A preliminary analysis of the process on floor plan recognition

Towards discovery of human's recognition mechanisms for complex structured images

Kenta Kamihori, Tomoki Shimano, Atsuto Oguro and Tomonobu Ozaki[†]

Department of Information Science, Nihon University,
Sakurajosui 3-25-40, Setagaya-Ku, Tokyo 156-8550, JAPAN
[†]tozaki@chs.nihon-u.ac.jp

Human experts can recognize a complex structured images easily by using appropriate domain knowledge. The ultimate goal of this research is to explore the fountainhead of this ability and to identify the human's recognition mechanisms for complex images. As a preliminary case study towards the ultimate goal, we focus on images of architectural floor plan in this research. By analyzing the errors occurred in the process of a certain task on floor plan recognition, we attempt to obtain significant insights and meaningful information on complex image recognition and processing by human.

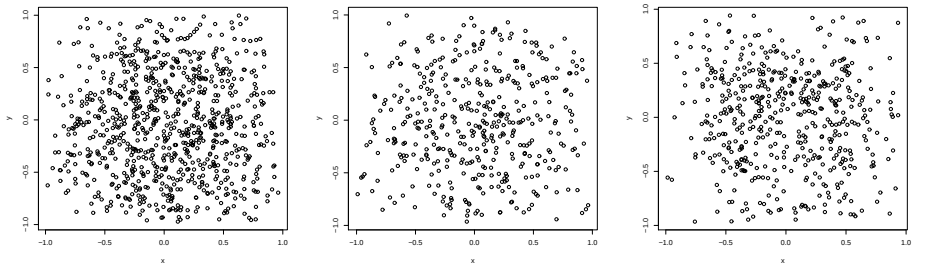
An architectural floor plan represents a whole layout of rooms. It also usually contains doors, windows and some graphical symbols representing sinks in a kitchen, wash stands, bathtubs, toilet seats and space for washing machines, etc. In addition, abbreviations of words, e.g. WIC for walk-in closet, LDK for area of living-dining kitchen, show the additional information of rooms and spaces. Since a floor plan image has rich information with a reasonable semantics as shown above, not only experts but also ordinary people can recognize the structure of the rooms and imagine the usability of the rooms easily. Due to the above reasons, we believe that floor plan images are suitable for an initial case study towards discovery of human's recognition mechanisms.

In this research, a certain task of floor plan recognition is prepared to capture the rough recognition process. In the task, given a floor plan image, subjects identify all rooms and their connectivity. Subjects are also required to specify the types of rooms such as living, kitchen and bathroom, as well as the types of the connection such as by door, by space and by wall. In other words, the objective of the task is to convert an image of floor plan to a labeled graph which consists of nodes representing rooms and edges for the connectivity.

We develop a GUI software to draw a graph on a given image by using mouse clicks and to record two kinds of log data during the task execution. The first one consists of the process of graph construction. We record the order of events on creation and deletion of nodes and edges, changing the label, and so on. The second log data contains physical conditions of subjects including head movements, eye movements, and blinks using wearable smart glasses.

As a preliminary experiment, we collect log data by three subjects using 642 of real floor plan images with size 578 x 618 pixels on average having at least

Subject A	Subject B	Subject C
Relative positions of errors in images		



example of errors with high conditional probability P(wrong label | correct label) (represented in the form of "correct label → wrong label")		
living room → LD	bathroom→bath	LK→living room
LDK→ LD	western style room→receipt	LD→LDK
LD→LDK	japanese style room→entrance	LDK→LD
kitchen→receipt	LD→living room	estroom→bath
door→wall	entrance→corridor	door→wall
bathroom→bath	bath→bathroom	entrance→corridor
Veranda→dining	kitchen→receipt	japanese style room→enterance

Fig. 1. Relative position of errors and rules with high conditional probability

three rooms. The average number of living rooms is 6.28 with standard deviation $s = 1.18$. As a result, three subjects draw 1284 of graphs in total having 27.2 vertices ($s = 7.80$) and 42.0 edges ($s = 15.19$) on average.

In this paper, we divide errors during the task execution into two kinds, noticed and not noticed by subjects, and conduct simple preliminary analysis of the former only.

The task might be difficult for the subjects and 73.8, 57.7 and 65.4 percent of graphs constructed by three subjects contained at least one error, respectively. We plot relative positions where errors occur in Fig.1. Since we cannot observe any particular tendency in the plots, we conclude that the position has no relationship on the error. In addition, we confirmed that the errors have no strong relationship with fatigue by using correlation measures on the duration time of tasks as well as the number of blinks. Some examples of errors having high conditional probability are shown in Fig.1. We can recognize that three subjects make similar errors in common.

The simple analysis of noticed errors is not sufficient for deriving strongly meaningful results. For future work, we plan to investigate on the deep analysis of the whole errors including not noticed ones by subjects for the discovery of recognition mechanisms of complex images.

Acknowledgement: In this paper, we used floor plan image dataset provided by NEXT Co., Ltd. and the National Institute of Informatics.