Extended Abstract for Poster Submission

Automated natural language description of indoor spaces

Saranya Kesavan and Nicholas A. Giudice*
Spatial Informatics program: School of Computing and Information Science
The University of Maine, Orono ME, USA.

Abstract: This research aims to investigate techniques for developing 3-dimensional cognitive models of indoor spaces through the use of natural language (NL) verbal descriptions of the environment. Our goal is to provide automated descriptions to blind/low-vision people, in order to provide a non-visual means of learning and navigating unfamiliar indoor spaces. These NL descriptions will be developed based on the spatial analysis of a 2-dimensional representation (i.e. photograph) of the space, which could be taken via the user's smart-phone. The metric and abstract values that are obtained as a result of spatial analysis will be systematically converted into NL descriptions by developing unique ontologies for generic indoor spaces and then applying formalized linguistic semantics. As a first step, the main objective of this preliminary research is to conduct behavioral experiments to better characterize the meanings of spatial prepositions and optimally structure the existing information through flexible and extensible description logic.

1 Motivation

Navigation involves a process of controlling and monitoring the movement of any physical entity from one place to another [1]. Most of the navigation systems that are currently employed target only outdoor, large scale environments and they are developed mainly for sighted users. For blind users, traditional canes and guide dogs provide mobility information (e.g., about obstructions to the path of travel) but provide little information about the spatial location of objects that are not in the path of travel or the configuration of these objects. Cognitive maps are often only considered for large-scale environments, but they also are relevant for knowing the global relations of objects in a room e.g. identifying objects in the space, knowing their orientation, their relative distances and their relative positions [2]. Without perceptual access to this information, blind users often rely on effortful cognitive strategies to learn and navigate their surrounds and are frequently dependent on others to get information about their target destination. This research aims to facilitate the development of cognitive maps of indoor spaces for blind users based on automatically generated natural language (NL) descriptions of the space. The information about a specific indoor space can be obtained from a 2D representation through direct visual observation or from a photographic observation. This research discusses a preliminary study that compares these two observation modes in order to characterize the extent and structure of information described.

2 Background

Spatial analysis techniques on photographs of indoor spaces will provide us with information about the type of objects, their spatial locations, and the global arrangement of these objects in space [5]. However, these details, if presented in an unorganized way, will not help a blind person to understand the space in a manner that supports accurate learning and navigation. Hence, we have to convert the photographic information to a meaningful and presentable format. Thus, this work plans to identify the factors that account for "the meaningful and presentable format", based on behavioral experiments that compare photographic and direct observations. The research also aims to automate the entire process of generating a NL description of an indoor space from a photograph of that space. This is possible when we preserve the information collected from the spatial analysis in a definite sequential and hierarchical structure [3, 4] (i.e. in terms of ontological structures).

*This work is funded by the National Science Foundation.

3 Research Goals:

Description strategies determine the order in which objects, in an indoor space, are verbally described. As there are many different strategies, determining the most effective strategy that creates a relatively accurate cognitive map is an important starting point and is a primary research goal in this study. The relationship between the spatial location of objects, their spatial representation, semantic representation of quantified spatial prepositions, and relationship between formal and linguistic semantics has to be stored in a database with hierarchical relationships. Another research goal of this project is to develop unique indoor space ontologies for generic indoor spaces based on these hierarchical relationships. Finally, this research aims to find a best practice approach for generating spatial descriptions automatically by retrieving appropriate spatial relationships from ontological relationships. This process could then be used to arrange framed sentences in an order that would help in generating an effective cognitive map.

4 Experiments and Hypotheses

Behavioral experiments are being conducted to understand the ways in which humans describe the spatial information of rooms when viewed from a photograph versus viewing the physical room. The hypotheses are as follows:

1. The verbal descriptions generated from realtime observations of indoor environments will lead to more accurate cognitive maps than verbal descriptions generated from observing pictures of the same environments.
2. The verbal descriptions that were developed based on a consistent description strategy and consistent lexical selection will aid spatial learning and the resultant cognitive map built up from these descriptions will be more accurate than from descriptions that were developed in the absence of consistent description strategy or lexical selection.
3. The cognitive map developed based on the verbal description of an indoor scene will be equally accurate as when developed from visual apprehension of that scene.

5 Expected Results

Behavioral experiments are currently being conducted based on the hypotheses stated above. From early pilot data, the accuracy of cognitive maps derived based on verbal descriptions that are generated from real time observations of indoor environments are the same as those generated from photographic observations. Larger scale experiments are now underway to further test this finding as well as the importance of description strategies and lexical selection. Once the Natural language generation is complete based on the principles of description logic, visual apprehension of indoor spaces will be compared with the Natural language description of indoor spaces in terms of Cognitive map generation.

References

1. Bowditch,N(1802). The American Practical Navigator. Defense Mapping Agency Hydrographic Topographic Center.

2.Lahav,O&Mioduser, D(2004).Exploration of Unknown Spaces by People Who Are Blind Using a Multi-sensory Virtual Environment. (2004). Journal of Special Education Technology, 19(3), 15-23.

3. Hois, J., Bhatt, M., Kutz, O., (2009). Modular Ontologies for Architectural Design. Cognition, 66-77.

4. Schütte, N. (2009). Generating natural language descriptions of ontology concepts. Proceedings of the 12th European Workshop on Natural Language Generation - ENLG ’09, (March), 106-109.

5. Zwarts,J&Winter,Y. 2000. Vector Space Semantics: a model-theoreticanalysis of locative prepositions. Journal of Logic, Language and Information,9:169–211.