Viewpoint selection in active visual search

Poster uploaded here!

Tiffany Wu, John K. Tsotsos
Department of Electrical Engineering and Computer Science, York University, Toronto, Canada
Correspondence: tiffwu@yorku.ca

Introduction

- Viewpoint selection during visual search has not been investigated in detail
- Existing active vision experiments have a lesser focus on visual search, and have not coupled gaze and 6DOF head pose to examine behaviour in an object search scenario[1,2]
- Active sensing is "the problem of *intelligent control strategies* applied to the data acquisition process which will depend on the current state of data interpretation and the goal or task of the process" [3]

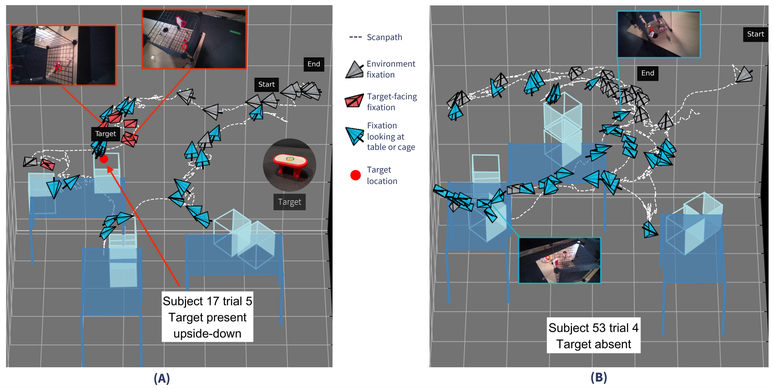
In order to understand visual search in an active environment, we must consider viewpoint selection

Highlights

- **Active visual search in a real-world environment provides a rich source of data for head and eye movements**
- **Gaze location sequences, head pose in 6DOF including crouches and head tilts were recorded, along with target orientation, to determine if object orientation influences head movements**
- **Subjects orient their head to match targets to their canonical orientation**
- **Subjects navigate to places where their viewpoints will be least obstructed to conduct search**

Research Questions

- **How do eye and head movements during active search vary by target presence, visibility, and set size?**
- **How are subject eye and head movements influenced by object placement and orientation?**



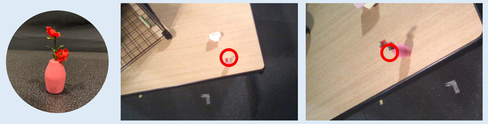
(A) (B)

The above plots show scanpaths of subject trials. The dotted line represents the path their head takes, with each frustum representing a fixation. Grey frusta are fixations looking at the environment, blue frusta are fixations looking at a table or a cage, and red frusta are fixations with the target in the central 30° FOV.

(A) A target present trial where the target was oriented upside-down. The subject sees the target, gets closer, then tilts their head to inspect the object before responding. The blow-ups show first-person views with fixation points, and they also show the tilting the subject performs.
(B) A typical target absent trial. The subject moves towards opening of cages and then looks in them. Two fixations are blown up with the fixation view from the first-person Tobii camera image. In all target absent trials, the subject does not fixate on every single object.

Motivation

- Most existing psychophysics research on visual search are *passive* and conducted on 2D computer screens [4,5]
- 3D Viewpoint selection is difficult to measure and induce on computer screens
- Viewpoint selection is a major component of active visual search strategies, allowing for:
  - Object disocclusion
  - Increasing/decreasing resolution
- Shepard & Metzler's [6] work suggests subjects perform mental rotations in order to complete a same-different task – do they similarly rotate their views when they can choose their viewpoints in a real-world search task?
  - Solbach & Tsotsos [7] conducted the same-different task in 3D, going into much more detail about the real-world counterpart of this study



Target Fixation 20 Fixation 21

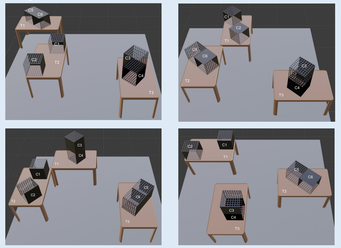
Example of a subject looking for a vase target. They first spotted the target from further, and in the next fixation crouched low and took a different view of the target to confirm.

Methodology

Active visual search task – subjects determine if a specified target is present or not

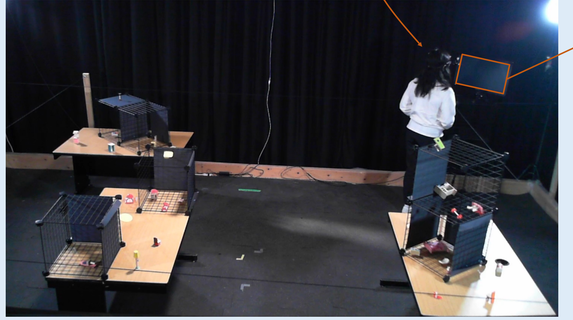
- 12 trials per subject, objects change in each trial

Layouts



Independent variables:
- Target: presence (y/n)
- Target visibility from starting location
- Target orientation
- Set size

Measurements:
- Response time
- Accuracy
- Eye and head movements



PESAO environment setup [8] with tables and cages
Toy furniture used as stimuli

Subject equipped with eye and head trackers

Equipment:
- *Tobii Pro Glasses 2*
- *6 OptiTrack cameras to track reflective markers attached to glasses*

Target image presented for 5 seconds

Subset of possible stimuli: set sizes 30, 40, 50, 60

Results

N=72

- 4 layouts, 6 trial conditions
- 3 subjects in each condition
- 864 trials, 66594 fixations, 12510m of head movement

		RT (s)		# fixations		Head movement (m)		Accuracy (%)	
Target presence		Present	Absent	Present	Absent	Present	Absent	Present	Absent
Correct	*Mean*	21.9	49.1	44	105.5	8.3	20		
	SD	14	25.8	31.2	61.5	5.1	9.3		
	Count	382	407	382	407	382	407	88.9 SD=11	94.2 SD=5.8
Incorrect	*Mean*	47.3	36.3	100.3	73.5	17.3	12.9		
	SD	27.9	24.2	64.1	54.3	7.1	10.4		
	Count	50	25	50	25	50	25		

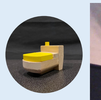
Descriptive stats summary table

Significant main effects of *set size* and *target presence* for:
- Response time
- Number of fixations
- Head movement
- Accuracy
- Number of revisits

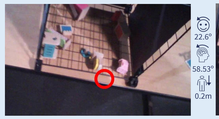
Significant interaction between *set size* and *visibility* for:
- Head movement
- Accuracy
- Number of revisits

Targets divided into orientation categories:

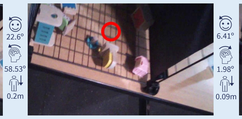
upright (canonical) front-up side-up diagonal upside-down



Target
In trial: side-up



Fixation 51
Roll: -9.26°
Pitch: 58.4°
Crouch: 1.41m



Fixation 53
Roll: -31.96°
Pitch: -0.13°
Crouch: 1.21m

Fixation 54
Roll: -38.37°
Pitch: 1.85°
Crouch: 1.12m

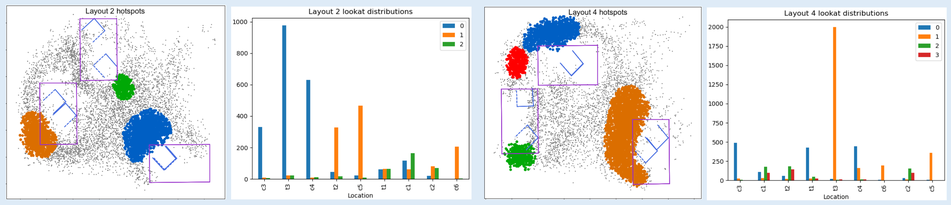
Head tilts and crouches:

roll pitch crouch

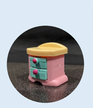
- *Diagonal* objects induced significantly less rolls and pitches than any other orientation
- *Front-up* objects induced significantly fewer pitch movements than *side-up* objects

Head movements in space:

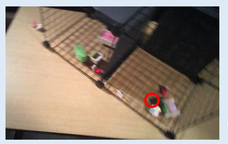
- Subjects tend to spend more time searching facing the openings of cages



Hotspots show subject head locations aggregated over each layout. All major clusters of head locations are around places with cage openings. Bar plots show which locations the subjects are looking at corresponding to each cluster. These line up with the openings of the cages as well as the table they are nearest at the time.



Target
In trial: diagonal



Fixation 7
Roll: -35.6
Pitch: -0.07°
Height: 1.36m

Fixation 9
Roll: -49.9°
Pitch: -7.8°
Height: 1.35m

Implications

- Many head and body movements can be induced by manipulating object placement and orientation
- Subjects tend to take several target views, from differing viewpoints, before confirming a target is present
- Subjects easily identify openings of cages and spend more time standing there, allowing for unobstructed views
- Subjects can identify the correct viewpoint to verify target correctness
- Subjects prefer to orient their views to the target's canonical orientation, rather than performing mental rotation

References

[1] Pelz, J., Hayhoe, M., and Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, 139:266–277.
[2] Hayhoe, M. M., Shrivastava, A., Mruczek, R., and Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, 3(1):6–6.
[3] Bajcsy, R. (1988). Active perception. *Proceedings of the IEEE*, 76(8):966–1005.
[4] Wolfe, J. M. (1998). What can 1 million trials tell us about visual search?. *Psychological Science*, 9(1), 33-39.
[5] Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12(1), 97-136.
[6] Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701-703.
[7] Solbach, M. D., & Tsotsos, J. K. (2023). The psychophysics of human three-dimensional active visuospatial problem-solving. *Scientific Reports*, *13*(1), 19967.
[8] Solbach, M. D., & Tsotsos, J. K. (2020). PESAO: Psychophysical Experimental Setup for Active Observers. *arXiv preprint arXiv:2009.09933.*

YORK UNIVERSITÉ UNIVERSITY
Centre for Vision Research
ICa@ Centre for Innovation in Computing at Lassonde