

Canonical views of scenes depend on the shape of the space

Krista A. Ehinger, Ken M. Haggerty, & Aude Oliva

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

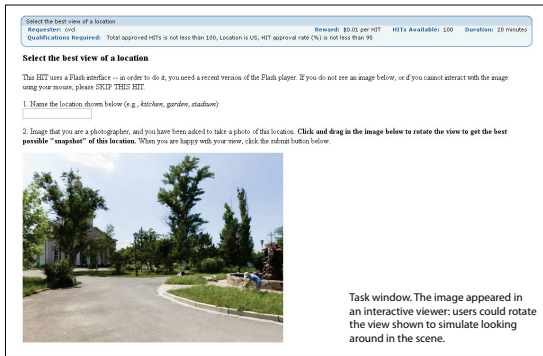
Introduction

When recognizing, depicting, or imagining objects, people show a preference for “canonical” views (Palmer, Rosch, & Chase, 1981). Canonical views of objects are generally views which maximize the number of visible surfaces and avoid self-occlusion. Are there canonical views of scenes? Is the canonical view determined by the shape or function (ie, navigational constraints) of the space?

Experiment

195 workers participated in online task on Amazon Mechanical Turk Stimuli were 624 panoramic photos, each shown to 10 different workers On each trial, workers performed two tasks:

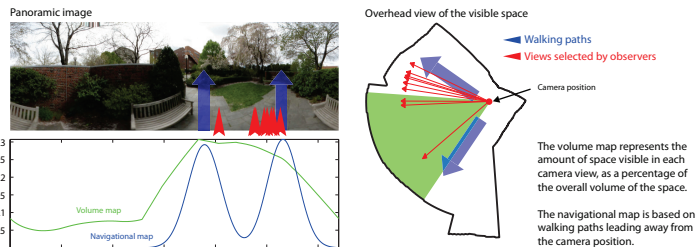
1. Name the location shown in the image (eg, “classroom”)
2. Rotate the image in a 360-degree viewer to show the “best view” of the location



Trials were dropped if the worker did not name the location (1% of trials) or did not use the panoramic viewer (3% of trials)

Modeling the shape of the space

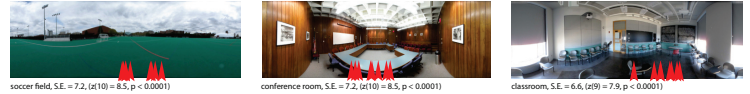
The boundaries of the space were obtained by outlining the ground plane and calculating the volume around the camera. Navigational paths were marked by Mechanical Turk workers.



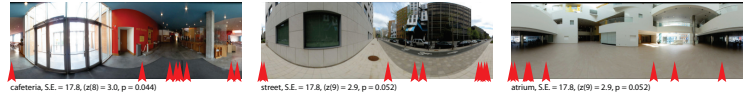
Results

Agreement on the “best view” was high for most scenes (Rayleigh’s test of nonuniformity gave $p < .01$ for 389 scenes (62%), $p < .05$ for 466 scenes (75%).

Examples of scenes at the 25th percentile of agreement:



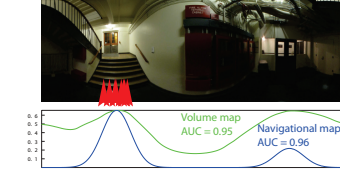
Examples of scenes at the 75th percentile of agreement:



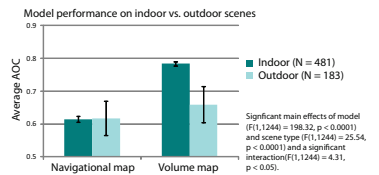
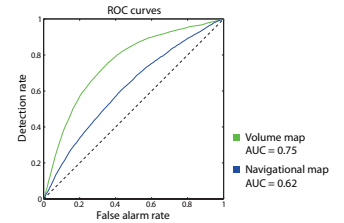
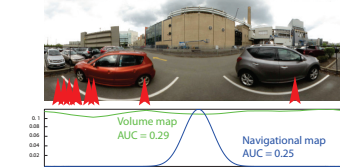
Agreement was higher in indoor scenes than in outdoor scenes ($t(245) = 6.12$, $p < 0.0001$), and was correlated with high range of visible depths (0.40), smaller volume (0.30), and high scene name agreement (0.44).

The volume map was the best predictor of the views chosen by observers. The navigational map performed above chance, but did not contribute any independent predictive power.

Example of a scene where both models performed very well



Example of a scene where both models performed very poorly



Conclusion

Observers generally agree on the “best view” of a scene. The best view of a scene is the one that shows as much of the space as possible, not necessarily the functional view for walking in that space. This is similar to findings with objects: eg, people do not prefer views of objects oriented for grasping (Blanz, Tarr, & Bülthoff, 1999).

References

Blanz, V., Tarr, M. J., & Bülthoff, H. H. (1999). What object attributes determine canonical views? Perception, 28(5), 575-599.
 Palmer, S., Rosch, E., Chase, P., (1981). Canonical perspective and the perception of 40 objects. In Attention and Performance IX, Ed. J. Long, A. Baddeley (Hillsdale, NJ: Lawrence Erlbaum), pp. 135-151.