answered. Remaining 52 images were failure cases due to incorrect characterization of wall material and incorrect labelling of rooms. The average time noted over the entire data set to complete execution was 30.52 seconds with Intel core i5 2.53 GHz processor (m460) and 8 GB RAM. For material segmentation, we compared our algorithm with Simple Linear Iterative Clustering (SLIC) as proposed in Achanta *et al.* [2012]. Comparative results are shown in Tab. 4.4 to show that our algorithm (in bold letters) for material segmentation outperformed SLIC. Since ours is the first attempt to this problem, we could not compare the narration synthesis part of our work with existing techniques. However, to quantify the quality of the description synthesis, we have computed the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) proposed by Lin [2004] score. For ROUGE evaluation, we invite volunteers to write descriptions of the floor plan image. Then we compute the ROUGE score by taking average of the corpus level score. ROUGE score (Tab. 4.5) show that the generated descriptions are concise in nature.

4.5 SUMMARY

In this chapter, we propose a novel framework for egocentric vision-based narration synthesis of a building floor plan image. We have also proposed a technique to identify various materials used to build different walls in a floor plan. These properties are helpful in studying the quality of the architecture of the house. We have performed experiments on real-world floor plan images with qualitative and quantitative analysis. In the next chapter, machine learning based approach for generating textual description from floor plan images is discussed in detail, while presenting novel feature descriptors BoD and LOFD for floor plan images.