



Brief article

# The pop out of scene-relative object movement against retinal motion due to self-movement

Simon K. Rushton <sup>a,b,\*</sup>, Mark F. Bradshaw <sup>b</sup>, Paul A. Warren <sup>a</sup>

<sup>a</sup> *School of Psychology, Cardiff University, Tower Building, Park Place, P.O. Box 901, Cardiff CF10 3YG, Wales, UK*

<sup>b</sup> *Department of Psychology, Surrey University, Guildford GU2 7XH, UK*

Received 21 June 2006; revised 6 September 2006; accepted 7 September 2006

---

## Abstract

An object that moves is spotted almost effortlessly; it “pops out”. When the observer is stationary, a moving object is uniquely identified by retinal motion. This is not so when the observer is also moving; as the eye travels through space all scene objects change position relative to the eye producing a complicated field of retinal motion. Without the unique identifier of retinal motion an object moving relative to the scene should be difficult to locate. Using a search task, we investigated this proposition. Computer-rendered objects were moved and transformed in a manner consistent with movement of the observer. Despite the complex pattern of retinal motion, objects moving relative to the scene were found to pop out. We suggest the brain uses its sensitivity to optic flow to “stabilise” the scene, allowing the scene-relative movement of an object to be identified.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Attention; Visual search; Motion; Optic flow; Pop out; Self-movement; Relative motion; Scene-motion; 3D

---

---

\* Corresponding author. Tel.: +44 29 208 70086; fax: +44 29 208 74858.  
*E-mail address:* [rushtonsk@cardiff.ac.uk](mailto:rushtonsk@cardiff.ac.uk) (S.K. Rushton).

## 1. Introduction

If you wish to silently attract the attention of a friend on the other side of the room, you might wave your hand. If that does not work, you may resort to jumping up and down and waving your arms. These actions are normally effective because objects that move are particularly conspicuous and can capture attention (e.g., Abrams & Christ, 2003; Franconeri & Simons, 2003).

A large body of research has examined object movement and visual attention. However, with very few exceptions (see Liu et al., 2005; Morvan & Wexler, 2005; Royden, Wolfe, & Klempen, 2001), it has only considered a special and ecologically rare case, the stationary observer.

When an observer is stationary, an object moving within the scene is uniquely identified by retinal motion and therefore a simple motion detector should suffice to detect its presence. However, during self-movement, every object in the scene changes position relative to the eye and consequently all objects are in motion within the retinal image. Under these circumstances, a simple motion detector would no longer be able to distinguish between scene-stationary and scene-moving objects. Identification of an object moving within the scene requires a different mechanism or strategy.

In a series of preliminary experiments we examined the ability of observers to identify a moving object during simulated head rotation (Rushton & Bradshaw, 2000), forward translation and lateral translation (Rushton & Bradshaw, 2004). Here, we report the results of a study examining what we believe to be the most interesting commonly occurring self-movement: a lateral translation and combined counter rotation of the head. Fig. 1 shows the pattern of retinal motion that results from viewing an array of objects distributed in depth whilst moving in this manner.

One of the objects shown is moving within the scene. If you examine Fig. 1, it is difficult to determine which object is moving within the scene from the white arrows indicating motion. Can the human brain locate the moving object in such a pattern of retinal motion? If so, how does it achieve this? Let us consider some potential solutions.

Irrespective of whether the observer is moving or stationary, at the instant an object begins to move, its retinal trajectory changes; therefore change in trajectory is a cue to scene relative movement. Observers can detect changes in the object trajectory, but performance drops off dramatically at four objects (Tripathy & Barrett, 2004). It follows that change of retinal trajectory is unlikely to be a useful cue to scene relative movement in a natural scene containing a large number of objects.

Another potential solution is based upon change in relative position: as the object moves within the scene, its 3D position relative to other scene objects changes; change of relative 3D position indicates movement within the scene. However, work on change blindness has shown that observers are poor at detecting changes of position of unflagged objects across instances of time (see Rensink, 2002). Furthermore, as the observer's viewpoint changes, some occluded objects will pop into view and this may increase the difficulty of the task as the appearance of objects has been shown to disrupt the detection of change (O'Regan, Rensink, & Clark, 1999). There-

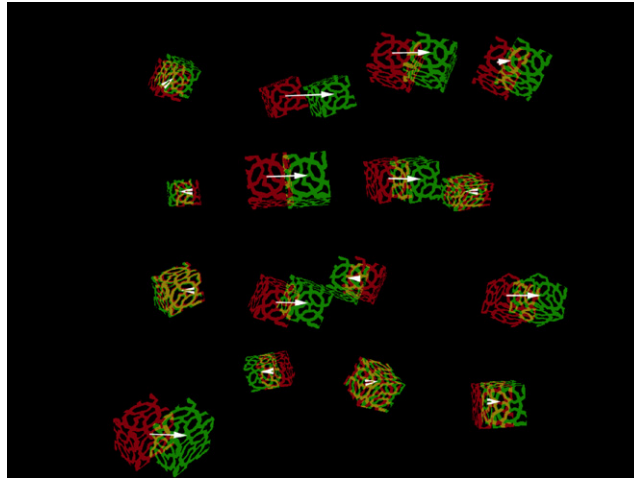


Fig. 1. The pattern of motion that results from a simulated lateral translation of the head and counter rotation of gaze to keep the centre of the array fixated. Images from a single moving viewpoint at two instants of time, the first instant in red, the second in green; arrows indicate the direction of motion of the objects between the two instants; the first instant in red (dark grey), the second in green (light grey). Note the pattern of motion is such that the task cannot be made easier by eye-movements or through search for local 2D velocity differences.

fore a solution based upon detection of a change in relative position is unlikely to be very effective.

An examination of the literature reveals that sometimes the brain can find ways to simplify a search task. For example, it may group objects that share a common feature such as depth (Nakayama & Silverman, 1986) or 2D motion (Duncan, 1995), or “filter” out a subset of objects e.g., use the absence of motion to remove static objects from consideration (McLeod, Driver, & Crisp, 1988). However, in the complex 3D flow patterns illustrated in Fig. 1, it is not obvious how grouping or filtering might aid identification of an object moving relative to the scene.

It therefore seems unlikely that the potential solutions above could enable an observer to detect a moving object during self-movement. Let us now approach the problem from a different perspective. When you move your head, even though images of objects in the scene trace out complicated patterns of motion across the retina (Fig. 1), the scene appears rigid; the dominant percept is of stability, not movement. Many explanations have been proposed for this phenomenon (Wallach, 1987), but in light of modern research findings, we can now propose a new one. The human brain has series of cortical areas along the visual pathway that are specialised for the recognition of the distinctive patterns of retinal motion associated with different components of self-movement, e.g., the radial pattern of motion that is characteristic of forward translation (see Wurtz, 1998). We will refer to these areas using the functional description, “optic flow detectors”. It is already known that optic flow detectors have a role in reducing retinal motion by driving reflexive eye-movements that stabilise the object of interest on the retina during self-movement (Bussetini, Masson,

& Miles, 1997). Might optic flow detectors have a further role to play as complex motion filters? If the components of retinal motion due to self-movement can be identified and ‘parsed out’<sup>1</sup> by such detectors, then the problem of identification of a moving object in the scene during self-movement reduces to the simple, observer-stationary case; any motion remaining after parsing of the retinal image by optic flow detectors can be attributed to movement of an object within the scene (Royden et al., 2001; Rushton & Warren, 2005).

In this paper, we explore the idea raised above. We first determine whether moving objects do pop out in the complex motion fields that result during self-movement. We go on to test a specific prediction of the *flow-parsing hypothesis* introduced above. In the experiments presented below, we use a variant of the well-established visual search paradigm and examine the time it takes to spot a single object moving through the scene as a function of the number of objects in the scene. We measure search times in both (simulated) observer-moving and observer-stationary conditions.

The assumptions that underpin the interpretation of results from pop out, “parallel”, “preattentive” or “efficient” visual search tasks are well known (see Tresiman & Gelade, 1980), but for clarity we repeat them here: if an object pops out, or to use slightly different terminology, is conspicuous, it should be spotted equally rapidly in either a cluttered or sparse scene. If the object does not pop out, the observer will have to actively search for it and the time taken will be a function of the number of objects in the scene. It is generally accepted that an object that moves pops out when the observer is stationary (Dick, Ullman, & Sagi, 1987), so performance in an observer-stationary condition provides a benchmark for evaluation of performance in the (simulated) observer-moving condition.

## 2. General methods

### 2.1. Observers

Six observers participated in the first experiment, five in the second. All were staff or students within the School of Psychology at Cardiff University. Two of the observers were authors; the remainder were naive as to the purpose of the experiment. All observers had normal or corrected to normal vision, none had previously been found to have any deficits in perception of stereo depth during previous experimental work. Observers’ participation in the experimental studies was regulated by the Ethics Committee of the School of Psychology, Cardiff University.

---

<sup>1</sup> It is not necessary to assume that optic flow detectors would literally separate the retinal flow into different components. An alternative description would involve searching within a filter (McLeod et al., 1988).

## 2.2. Displays and procedure

In the experiments that follow we generated patterns of retinal motion through *simulation* of self-movement. At all times the observer remained stationary (head on a chin rest).

Nine, sixteen or twenty-five red textured objects of approximately  $2 \times 2 \times 2$  cm were randomly oriented and placed within a volume of  $26 \times 26 \times 50$  cm (see Fig. 1), with the centre of the volume 105 cm directly ahead of the observer. The observer viewed the objects displayed on a 22" Viewsonic p225F CRT, placed at a distance of 0.75 m from the observer, in a pitch-black room. Perspective-correct images were rendered in OpenGL and anti-aliased. Objects were drawn in red (because it has the fastest phosphor) and a red filter was placed in front of the screen (to increase contrast). Left and right eye views were each generated at 50 Hz, temporally interleaved and viewed through synchronised shutter glasses (Stereographics, CA). The result was a display that produced a compelling percept of 3D objects floating in space (note that these display conditions differ markedly from those used by other investigators in previous studies on search and movement).

The computer-rendered objects either remained stationary (stationary observer condition, "static") or were moved, rotated and scaled in a manner that was consistent with a lateral translation of the head at 3 cm/s, combined with a counter-rotation of the head to keep the centre of the volume fixated (simulated moving observer condition, "move"). The resultant pattern of motion corresponds to that you might experience if you were sat at your desk and moved sideways in your chair whilst studying a collection of flies resting on (deep) bookshelves beyond your desk.

During a trial, after 0.4–0.7 s, one of the elements began to move laterally at 1 or 1.5 cm/s (relative to the other scene-stationary objects; randomly with or against the direction of simulated self-movement). We refer to these as "slow" and "fast" in the remainder of the paper. The target element was therefore identified by both movement and an onset of movement relative to the scene. Observers were instructed to press a button indicating movement direction (leftwards or rightwards) as soon as the moving element was detected. Each observer took part in four sessions of one hundred trials. On each trial the number of objects (9, 16 or 25) and display condition were chosen randomly.

## 2.3. Analysis

The distribution of times in reaction time studies is typically positively skewed (and we confirmed it was here) therefore, we used medians as the measure of central tendency and log transformed response times prior to statistical testing. Median response time and the percentage of wrong responses (error rates) were calculated for each observer, for each experimental condition (condition  $\times$  number of objects). Reaction times and error rates were compared by ANOVA. Slope of the response time curves (ms per object) was also calculated for each condition.

### 3. Experiment 1

The average response times are shown in the left panel of Fig. 2.

The response times were slightly longer in the moving observer conditions (broken lines), which is to be expected from the difference in response times in a comparable single target motion detection task (Rushton & Warren, 2005). However, the critical finding is that, as in the stationary observer condition, the search time does not increase with the number of scene objects<sup>2</sup> (and there was no difference in error rates between conditions). The slopes of the curves for each condition were:  $-1.2$  ms/object (slow-move),  $-0.9$  ms/object (slow-static),  $-0.8$  ms/object (fast-move),  $-1.7$  ms/object (fast-static). Confirmatory statistical testing showed no interaction between condition and number of objects,  $F(6, 30) = 0.84$ ,  $p = 0.55$ ,  $\eta_p^2 = 0.144$ , and no effect of condition on error rates  $F(3, 15) = 0.66$ ,  $p = 0.59$ ,  $\eta_p^2 = 0.117$ . Therefore, we can conclude that an object that moves within the scene pops out in both our moving and stationary observer conditions.

### 4. Experiment 2

A second experiment was motivated by the flow-parsing hypothesis introduced above. From a computational perspective, identification of the type of self-movement that caused a pattern of retinal motion becomes more difficult when there is no information about depth relations within the scene (Cutting & Readinger, 2002; van den Berg & Brenner, 1994). If flow-parsing was implicated, then we would predict that the search task would become more difficult in the moving observer case when the binocular disparity (depth order) information is removed. We removed this information by presenting the same image to both the left and right eyes (a “synoptic” presentation; the natural analogue is when objects are distant and binocular disparities become too small to be registered). To reduce the number of trials, we used a single speed (intermediate between the fast and slow speeds from the previous experiment).

The right panel of Fig. 2 shows that when there is no disparity depth, the response times for the moving observer condition (synoptic-move) increase as a function of the number of elements, i.e., the moving element can no longer be said to pop out.

The slopes of the curves for each condition were:  $65.4$  ms/object (synoptic-move),  $0.3$  ms/object (synoptic-static),  $-4.2$  ms/object (stereo-move),  $0.1$  ms/object (stereo-static). Confirmatory statistical analysis showed a significant interaction between condition and number of scene objects,  $F(6, 24) = 7.22$ ,  $p = 0.00017$ ,  $\eta_p^2 = 0.64$ , reflecting a significant main effect of number of scene objects in the synoptic-move condition,  $F(2, 8) = 7.74$ ,  $p = 0.013$ ,  $\eta_p^2 = 0.66$ , but none of the others: synoptic-static

<sup>2</sup> Note that the element density varied with the number of elements. In a replication with a constant density, response time did increase gradually with the number of elements, but the gradients remained similar across conditions.

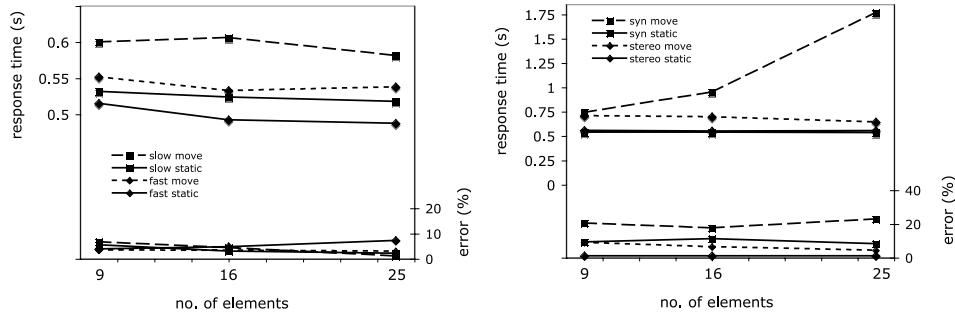


Fig. 2. Response times and error rates as a function of the number of elements. Data points indicating response times are found towards the top half of the graphs. Data points indicating percentage of incorrect direction judgements are found towards the bottom half of the graphs. Note: different axis scalings are used to aid illustration of the data. Left panel shows data for slow (square markers) and fast (diamond markers) targets, and static (solid lines) and moving viewpoints (broken lines). Right panel shows the influence of stereo depth information on response times. Square markers indicate no disparity conditions, broken lines indicate moving conditions.

( $F(2, 8) = 0.013$ ,  $p = 0.987$ ,  $\eta_p^2 = 0.003$ ), stereo-move ( $F(2, 8) = 3.748$ ,  $p = 0.071$ ,  $\eta_p^2 = 0.484$ ; note decrease in response time with number of objects), stereo-static ( $F(2, 8) = 0.032$ ,  $p = 0.968$ ,  $\eta_p^2 = 0.008$ ). Analysis of error rates also showed a main effect of condition,  $F(3, 12) = 14.95$ ,  $p = 0.00023$ ,  $\eta_p^2 = 0.79$ , reflecting a significant difference between synoptic-move and the other three conditions (pairwise comparisons all  $p < 0.05$  between synoptic-move and other conditions).

The impact of the presence or absence of depth order information is therefore compatible with the flow-parsing hypothesis. We note that this result also serves as a useful control, demonstrating that observers were not able to complete the previous experiment on the basis of differences in 2D motion. The synoptic results are in line with the results of an earlier study (Royden et al., 2001) which did not include depth order information.

## 5. Conclusions

These results demonstrate that an object that moves pops out even against the complex pattern of retinal motion that results from self-movement. Let us consider how this finding fits with some previous research in visual attention.

Nakayama and colleagues (e.g., Nakayama & Silverman, 1986) suggested that attention is deployed over a high-level representation of the visual scene rather than a low-level featural representation and, furthermore, that this high-level representation is a collection of surfaces. The results reported here are compatible with the general thrust of their proposal (attention is deployed on a high-level representation), but difficult to reconcile with the surface proposal. Here, target objects are identified not by motion relative to a surface, but rather motion relative to a rigid 3D array of objects.

A moving object is found amongst stationary objects more rapidly than a stationary object amongst moving objects. MacLeod and colleagues (e.g., MacLeod et al., 1988) proposed a “motion filter” model to account for this and other findings. A motion filter “sees” moving objects but does not “see” static objects. Therefore, the motion filter can “ignore” any number of static objects and simply “look” for a moving object. In contrast a static object amongst moving objects can only be identified by inspecting each object to find an absence of motion. The results reported here could be described within a motion filter framework, but the account would have to be revised to include a more complex filter than previously envisaged. Whether this filter could still be neural area MT as originally suggested by MacLeod and colleagues is open to question. The disparity tuning of MT (DeAngelis, Cumming, & Newsome, 1998) may allow it to identify a moving object if it moves at a different velocity to others with a similar depth range, but area MST, which is sensitive to more complex motion patterns than MT (see Smith, Wall, Williams, & Singh, 2005), might be a better candidate.

Pylyshyn suggested that the ability to track multiple moving objects may underpin a stable representation of the visual world during observer movement (e.g., Pylyshyn & Storm, 1988). We can turn this proposal around and suggest that the motion-parsing we propose here would allow a multiple-object tracking mechanism to act on an observer-movement-independent stable representation of scene-relative motion.

To conclude, we turn our attention back to research on optic flow. Previous work has been motivated by the hypothesis that patterns of retinal motion characteristic of observer movement are recognised and then used in the visual guidance of locomotion (Gibson, 1979; Warren & Hannon, 1988 but see Rushton, Harris, Lloyd, & Wann, 1998). We propose a different role: the brain identifies components of retinal motion due to self-movement so that it can separate them out; remaining motion can then be attributed to movement within the scene. Consequently, we suggest that it is our ability to recognise and account for (retinal motion due to) self-movement that underpins our ability to recognise scene relative object movement.

### Acknowledgements

Mark Bradshaw passed away before the completion of this manuscript. We hope that he would have been pleased with its final form. We thank Wendy Adams, Erich Graf, Alex Holcombe, Bob Snowden, Petroc Sumner, Mei-Ling Huang and two anonymous reviewers for useful comments on drafts of this paper.

### References

- Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, *14*, 427–432.
- Bussetini, C., Masson, G. S., & Miles, F. A. (1997). Radial optic flow induces vergence eye movements with ultra-short latencies. *Nature*, *390*, 512–515.



- Cutting, J. E., & Readinger, W. O. (2002). Perceiving motion while moving, or how pairwise nominal invariants make optical flow cohere. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 731–747.
- DeAngelis, G. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature*, *394*, 677–680.
- Dick, M., Ullman, S., & Sagi, D. (1987). Parallel and serial processes in motion detection. *Science*, *237*, 400–402.
- Duncan, J. (1995). Target and non-target grouping in visual search. *Perception & Psychophysics*, *57*, 117–120.
- Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics*, *65*, 999–1010.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.
- Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., et al. (2005). Multiple-object tracking is based on scene, not retinal, coordinates. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 235–247.
- McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature*, *320*, 154–155.
- Morvan, C., & Wexler, M. (2005). Reference frames in early motion detection. *Journal of Vision*, *5*, 131–138.
- Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*, 264–265.
- O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change blindness as a result of 'Mudsplashes'. *Nature*, *398*, 34.
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.
- Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277.
- Royden, C. S., Wolfe, J. M., & Klempen, N. (2001). Visual search asymmetries in motion and optic flow fields. *Perception & Psychophysics*, *63*, 436–444.
- Rushton, S. K., & Bradshaw, M. F. (2000). Visual search and motion - is it all relative? *Spatial Vision*, *14*, 85–86.
- Rushton, S. K. & Bradshaw, M. F. (2004). Object motion from structure: the detection of object motion by a moving observer [Abstract]. *Journal of Vision*, *4*(8), 795a, <<http://journalofvision.org/4/8/795/>>. doi:10.1167/4.8.795.
- Rushton, S. K., Harris, J. M., Lloyd, M. R., & Wann, J. P. (1998). Guidance of locomotion on foot uses perceived target location rather than optic flow. *Current Biology*, *8*, 1191–1194.
- Rushton, S. K., & Warren, P. A. (2005). Moving observers, relative retinal motion and the perception of object movement. *Current Biology*, *15*, R542–R543.
- Smith, A. T., Wall, M. B., Williams, A. L. W., & Singh, K. D. (2005). Sensitivity to optic flow in human cortical areas MT and MST. *European Journal of Neuroscience*, *23*, 561–569.
- Tresiman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Tripathy, S. P., & Barrett, B. T. (2004). Severe loss of positional information when detecting deviations in multiple trajectories. *Journal of Vision*, *4*, 1020–1043.
- van den Berg, A. V., & Brenner, E. (1994). Why two eyes are better than one for judgements of heading. *Nature*, *371*, 700–702.
- Wallach, H. (1987). Perceiving a stable environment when one moves. *Annual Review of Psychology*, *38*, 1–27.
- Warren, W. H., & Hannon, D. J. (1988). Direction of self-motion is perceived from optical flow. *Nature*, *336*, 162–168.
- Wurtz, R. H. (1998). Optic flow: a brain region devoted to optic flow analysis? *Current Biology*, *8*, R554–R555.