HUMAN VISION SIMULATION IN THE BUILT ENVIRONMENT

Qunli Chen, Chengyu Sun
DDSS Group, Faculty of Architecture, Bulding and Planning, Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Bauke de Vries
DDSS Group, Faculty of Architecture, Bulding and Planning, Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Keywords: Architectural cue recognition, Human vision simulation, Visual perception, Built environment.

Abstract: This paper first presents a brief review on visual perception in the built environment and the Standard Feature Model of visual cortex (SFM); following experiments are presented for architectural cue recognition (door, wall and doorway) using SFM feature-based model. Based on the findings of these experiments, we conclude that the visual differences between architectural cues are too subtle to realistically simulate human vision for the SFM.

1 INTRODUCTION

In our daily life, we evaluate the building and the built environment from two different aspects, namely the structure of the building and the human behavior in the building. In the past research was focused on the structure of the building. However, as the development of the new technologies, nowadays the buildings can be designed as to meet different people's requirements. Therefore we should turn to the other evaluator, the behavior of human beings in the building or built environment. It is important to develop a dynamic model in which we can simulate human behavior with the help of agent-based systems in a virtual built environment. Unlike most of the previous vision and visual interpretation researches, the vision and actions of the agent in our project will be determined by applying human visual perception simulation based on real physical perception occurring in human brains.

As the first step we propose to link the architectural cues and human vision simulation for the development of an architectural cue recognition system. After a careful research and consideration, we select the Standard Feature Model of visual cortex (SFM) for architectural cure recognition.

2 VISUAL PERCEPTION IN BUILT ENVIRONMENT

Visual perception is always regarded as the most important type of human perceptions in the built environment by architects (Crosby, 1997) and environmental psychologist (Arthur & Passini, 1992). In this article, we discuss the visual perception as the background of human perception simulation in the built environment.

The research of human beings' visual perception in the built environment starts from Gibson's Affordance theory (Gibson, 1966), which explains how an object in the built environment is (re)cognized as an object notion and its potential usage in the brain. He interprets the built environment as a set of various affordances, which are the units of a human being's (re)cognition. With a motivation, the visual stimuli, the light pattern of an object, is (re)cognized as a visual affordance according to the "schemata" of this object notion in the brain.

Based on Gibson's theory, Hershberger (Hershberger, 1974) develops the mediational theory of environmental meaning. He splits the notion of schemata into two kinds of knowledge explicitly. One is the linkage from the light pattern of the

Chen Q., SUN C. and de VRIES B. (2009).
HUMAN VISION SIMULATION IN THE BUILT ENVIRONMENT.
In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 385-388
DOI: 10.5220/0001768103850388

object to the object notion in the brain. The other is the linkage from the object notion to its potential usage.

With more biological support, Lam (Lam, 1992) explains the visual perception as an active information-seeking process directed and interpreted by the brain, which can be explained in four transitions: T1, T2, T3, and T4 between five concepts: Object, Light Pattern, Electric Signals, Object Notion, and Cue (Figure.1).

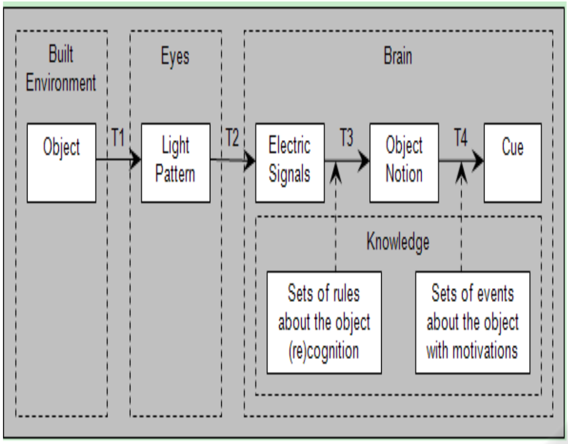

Figure 1: Visual perception process.

Through the above process, some objects in built environment are selectively perceived by human beings as cues for the specific motivation. In the following introduction on the human visual perception simulation, the object, the transition T1, the light pattern, the transition T2, and the electric signals are compacted as an image input in pixel grid. The simulation focuses on the transition T3, namely from the pixel grid images to object notions.

3 STANDARD FEATURE MODEL

In recent years, some researchers turned back to look at the object recognition problem from the biology science side, and obtained very good results. Among them, Serre developed a hierarchical system which can be used for the recognition of complex visual scenes, called SFM (Standard Feature Model of visual cortex) (Serre et al., 2004, 2007). The system is motivated by a number of models of the visual cortex. Earlier, object recognition models aimed at improving the efficiency of the algorithms, optimize the representation of the object or the object category. Not much attention was focused on the biological features for higher complexity, not to mention applying the neurobiological models of object recognition to deal with real-world images.

The SFM model follows a theory of the feed forward path of object recognition in the cortex, which accounts for the first 100-200 milliseconds of processing in the ventral stream of the primate visual cortex (Riesenhuber et al., 1999); (Serre et al., 2005). The SFM model tries to summarize what most of the visual neuroscientists agree on: firstly, the first part of visual processing of information in the primate cortex follows a feed-forward way. Secondly, the whole visual processing is hierarchical. Thirdly, along this hierarchy the receptive fields of the neurons will increase while the complexity of their optimal stimuli will increase as well. Last but not least, the modification and the learning of the object categories can happen at all stages.

In the SFM model there are four layers, each containing one kind of computational units. There are two kinds of computational units, namely S (simple) units and C (complex) units. The function of the S unit is to combine the input stimuli with Gaussian-like tuning as to increase object selectivity and variance while the C unit aims to introduce invariance to scale and translation. We simply call the four layers as S_1, C_1, S_2, and C_2. A brief description of the functions, input and output to the four layers are listed in Table 1.

This biological motivated object recognition system has been proven to be able to learn from few examples and give a good performance. Moreover, this generic approach can be used for scene understanding. Last but not least, the features generated by the model can work with standard computer vision techniques; furthermore it can be used as a supporting tool to improve the performance of those computer vision techniques.

4 EXPERIMENT AND FINDING

There are different cues in the built environment, which can be divided into three groups, namely non-fixed cues, semi-fixed cues and fixed cues. Non-fixed cues are defined as a type of information perceived from the dynamic objects. Objects like maps, signage, and different decorations are semi-fixed cues, and the architectural cues are fixed cues.

We conducted two experiments applying the SFM model to training sets for architectural cues recognition. Sets of images of scenes containing architectural cues are used as the training examples.

Table 1: Brief description of the four layers of the SFM model.

	Layer number	Brief actions	Input	Output
S_1	1	Apply Gabor filters to the input, obtain maps of orientations and scales	Gray-scale images	maps of different positions, scales, and orientations
C_1	2	Use a max operation, obtain the position invariant features for each band	Bands of maps from S_1	Position invariant features for each band (C_1 features)
S_2	3	Pool C_1 features in patches (different scales in the same orientation)	Image patches at all positions from C_1	S_2 patches
C_2	4	Combine a max operation and the S_2 patches, find the scale invariant features	S_2 patches	Position & scales invariant features (C_2 features)

In the first experiments, flat-shaded images of simple scenes consisting of a room and a doorway were used as training images. We aimed to test whether the SFM features are sufficient to distinct the difference between the walls and the doorways. Images containing a doorway are used as positive examples while images containing walls as negative examples. A good result was achieved using this set of images (higher than 90%).

In the second set of training images, the test scenes where rendered using different materials and shadows, to add an additional level of realism. A light source is fixed in the middle of the ceiling for each of the room. To standardize the examples, all the doorways have similar width and height. Some samples for the second set of the training images are shown in Figure 2. The first two rows show the examples of the doorway (the doorway can be designed to the left of the wall, or to the right of the wall), with the last row showing examples of the wall. Table 2 presents the results of the first experiment with two sets of training examples of standardized input.

Furthermore, we explored the effect of other parameters: we find that the width of the doorway, the shape of the room and the proportion of the width of the doorway and the room can affect the distinction.

Table 2: Results of the first experiment.

	Input samples	Classification
1	Single room with doorway	>90%
2	Assigned with materials and light source	>80%

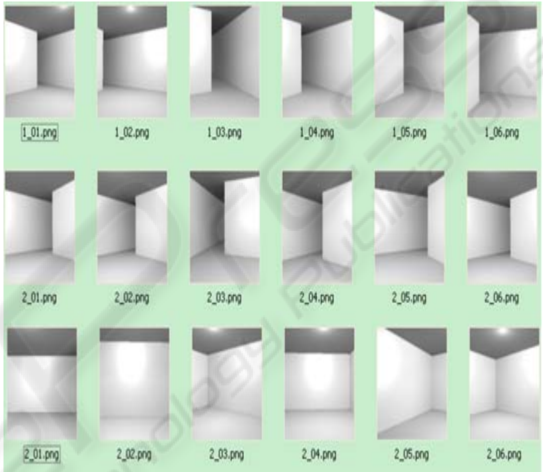

Figure 2: Samples for the first experiment.

In the second experiment, an additional simple building element was added to the scenes: next to the doorways from the first experiment, the test scenes now also contained doors. We aim to test the distinction between the door and the wall, the doorway and the wall, the door or doorway and the wall, finally the door and the doorway. Samples for this experiment of the training images are shown in Figure 3. In Figure 3, three types of architectural cues are shown in rows sequentially, namely door, doorway and wall. We chose images of door as positive example and images of wall as negative examples when we aim to use SFM to test the distinction between door and wall.

The results of this second experiment were a lower percentage of successful recognitions. Table 3 shows the samples and the results. It is quite clear from this table that the SFM gives good distinction between door and wall, door or doorway and wall, and doorway and wall. However it cannot find efficient distinction for the recognition between door and doorway; in other word the distinctions between the features extracted for door and doorway are too subtle for efficient recognition.

Table 3: Results of the second experiment.

	Distinction	Classification
1	Door & Wall	91.66%
2	Doorway & Wall	83.33%
3	Door or Doorway & Wall	95.75%
4	Door & Doorway	66.67%

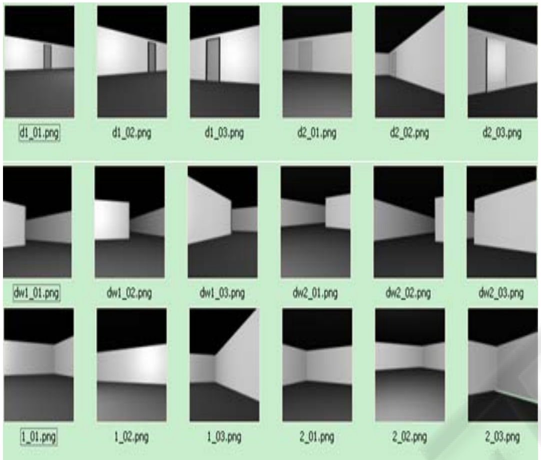

Figure 3: Samples for the second experiment.

5 DISCUSSION AND FURTHER WORK

Based on the findings, we conclude that a limited recognition can be achieved using SFM-based features for architectural cue recognition. Our interpretation of these results is that the visual difference between doors and doorways are too subtle to serve as a significant discrimination factor for the SFM.

Our aim is to simulate real human vision including its limitations. The SFM model deviates considerably from real human performance. Therefore we will research a probabilistic approach that allows us to estimate vision variables for object recognition from experiments with real humans. We hope to report this in the near further.

REFERENCES

Arthur, P., & Passini, R., 1992. Wayfinding: People, Signs, and Architecture. New York: McGraw-Hill Book Company.

Burl, M., Weber, M., & Perona, P., 1998. A probabilistic approach to object recognition using local photometry and global geometry. Proc. ECCV (pp. 628-641).

Crosby, A. W., 1997. The measure of reality: Quantification and Western society 1250-1600. Cambridge: University Press.

Gibson, J. J., 1966. The senses considered as perceptual systems. Boston: Houghton Mifflin.

Hershberger, R. G., 1974. Predicting the meaning of architecture, in Jon Lang et al. (Eds.), Designing for Human Behavior: Architecture and the Behavioral Sciences (pp. 147-156). Stroudsburg: Dowden, Hutchinson and Ross.

Lam, W. M. C., 1992. Perception & Lighting as formgivers for architecture. New York: Van Nostrand Reinhold.

Riesenhuber, M., & Poggio, T., 1999. Hierarchical Models of Object Recognition in Cortex, Nature Neuroscience, vol. 2, no. 11 (pp.1019-1025).

Serre, T., & Riesenhuber, M., 2004. Realistic Modeling of Simple and Complex Cell Tuning in the HMAX Model, and Implications for Invariant Object Recognition in Cortex," CBCL Paper 239/AI Memo, Massachusetts Inst. of Technology, Cambridge.

Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., & Poggio, T., 2005. A Theory of Object Recognition: Computations and Circuits in the Feed forward Path of the Ventral Stream in Primate Visual Cortex. AI Memo/CBCL Memo.

Serre, T., Wolf, L., & Poggio, T., 2004. A new biologically motivated framework for robust object recognition, CBCL Paper #243/AI Memo #2004-026, Massachusetts Institute of Technology, Cambridge, MA.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T., 2007. Robust Object Recognition with Cortex-Like Mechanisms, IEEE Transactions on pattern analysis and machine intelligence (pp.411-426).