Journal of Experimental Psychology: Human Perception and Performance

Embodied Memory: Effective and Stable Perception By Combining Optic Flow and Image Structure

Jing Samantha Pan, Ned Bingham, and Geoffrey P. Bingham

Online First Publication, March 11, 2013. doi: 10.1037/a0032070

CITATION

Pan, J. S., Bingham, N., & Bingham, G. P. (2013, March 11). Embodied Memory: Effective and Stable Perception By Combining Optic Flow and Image Structure. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. doi: 10.1037/a0032070

Journal of Experimental Psychology:
Human Perception and Performance

© 2013 American Psychological Association
0096-1523/13/$12.00 DOI: 10.1037/a0032070

Embodied Memory: Effective and Stable Perception By Combining Optic Flow and Image Structure

Jing Samantha Pan
Indiana University

Ned Bingham
Cornell University

Geoffrey P. Bingham
Indiana University

Visual perception studies typically focus either on optic flow structure or image structure, but not on the combination and interaction of these two sources of information. Each offers unique strengths in contrast to the other's weaknesses. Optic flow yields intrinsically powerful information about 3D structure, but is ephemeral. It ceases when motion stops. Image structure is less powerful in specifying 3D structure, but is stable. It remains when motion stops. Optic flow and image structure are intrinsically related in vision because the optic flow carries one image to the next. This relation is especially important in the context of progressive occlusion, in which optic flow provides information about the location of targets hidden in subsequent image structure. In four experiments, we investigated the role of image structure in "embodied memory" in contrast to memory that is only in the head. We found that either optic flow (Experiment 1) or image structure (Experiment 2) alone were relatively ineffective, whereas the combination was effective and, in contrast to conditions requiring reliance on memory-in-the-head, much more stable over extended time (Experiments 2 through 4). Limits well documented for visual short memory (that is, memory-in-the-head) were strongly exceeded by embodied memory. The findings support J. J. Gibson's (1979/1986, *The Ecological Approach to Visual Perception*, Boston, MA, Houghton Mifflin) insights about progressive occlusion and the embodied nature of perception and memory.

Keywords: embodied memory, progressive occlusion, visual short-term memory (STM), optic flow, image structure

Two camps have been distinguished in the study of visual perception, based on theoretical approach, topics of research, and corresponding methodologies, namely, the constructivist and the ecological camps (Norman, 2002). The constructivist focuses on the perception of visual targets (e.g., their identity, shape, size, orientation) based on static image-based information, in which observers are treated as passive receivers of sensory stimulation that they use to construct cognitive representations of the perceptual targets. The ecological camp focuses on the perception of events and actions using optic flow information that is generated by motions in the interaction between the observer and the environment. In this case, optic flow information is contained in the observer–environment system and perception is direct, involving no mental construction but only detection and use of information.

The two visual systems theory has accommodated these two aspects of vision anatomically by hypothesizing two separate neural pathways, respectively (Milner & Goodale, 1995, 2006). A ventral pathway was hypothesized to deal with image-based processing to perform visual recognition and identification, receiving input from the primary visual cortex and located mainly in the inferotemporal cortex. In contrast, the dorsal system was hypothesized to processes motion and optic flow information to guide actions, receiving visual input from the primary visual cortex and subcortical areas, and located mainly in the posterior parietal cortex. Neuropsychological studies of patients with optic ataxia (i.e., damaged dorsal system) and patients with visual agnosia (i.e., damaged ventral system) have shown that deficits in one system minimally affect functions of the other system (Milner & Goodale, 1995, 2006; Milner, Paulignan, Dijkerman, Michel, & Jeannerod, 1999). Hence, Milner and Goodale have suggested that the two systems can be dissociated. The theory has inspired a great deal of research focused on either image-based vision or optic-flow-based vision (see Norman, 2002, for review).

Psychophysical studies of visual perception also exhibit this dissociation between image-based and optic flow-based approaches. A canonical line of research in the image-based approach is object recognition, which is concerned with how individuals identify and categorize objects. Most of this research has been dedicated to the problem of recovering 3D structure from

Jing Samantha Pan, Department of Psychological and Brain Sciences, Indiana University; Ned Bingham, School of Electrical and Computer Engineering, Cornell University; Geoffrey P. Bingham, Department of Psychological and Brain Sciences, Indiana University.

Correspondence concerning this article should be addressed to Geoffrey P. Bingham, Psychological and Brain Sciences, Indiana University, 1101 East Tenth Street, Bloomington, IN 47405-7007. E-mail: gbingham@indiana.edu

static 2D retinal stimulation (see Riesenhuber & Poggio, 2000, Tarr & Bulthoff, 1998, and Tarr & Vuong, 2002, for reviews). The two predominant approaches are "view based" and "structural description" models. View-based models consider object representation as collections of view-specific features that are viewer-centered and viewpoint dependent (Edelman & Bulthoff, 1992; Tarr, 1995). Structural description models, on the other hand, contend that objects are represented as spatial arrangements of their parts in 3D and, hence, that the representation is object centered and view invariant (Biederman, 1987; Hummel & Biederman, 1992). Both approaches have been supported by evidence from psychophysical experiments and neuroimaging studies (see Logothetis & Sheinberg, 1996, Riesenhuber & Poggio, 2000, and Tarr & Bulthoff, 1998, for reviews). Despite the long and ongoing debate between these approaches, both of them have effectively assumed that object recognition is based only on information that static 2D images offer. Although a common paradigm in object recognition studies is to ask observers to compare two objects after some rotation in depth (obviously a continuous process involving progressive occlusion), in these studies viewers are typically presented with discrete views of the object before and after rotation. The optic flow from the rotation is not presented to observers. Thus, in the object recognition literature, perception is traditionally studied as a process that involves converting discrete static 2D visual inputs—that is, images—into 3D objects.

In contrast, studies of visually guided action—for instance, the visual control of locomotion—focus exclusively on use of motion-generated optic flow for perception of heading (e.g.Warren & Hannon, 1990), control of steering (e.g., Li & Warren, 2000, 2002; Wann & Land, 2000; Wilkie & Wann, 2006), or control of braking (e.g., Anderson & Bingham, 2010, 2011; Fajen, 2005a, 2005b, 2008; Yilmaz & Warren, 1995).[1] When an individual locomotes through the environment, the motion generates a global pattern of optic flow that surrounds the moving observer. This pattern was described in an early study by Nakayama and Loomis (1974) as including radial outflow from a focus of expansion (FOE) in the direction of heading, radial inflow to a focus of contraction (FOC) in the direction of retreat (opposite to the direction of heading), and parallel flow in the directions perpendicular to the axis between heading and retreat. The optic flow is also structured by variation in the distances of surfaces in the environment surrounding the locomoting observer. This produces motion parallax that, as Nakayama and Loomis noted, exists everywhere in the global flow pattern, with the exception of only two points, the FOE and FOC. In motion parallax, flow speeds projected from closer surfaces are faster than those projected from farther surfaces. Because the environment is typically populated by opaque surfaces (or is cluttered, e.g., Gibson, 1979/1986), such motion parallax yields progressive occlusion that, therefore, occurs nearly everywhere around the locomoting observer. When a closer surface passes in front of one that is farther away, the optical texture projected from the latter is deleted along the contour projected from the relevant edge of the front surface.[2]

Studies of the perception of heading feature radial expansion in optic flow and the FOE (e.g., Warren & Hannon, 1990). Studies of braking also focus on the radial expansion in optic flow, but, in this case, it is all about the changing rates of flow produced as the approach distance shrinks (Yilmaz & Warren, 1995). Steering requires that obstacles along the path of locomotion be avoided, and models thus far have featured the steering dynamics rather than analysis of the relevant optic flows (Fajen & Warren, 2003). The relevant flow structures necessarily include, in addition to radial outflow from the FOE, motion parallax and the deletion/accretion of progressive occlusion/disocclusion. So, for instance, as an observer runs through the forest (as in many scenes from *The Last of the Mohicans*, for instance), he or she must anticipate an obstacle (another tree), visible moments ago but now hidden behind the tree around which one is about to turn. The perception of, and memory for, hidden objects is relevant to a wide variety of common tasks, from driving a car on the freeway, to running across campus among all the busy students during a change in classes, to the defensive lineman searching for the quarterback, to the Hoosier hunter seeking the squirrel in the forest canopy. It is what enables a child to cheat at hide-and-seek. Despite its universal occurrence, the perception of, and memory for, hidden objects is a topic that has received little attention in research. It is our topic here, and it requires a study of vision that entails both optic flow and image structure. The child peeking in hide-and-seek experiences optic flow as she watches all the other children running for their hiding places. This includes progressive occlusion as the children finally become hidden. However, once they are gone, the seeker is left with only the static image structure, projected from the surrounding surfaces behind which all the other children are hiding. In this circumstance, this static image structure is embodied memory that helps the seeker find the hidden children. This is what enables the child to cheat by having watched everyone run out of sight (and detecting the corresponding optic flow).

With the above examples in mind, the dichotomy in research on visual perception between image-based and optic-flow-based approaches is a historical artifact that is not representative of natural visual functioning. Normally, vision entails use of both images and optic flow. On the one hand, images and optic flow are different and offer different advantages. On the other hand, they are inalienably related because optic flow is what carries one image to the next, as a result of motion among or relative to surfaces surrounding the observer. We argue that progress in the study of vision requires that image structure and optic flow be rejoined in analyses of visual function. In the current work, we introduce a new paradigm that involves recalling and identifying multiple hidden targets in a dynamic environment. The most effective performance relies on both static image structure and motion-generated optic flow.

Image structure and optic flow each entail advantages and disadvantages, but when combined, they complement each other to yield effective performance. Optic flow provides immediate and powerful information about the 3D structure of the surroundings,

[1] It is important to note that object recognition does not normally involve only image structure and visually guided locomotion does not normally entail only optic flow. On the one hand, optic flow provides important information about 3D object structure relevant to the recognition of objects and investigated in SFM studies (e.g., Lee, Lind, Bingham, & Bingham, 2012; Tittle & Braunstein, 1993). On the other hand, an unmoving person may decide in what direction to locomote by using information available in static image structure before they take their first step (e.g., Rushton, Harris, Lloyd, & Wann, 1998).

[2] When a surface comes back into view, the optical texture projected from its surface is accreted along the contour projected from the relevant edge of the surface in the front.

the layout of surfaces in cluttered terrain (Domini & Caudek, 2003; Todd, 1995), and the relative speeds and directions of motions (see Warren, 2008, for a review). For an observer translating through rigid surroundings, the speed of the optic flow covaries with the 3D distance of surfaces, providing a direct and immediate map of environmental layout (Simpson, 1993). However, optic flow information is ephemeral. It varies in quality with the relative speeds of motion and becomes unavailable when motion stops. For example, strong optic flow specifying 3D layout is generated when one walks into a workspace (e.g., an office or kitchen) to perform subsequent manual tasks, but the optic flow disappears, or at least is significantly weaker, when one stops locomoting. While a perceiver remains standing or seated, and is thus without strong optic flow, must he or she retain all previously detected optic flow information about the surroundings strictly in the head (that is, in memory, as traditionally construed)? Perhaps not, given the intrinsic relation between optic flow and image structure, and given the fact that the image structure would remain present.

Image structure is weaker in its ability to specify the 3D layout of an environment, but it is persistent. Image-based vision relies on cues (such as image size, texture gradients, or height in the visual field) to deduce depth relations based on experience. It reduces the 3D dynamical environment to flat and static snapshots from which perceivers extract useful cues to judge depth and distance. Problems like figure–ground ambiguity often result. Although image structure is weak in specifying depth relations, it is stable, and, given the symmetry between image structure and optic flow, image structure can be used to preserve information about 3D structure provided by optic flow. Because optic flow carries one structured image into the next structured image, optic flow and image structure are intrinsically related and largely symmetric in respect to the layout of surfaces from which the structure is projected. In part, the relation could be cast as a calibration of image-based information about 3D structure by the more powerful optic flow information. More than this, optic flow can also specify the changes in 3D spatial structure that, in turn, relates sequential images. The object recognition problem, studied in so many characteristic experiments using only the static images before and after a target object has rotated in depth, would become much easier if the optic flow produced by that rotation were made available to the observer (Bingham & Lind, 2008; Lee, et al., 2012). Still, once the optic flow has ceased, the static images remain, and thus help preserve the information provided by the optic flow.

The combination of optic flow and image structure would make perception most effective. Offloading the information provided by transient optic flow to external stable image structure allows individuals to access and act upon spatial information provided by optic flow without having to hold it all in the head. In this way, image structure becomes an embodied memory system for situated, active observers. It is embodied because the image structure is projected from the substantial surfaces of the environment in which the (substantial or embodied) observer is situated and, therefore, to which the observer is related. Gibson (1950, 1979/1986) pointed out that observers are embodied (or are substantial or physical) and, as such, must be supported by a (substantial or physical) ground or support surface in the environment (see also Carr, 1935). Otherwise, observers would be in free fall. Observers are also typically surrounded by a layout of substantial surfaces. The gravity that constrains this configuration of embodied or substantial entities also relates them by "forcefully" providing a common orientation. The surrounding layout of substantial surfaces projects structured light (or image structure) to the observer. This resulting image structure can then serve as an embodied memory, as described and investigated in the current studies.

This optic flow and image structure synergy has been shown to allow the perception of metric object shape and the guiding of accurate reaches-to-grasp (Lee & Bingham, 2010). Using either judgment (e.g., Johnston, 1991; Norman & Todd, 1993) or reach-to-grasp (Lee, Crabtree, Norman, & Bingham, 2008) measures, perception of metric shape has been shown to be inaccurate when only small perspective changes (~ 10° to 15° change) are available in optic flow. However, large perspective changes of 45° (or more) have been found to allow accurate perception of metric shape (e.g., Bingham & Lind, 2008; Brenner & van Damme, 1999). Such large changes are typically available to locomoting observers but not seated ones. Lee and Bingham (2010) investigated whether large perspective changes (>45 degrees) would enable seated observers to perceive metric shape and use the information to guide accurate feedforward[3] reaches-to-grasp after optic flow had stopped and only static image structure remained available. They found that the large perspective changes allowed accurate reaches-to-grasp performed immediately after motion stopped, but would it still work after a delay during which only static image structure was available? Hu, Eagelson, and Goodale (1999) found that reaches became inaccurate within 5 s after removal of visual image structure, but what happens if the image structure remains and only the optic flow ceases (and is thus removed)?

Lee and Bingham (2010) found that participants were still able to perform accurate reaches-to-grasp after a 5-s delay, during which image structure, but not optic flow, was available. Then, when multiple objects were viewed with optic flow, followed by subsequent reaches-to-grasp that were performed in series with only image structure available, performance remained accurate over the much longer delays that were incurred. This experiment, combined with the results of Hu, Eagelson, and Goodale (1999) and the previous results of Lee, et al. (2008), demonstrated that optic flow information was necessary to enable accurate shape perception (and thus accurate reaches-to-grasp), but that persistent image structure was necessary and sufficient for continued accuracy of performance once optic flow had ceased.

We hypothesize an optic flow and image structure synergy in which optic flow provides information about the 3D layout of objects and surfaces, and image structure (remaining at the terminus of flow) allows the resulting perception to remain stable. In the current study, we test whether this can facilitate perception of multiple objects in a cluttered environment in which perspective changes can take visible objects out of view while providing information in resulting optic flow about where they have gone. Information carried in optic flow is only available while the interaction is ongoing, but this information could be preserved in the remaining stable image structure for future access. We propose that when objects are perceived in a

[3] Feedforward reaches are performed without being able to see the hand, only the target object, and thus online visual guidance cannot be used. The reach-to-grasp is controlled using only the available static image structure specifying the location, shape, size and orientation of the target object.

3D space, image structure is calibrated by optic flow information generated by motion of the objects and/or the observer. Subsequently, spatial information in optic flow is preserved in image structure and remains accessible after optic flow ceases. The point is that the stability of the information would allow performance in recalling the spatial layout to be more accurate than if supported merely by spatial memory without image structure remaining available.

To test our claim, we modified Kaplan displays (Gibson, 1979/1986, p. 189; Kaplan, 1969), which were originally designed to demonstrate perception of progressive occlusion using optic flow. In the Kaplan display, a randomly textured square was perceived to move in front of a background composed of identical random textures. When viewing only any single frame from the display (i.e., a static picture of the square and the background), an observer could only see a single textured surface (that is, the square was invisible). However, if the continuous motion was presented, an observer could immediately see both surfaces, one in front of the other, separated in depth. In the current study, the displays also involved two moving planar surfaces separated in depth, one in front of the other. Targets on the back surface could be seen through windows (or holes) in the front surface. The surfaces moved relative to one another, producing progressive occlusion taking the targets out of view beyond the windows. In some conditions, we added image structure to the original Kaplan display by placing visible contours around the windows. We tested recall of the locations of targets in conditions with only image structure, only optic flow, or both in combination. We used a spatial memory task to test participants' perception and retention of object locations. The response task was performed once optic flow had ceased, and entailed the identification of previously perceived, but now occluded, targets.

In this study, we performed four experiments to investigate how information available in optic flow and in image structure interacted to allow participants to identify locations of multiple targets. In Experiments 1, 2, and 3, we compared how well participants could recall locations of target objects when (a) only optic flow information was available, (b) only image structure was available, or (c) both optic flow and image structure were available. In Experiment 4, we further explored how well spatial information was preserved, given both optic flow and image structure with extended time delays up to nearly half a minute. The tasks were representative of naturally occurring conditions under which spatial information must be used to identify locations of objects previously observed to go out of view.

Experiment 1

Experiment 1 was designed to test how well participants could recall locations of target objects when only optic flow was available without corresponding static image structure.

Method

Participants. Fifteen adults (seven males and eight females between 22 and 33 years of age) participated in the experiment. All participants had normal or corrected-to-normal vision. Participants were paid $7 per hour for completion of the experiment.

Apparatus. Participants sat in front of a computer monitor (display width = 43 cm; height = 27 cm) with a viewing distance of 50 cm. The refresh rate of the monitor was 60 Hz.

Procedure. Participants read and signed consent forms and then sat in front of the test computer to complete three to 10 practice trials in the presence of the experimenter to become familiar with the task. The basic display consisted of two rectangular surfaces, one smaller and in front of the other, and both parallel to the computer screen (thus, frontoparallel). The surfaces were randomly textured in exactly the same way, that is, identical density of binary (black/white) texture. The rear surface extended well beyond the edges of the computer screen and its edges never appeared on screen during the display. The front surface was smaller (27° × 27° visual angle), so that it occluded the rear surface only in the central portion of the display, that is, portions of the rear surface could be seen beyond edges of the front surface. The front surface also contained cutout holes or windows through which the other back surface could be seen. In some of these windows, pink squares could be seen lying on the rear surface. The pink squares were targets. Windows showing only random texture were distracters. Both the windows and the targets were 1.5 cm × 1.5 cm squares (a little smaller than 1° × 1° visual angle). The two surfaces were separated in depth (although when the display was static, the depth separation could not be seen and the two surfaces appeared to be one, just as in the original Kaplan, 1969, displays).

Each experimental trial consisted of four phases: rotation, translation, delay, and response. An experimental trial started with the two planar surfaces rotating in depth. This structure-from-motion revealed the depth relation between the surfaces. Participants watched the surfaces rotate for 7 s and studied the locations of the targets. After the rotation stopped, with both surfaces once again lying frontoparallel, the rear surface translated rigidly in one of the eight directions (i.e., one of the N, NE, E, SE, S, SW, W, or NW directions) parallel to the surface itself, while the front surface remained stationary. This rigid translation of the rear surface could be seen both through the windows in the front surface and beyond the four edges of the front surface. As the rear surface translated, the pink squares on it passed beyond the windows in the front surface and thus became occluded behind the front surface. When the translational motion stopped, the hidden locations of the targets would be seen in terms of the distance and direction that the entire rear surface had moved rigidly. This movement took 3 s. At the end of the translation, targets were completely occluded by the front surface and only random texture of the rear surface could be seen through the windows in the front surface. A delay of 2 s was introduced after the translation occurred, during which participants saw a static image of the surfaces (which now appeared to be one homogeneous field of random textures). Participants then used the mouse to click on the locations of the targets, which were now hidden behind the front surface.

The structure of these displays was similar to that of the Kaplan displays (Kaplan, 1969, see also Gibson, 1979/1986, pp. 189–191). The windows and the presence of the two surfaces could only be seen with optic flow (that is, during the rotation and translation phases). Once motion stopped, only a single static display of random texture could be seen. Thus, when the optic flow stopped, the windows were lost as landmarks (see Figure 1).[4]

We encouraged participants to click accurately instead of randomly by introducing a point system: starting with 200 points, if they identified a target correctly (that is, a “hit”), they gained a point; if they identified a target incorrectly (that is, clicking were there was no target or a “false alarm”), they lost a point; if they did not attempt to identify, there would be no point change. At the end of the experiment, participants received bonus payment (in addition to the standard participation payment) proportional to their final points. This was designed to prevent guessing and to promote accurate performance. The method was effective. In all conditions of the four experiments, there were very few false alarms. In fact, the median number of false alarms in each cell for each participant was zero. This means that in all conditions tested, there were no false alarms in more than half of the trials. The extremely small number of false alarms in these experiments suggested that participants were careful and conservative when making responses. They did not guess, although sometimes they apparently did misremember. For instance, in Experiment 4, in the entire set of 1,032 trials, 26 exhibited a high number of false alarms and no hits. In these trials, it was likely that participants misremembered the direction in which the back surface had moved, taking the targets out of view beyond the windows. They probably recalled the target windows correctly, but systematically clicked on the wrong locations relative to those target windows. So participants made some errors in memory, but they did not guess. Therefore, we simply analyzed the number of targets identified correctly (that is, hits) as a measure of memory.

In this experiment, there were two versions. In Version 1 (nine participants), on each Trial 2, 3, 4, or 5, targets or distracters were shown on the display at random locations. With a crossed design (two targets with 2, 3, 4, or 5 distracters; three targets with 2, 3, 4, or 5 distracters, and so on), this yielded 16 combinations of targets and distracters. With two trials for each combination, each participant completed 32 trials in this experiment. There was a delay of 2 s between stimulus display and target identification, during which the ending scene of the display (random textures) remained on screen. In Version 2 (six other participants), 6, 9, 12, or 15 targets or distracters were presented on each trial (crossed to yield 16 unique combinations). There were 5-s delays between stimulus display and response. In one block of 16 trials, random textures were displayed on the screen during delay (No Blank condition), and in the other block of 16 trials, there was a black screen during delay (Blank condition). All six participants completed three repetitions of both No Blank and Blank trials. In Version 2, the numbers of targets and distracters were increased and blanks were introduced to make the conditions comparable with those in Experiment 3. This allowed direct comparison of performances between the optic-flow-only experiment (Experiment 1) and the optic flow and image structure experiment (Experiment 3). For the specific purpose of the current experiment, we took the No Blank trials of Version 2 and performance in Version 1 to analyze target identification with optic flow information only. The results are reported in the next section (see Table 1).

Results

In this experiment, with optic flow information alone, only a small number of targets were identified ($M = 1.62$, $SD = 1.00$). Specifically, in Version 1, with up to 5 targets and no blank during delay, the mean number of targets identified was 1.56 ($SD_{\text{Version 1}} = 0.91$); in Version 2, with up to 15 targets and no blank during delay, the mean number of targets identified was 1.68 ($SD_{\text{Version 2}} = 1.08$). The means were not significantly different, $t(556) = 1.50$, $p = .133$. In Version 1 (with 2, 3, 4, or 5 targets and no blank), no participant hit 5 target locations correctly; and among the 144 trials that had 4 or more targets, participants clicked on 4 target locations correctly only 4 times (see Figure 2, left). Although hits were significantly affected by number of targets available, $F(3, 24) = 4.01$, $p < .02$, there was no consistent linear trend (either increasing or decreasing) between number of targets identified and number of targets available. See Table 2a. In the No Blank trials of Version 2, with 6, 9, 12, or 15 targets available, the maximum number of targets identified was 6 (1 trial), and in only 16 out of 288 trials, participants identified 4 targets or more (Figure 2, right). There was no consistent trend between targets identified and targets available (see Table 2a). In combination, results from the two versions showed that, regardless of how many targets were available, participants’ performance was equally poor and the number of targets one could identify maxed out at around 4 or 5 items. Therefore, given only optic flow information, participants were able to perceive depth structure and distinguish targets from distracters to identify target locations, but when motion stopped after targets were occluded, they had to rely entirely on memory-in-the-head, with the result that performance was at or below levels characteristic of visual short-term memory.

Experiment 2

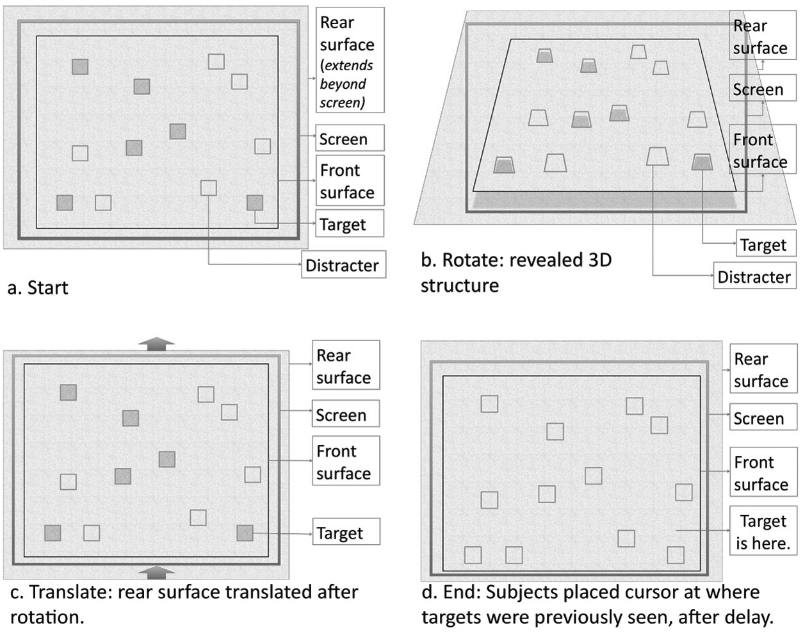
Because target identification was poor when only optic flow information was present, in this experiment, we tested how well participants could recall locations of target objects (a) when only image structure information was available, and (b) when static image structure was available in addition to optic flow information.

Method

Participants and apparatus. The same nine participants who completed Experiment 1, Version 1, completed this experiment on a different day using the same computer monitor.

Procedure. The basic stimuli and procedures were the same as those in Version 2 of Experiment 1, with a few changes. First, in this experiment, we added stable image structures. Green lines were drawn around the border of each window in the front surface. The image structure remained visible throughout all experimental trials. With the presence of stable image information, we expected

[4] Try out the full experiments (Experiment 1, 2, and 3) online at http://www.indiana.edu/~palab/research.php. Click to expand the “Perception and Embodied Memory” tab, then download the demos under “Experiment Demos.” Demos 1, 2, and 3 correspond to stimuli used in Experiments 1, 2 and 3, respectively. Demos work on Mac OS’s only. Speed of motion may change depending on graphics setting and/or the operating system.



a. Start

b. Rotate: revealed 3D structure

c. Translate: rear surface translated after rotation.

d. End: Subjects placed cursor at where targets were previously seen, after delay.

Figure 1. An illustration of stimulus display (not drawn to scale). This illustration shows the display containing both optic flow and image structure information (as used in Experiments 3 and 4). In the experiment testing optic flow only (Experiment 1), there was no visible border around targets and distracters. In the experiment testing image structure information only (Experiment 2), in half of the trials, there was no continuous translation, or Step C was skipped (Instant Shift condition); and in the other half, all steps were shown (Continuous Shift condition).

better performance in the recollection task. Second, to compare performance when optic flow was available to performance when it was absent (half the trials), the translation of the rear surface, and hence the progressive occlusion of targets on the rear surface, was not visible to participants (Instant Shift condition). Participants only saw an abrupt disappearance of targets as they jumped discretely to new, occluded locations. The display shifted instantly from one static image to another, with no continuous optic flow taking the first image to the second one. This was the same as the discrete change in rotated views widely used in object recognition paradigms that include only image-based information. As a result, observers were unable to see the direction and distance to which the targets moved. In the other half of the trials, the translation was shown and progressive occlusion was perceivable (Continuous Shift condition; see Table 1). As before, the numbers of targets and of distracters was 6, 9, 12, or 15, which again yielded 16 combi-

Table 1
Summary of Experimental Design

	N targets	*N* distracters	Delay	Blank	Border	Instant shift	Repetitions
Experiment 1							
Version 1	2, 3, 4, 5	2, 3, 4, 5	2 s	No	No	No	2
Version 2	6, 9, 12, 15	6, 9, 12, 15	5 s	No	No	No	3
	6, 9, 12, 15	6, 9, 12, 15	5 s	Yes	No	No	3
Experiment 2							
Instant Shift condition	6, 9, 12, 15	6, 9, 12, 15	2 s	No	Yes	Yes	1
Continuous Shift condition	6, 9, 12, 15	6, 9, 12, 15	5 s	No	Yes	No	1
Experiment 3							
No Blank condition	6, 9, 12, 15	6, 9, 12, 15	5 s	No	Yes	No	2
			10 s	No	Yes	No	2
			15 s	No	Yes	No	2
Blank condition	6, 9, 12, 15	6, 9, 12, 15	5 s	Yes	Yes	No	2
			10 s	Yes	Yes	No	2
			15 s	Yes	Yes	No	2
Experiment 4							
No Blank condition	9, 12, 15, 18	12	5 s	No	Yes	No	5
			25 s	No	Yes	No	5
Blank condition	9, 12, 15, 18	12	5 s	Yes	Yes	No	5
			25 s	Yes	Yes	No	5

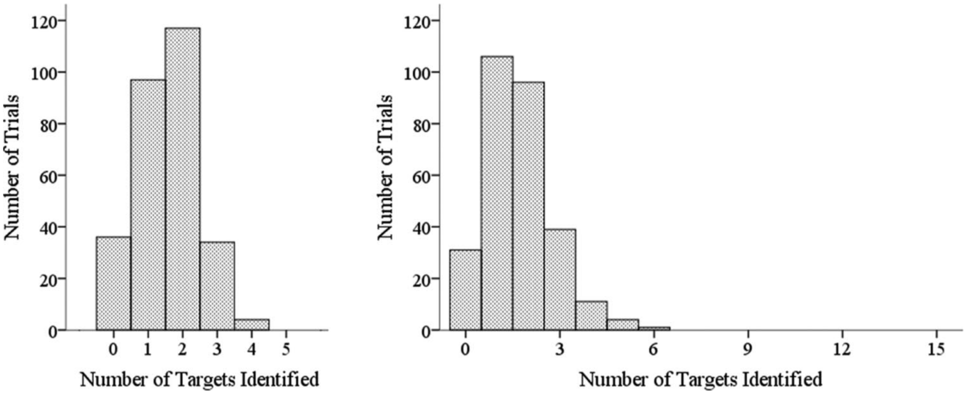


Figure 2. Frequency distributions of targets correctly identified with optic flow information only. Left: Version 1 (9 participants) with 2, 3, 4, or 5 targets and distracters, 2-s delays, and no blank during delay (total = 288 trials). Right: Version 2 (6 participants) with 6, 9, 12, or 15 targets and distracters, 5-s delays, and no blank during delay (total = 288 trials). The means, standard deviations, and general shapes of frequency distributions were very similar between the two versions.

nations of targets and distracters. The delay between stimulus display and participants' response was 5 s.

Results

When the continuous translation of the rear surface was not shown to the participants (Instant Shift condition), participants did not respond (mean number of clicks for targets, distracters, and empty spaces was zero, as shown in Table 2c). With no available information generated by progressive occlusion, participants were unable to perceive direction of movement and the relation between the front and the rear surfaces. In addition, because they would be penalized for incorrect responses, participants did not guess the direction of translation in this condition. They did not identify any target location in this condition without optic flow.

Table 2
A Summary of Mean Number of Targets Identified and Standard Deviations in Experiment 1 (Version 1 and Version 2, No Blank Trials) and Experiment 2 (Two Conditions)

Number of targets available	Mean targets identified	*SD* targets identified
a. Experiment 1: Optic flow only		
2	1.32	0.728
3	1.49	0.888
4	1.74	0.964
5	1.69	0.973
Average	1.56	0.905
6	1.29	0.863
9	1.79	0.838
12	1.63	1.130
15	2.01	1.316
Average	1.68	1.082
b. Experiment 2: Image structure and optic flow (Continuous Shift condition)		
6	5.50	1.231
9	7.39	1.626
12	8.72	2.133
15	9.39	3.524
Average	7.75	2.719
c. Experiment 2: Image structure only (Instant Shift condition)		
6	0	0
9	0	0
12	0	0
15	0	0
Average	0	0

On the other hand, when continuous translation of the rear surface was shown to the participants (Continuous Shift condition), providing optic flow information, and with the availability of stable image structure (target and distracter windows had green borders, which remained in view throughout the trials), more targets were identified compared with when there was only optic flow but not image information (i.e., Experiment 1), or when there was only image information but not optic flow information (i.e., Instant Shift condition). Moreover, there was a positive linear relationship between hits and number of targets, which suggested that the limit of embodied memory was not reached at 15 targets.

Participants in the Continuous Shift condition correctly identified more target locations on average ($\text{mean}_{\text{Continuous Shift}} = 7.75$) than in the Instant Shift condition of this experiment ($\text{mean}_{\text{Instant Shift}} = 0$; repeated measures $t[8] = 22.92$, $p < .001$). Performance in the Continuous Shift condition was also better than that in Experiment 1, Version 2, with a comparable number of targets available and no blank during delay ($\text{mean}_{\text{Exp1-V2}} = 1.68$; independent samples $t[166] = 25.77$, $p < .001$). Note that in the Continuous Shift condition, both optic flow and image structure information are available; in the Instant Shift condition, only image structure information was available; and in Experiment 1, Version 2, only optic flow information was available. Therefore, participants' performance in the condition with both sources of information was significantly better than that in either condition in which only one source of information was available.

Unlike in Experiment 1, in which number of hits failed to exceed approximately 2 (as shown in Table 2a), in the Continuous Shift condition of Experiment 2, number of hits increased with the number of targets available, $F(3, 24) = 17.82$, $p < .001$, as shown in Table 2b, and this increase was linear with a constant rate of

0.43, $F(1, 142) = 57.38$, $p < .001$, $r^2 = 0.3$, as shown in Figure 3. (We tested a second-order polynomial fit, but the quadratic term was not significant, $p > .1$.) The linear trend suggested that participants' ability to recall locations of targets did not max out at 15 targets. This clearly exceeded the frequently documented capacity of visual STM as about four items (Luck & Vogel, 1997) and motivated our next experiment in which we tested whether, given both optic flow and image structure, participants could recall locations of targets after a delay, and if so, how well, depending on the persistent availability of the image structure or the lack thereof.

Last, mean hits decreased with an increasing number of distracters, $F(3, 24) = 5.75$, $p < .005$, as shown in Figure 3. Furthermore, the number of distracters and number of targets interacted in determining the number of hits, $F(9, 72) = 2.47$, $p < .02$. However, this effect only appeared as the number of distracters reached 15. When there were 6, 9, or 12 distracters, the rate of increase in hits with number of targets remained steady at approximately 0.5 ± 0.1. However, when there were 15 distracters, this rate of increase dropped to 0.16.

Experiment 3

In the previous experiments, we showed that with motion-generated optic flow information alone, participants were able to identify targets and recall their locations after they became occluded. Optic flow itself was enough to inform participants about the spatial relations among objects and surfaces. However, the performance with optic flow alone was extremely poor: Participants were, on average, able to identify less than two targets. This number was even smaller than the suggested capacity for visual STM. We postulated that to do this task, participants had to fixate on objects and track their movements. Thus, the number of targets they could identify would have been limited to what might fall within the foveal span. We also found that with image structure information only, participants were unable to identify any targets. This was because, without optic flow, image structure alone was unable to specify progressive occlusion and, thus, the eventual location of hidden targets. Although the landmarks were persistently available, without optic flow, they were useless because participants had no information about where the hidden targets were relative to the landmarks.

In this experiment, we tested the stability of perception, given both optic flow and image structure. Specifically, we introduced three levels of time delay (5, 10, and 15 s) between presentation of optic flow information and target identification. Additionally, during the delay, participants either saw a black screen (the Blank condition) or continued to see the ending scene of the structure-from-motion (SFM) display containing bordered windows (the No Blank condition). In the former case, the black screen between perceiving and recalling interrupted the image structure, whereas in the No Blank condition, the image structure was continuous and persistent. We hypothesized that if image structure was crucial in preserving perceived locations, performance would be better in the condition with continuous image structure (the No Blank condition) than in the condition with interrupted image structure (the Blank condition). We would also compare performance across the three levels of delays to learn the temporal stability of perception formed with the presence of continuous or interrupted image structures.

Method

Participants and apparatus. Ten participants (five females and five males, between 22 and 29 years of age) completed this experiment, including the nine from Experiment 1 and Experiment 2, using the same computer monitor.

Procedure. Participants observed the same display of two spatially separated, random-textured planar surfaces, containing

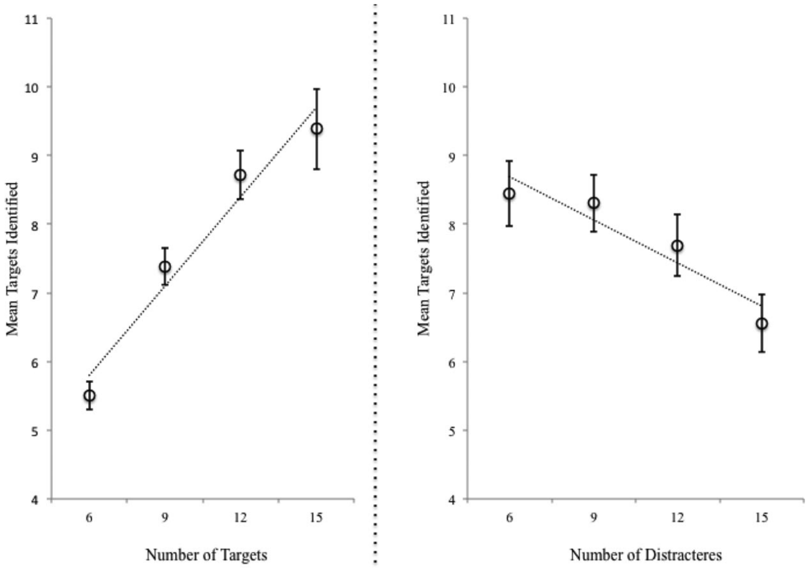


Figure 3. Mean number of target locations identified increased with number of targets and decreased with number of distracters, when both optic flow and image structure information was available (Continuous Shift condition, Experiment 2). Error bar = ±1 *SE*.

targets and distracters of size 1.5 cm × 1.5 cm. The planes rotated for 7 s to reveal depth relations between them. All windows were outlined with green borders. At the end of rotation, the rear surface translated in one of eight directions around the clock, and pink targets on the rear surface became occluded while the green borders on the windows remained in view. After the 3-s translation, participants waited for 5, 10, or 15 s to respond. During the delay, participants either saw a black screen (Blank condition) or the static image of the surfaces (No Blank condition). Subsequently, after the delay, participants responded by using the mouse to click on target locations behind the front surface.

The numbers of targets and distracters used in this experiment were 6, 9, 12, and 15. With a fully crossed design, there were 16 combinations of targets and distracters. For each combination, we tested the Blank and No Blank conditions and delay durations of 5, 10, and 15 s. This yielded 96 unique trials in one experimental block. Each participant completed two experimental blocks (192 trials total) on two separate days (refer to Table 1).

Results

In this experiment, performance, measured by the number of hits, increased with number of targets available and decreased with number of distracters. Hits were greater when the image structure information was continuously available in the No Blank condition than in the Blank condition, in which it was made unavailable during the delay (see Figure 4).

An omnibus repeated measures ANOVA was performed with blank (two levels), number of targets (four levels), number of distracters (four levels), and delay duration (three levels) as the within-subject factors. This revealed, first, a significant effect of blank versus no blank, $F(1, 9) = 18.5$, $p < .002$. In the No Blank condition (with continuous image structure during the delay), the mean number of hits was 7.44 ($SD = 2.78$), whereas in the Blank condition (in which image structure was removed during the delay), the mean number of hits was 7.06 ($SD = 3.00$). Thus, the continuously available image structure led to better recollection of perceived object locations.

In addition to blank, hits were significantly affected by number of targets, $F(3, 27) = 55.6$ $p < .001$. As shown in Figure 4, participants were able to identify more targets as the number of targets increased. In both Blank and No Blank conditions, mean hits increased by 0.4 per target object added ($r^2 = 0.97$, $F[1, 2] > 80.0$, $p \leq .01$ in both conditions). Participants' ability to identify occluded targets clearly failed to max out when there were 15 targets and 15 distracters, given these trends.

As the number of distracters increased, the number of hits decreased, $F(3, 27) = 20.3$, $p < .001$, as shown in Figure 4. Unsurprisingly, the task became more challenging with more distracters.

Furthermore, number of targets interacted with number of distracters, $F(9, 81) = 3.42$, $p < .01$. The rate of increase in hits with number of targets decreased with an increase in the number of distracters. Specifically, when there were 6 distracters, hits increased as 0.55 times the number of targets present, whereas with 15 distracters, the rate dropped to 0.31 times the number of targets present (see Figure 5).

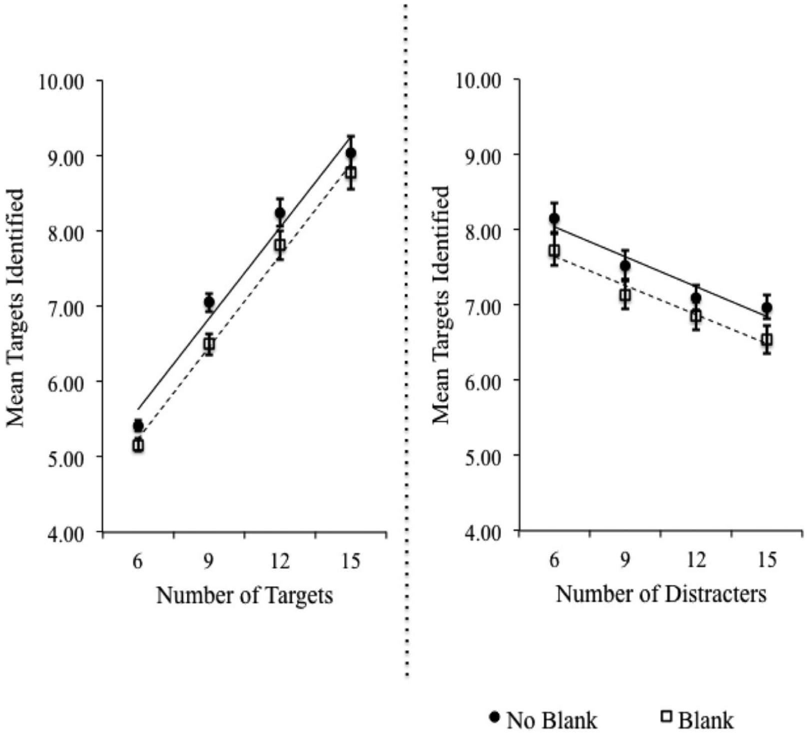


Figure 4. In Experiment 3, the mean number of target locations identified increased with number of targets (left) and decreased with number of distracters (right). Hits were consistently higher in the No Blank condition than in the Blank condition. Error bar = ±1 *SE*.

Many participants reported, during the postexperiment debriefing, that they started to work to perceive patterns with the targets and distracters when there were more items displayed at one time, and they used this configurality to facilitate later recollection. Hence, in the next experiment, we reduced the size of targets and distracters, and increased the number of potential target or distracter locations so that targets and distracters were more sparsely spread out on the screen. This was a measure to control for the pattern-detection strategy.

Finally, the omnibus ANOVA showed that lengths of delay (5, 10, or 15 s) did not significantly affect the number of hidden targets participants identified, $F(2, 18) = 2.07$, $p = .16$. Although the interaction between delay duration and blank or no blank was not significant, $F(2, 18) = 1.04$, $p = .37$, the lowest number of hits occurred in the Blank condition with the longest delay (mean hits = 6.92, SD = 3.09). Results from the current experiment suggested that performance did not drop after 15 s of delay, and this trend was not different in trials with continuous or interrupted image structures. This motivated the next experiment, in which we explicitly studied the potential interactive effect of extended delay duration and the persistence of image structure on performance.

Experiment 4

With both optic flow and image structure, participants were clearly able to outperform the visual STM capacity of four items, as expected. Our proposed explanation for this is that when optic flow and image structure are combined, information about locations of targets does not need to be kept strictly in memory-in-the-head. Instead, it can be partially offloaded to the invariant structure outside of the head and preserved in the image structure. This form of embodied memory boosted performance in this task.

To test how well the embodied memory could work, we conducted the following experiment with significantly prolonged delays, increased number of targets, and reduced target and distracter size. Small-sized targets and distracters were used, yielding a larger number of possible locations for targets and distracters and more space between them (i.e., sparse distribution despite larger numbers of targets and distracters). This was intended to make pattern detection, as a plausible aid for remembering target locations, more difficult and less likely. In the following experiment, the numbers of targets tested were 9, 12, 15, and 18, each paired with 12 distracters. Two levels of delay were tested, namely, 5 and 25 s, during which participants were presented either with continuous image structure of the visual scene or with a black screen. We hypothesized that performance should not deteriorate with increased delay in the No Blank condition because the stable external structures should enable embodied memory to remain stable, as the memory burden is offloaded to image structure, allowing target locations to be determined with respect to landmark features. However, in the Blank condition, in which traditional visual memory was required during delay, participants should identify an increasingly smaller number of target locations as the delay duration increased. Traditional memory-in-the-head is known to decay with time and is therefore relatively unstable, whereas, by hypothesis, embodied memory should be stable. We expected that delay should have a significant effect in the Blank condition, but not in the No Blank condition.

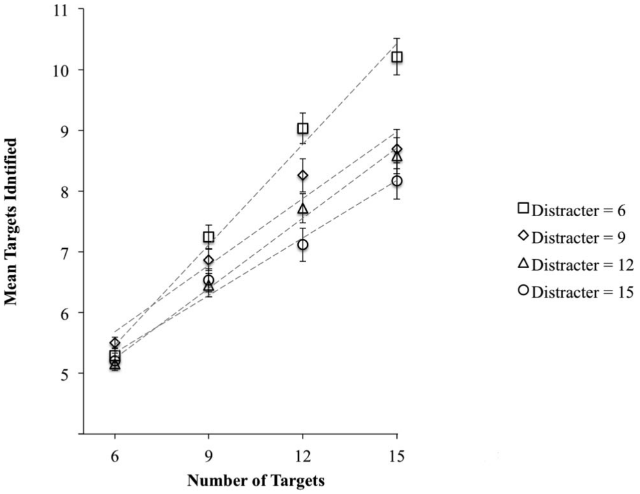


Figure 5. In Experiment 3, hits were affected by the interaction between number of targets and number of distracters. Error bar = ±1 *SE*.

Method

Participants. Thirteen adults (five females and eight males, between 20 and 33 years of age) participated in this experiment. All participants had normal or corrected-to-normal vision. Participants were paid $7 per hour for completion of the experiment.

Apparatus. As in the previous experiments, participants sat in front of a 20-in. computer monitor, with viewing distance and height adjusted to place the observer with eyes at midscreen height at 50 cm viewing distance. The refresh rate of the monitor was 60Hz.

Procedure. General experimental procedures were the same as those in Experiment 3, with the following modifications. First, sizes of targets and distracters were reduced to 0.8 × 0.8 cm squares. There were 9, 12, 15, or 18 targets, and the number of distracters was fixed at 12. As in previous experiments, targets were pink squares on the rear surface, and distracters were empty windows outlined with green borders. Two delay durations of 5 and 25 s were used. All other variables were the same as in previous experiments, including rate and duration of rotation and translation, and distance between front and rear planes. In this experiment, each participant completed five blocks of 16 unique trials that covered the four levels of target numbers, two levels of delay, and Blank or No Blank conditions (see Table 1). All participants completed the five blocks (or 80 trials) in a single session lasting approximately 1.5 hr, with short breaks taken in between blocks.

Results

Hits increased with the number of targets (see Figures 6 and 7). Variation in the duration of delay failed to affect performance in the No Blank condition (mean hits$_{\text{5s Delay}}$ = 7.90; mean hits$_{\text{25s Delay}}$ = 7.63), in which static image structure remained available throughout the delay period. However, as shown in Figure 6, increase of delay yielded a significant decrease in hits in the Blank condition (mean hits$_{\text{5s Delay}}$ = 7.53; mean hits$_{\text{25s Delay}}$ = 6.44), in which static image structure was removed during the delay and in which participants were forced to rely on memory-in-the-head.

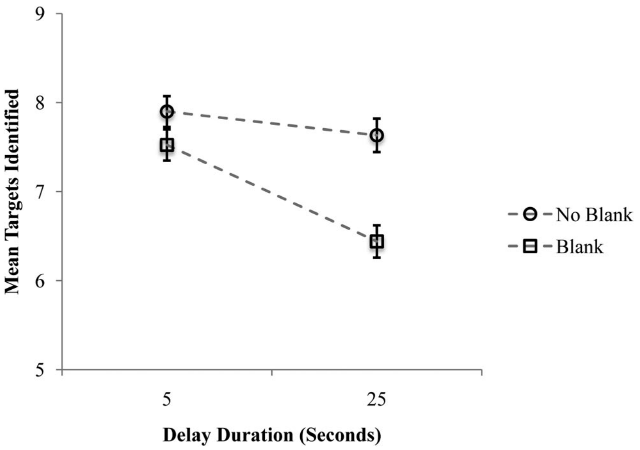


Figure 6. In Experiment 4, mean number of target locations identified was affected by the interaction between blank and delay. Error bar = ±1 *SE*.

A repeated measures ANOVA on hits, with blank (two levels), delay (two levels), and number of targets (four levels) as factors, yielded significant main effects of number of targets, $F(3, 36) = 36.20$, $p < .001$, blank versus no blank, $F(1, 12) = 19.22$, $p < .001$, and delay, $F(1, 12) = 16.23$, $p = .0017$. Overall, performance was better in trials with continuous image structure ($\text{mean}_{\text{No blank}} = 7.8$; $\text{mean}_{\text{Blank}} = 7.0$) and better in trials with short delays ($\text{mean}_{\text{5s-delay}} = 7.7$; $\text{mean}_{\text{25s-delay}} = 7.0$). More to the point, the ANOVA yielded a significant interaction of delay and blank, $F(1, 12) = 7.09$, $p < .03$. As shown in Figure 6, in trials with no blank, the variation in delay failed to affect the number of hits, whereas in the Blank condition, the number of hits decreased with increased delay. This showed that the availability of persistent image structure rendered performance independent of length of delay, whereas absence of image structure required dependence on memory-in-the-head, which yielded decreases in performance.

The number of targets identified increased significantly with number of targets available. The average number of hits with a 25-s delay and no blank was 9.2. In trials with 18 targets, 3 out of the 13 participants identified more than 12 targets on average, and the maximum number of targets identified was 18. These performance levels were much greater than the established upper bound of visual STM capacity, which was approximately 4 items.

When plotting the number of hits as a function of the number of targets available, the increase was linear with no asymptotic trend apparent (see Figure 7). The linear relation between number of targets and number of hits suggested that the number of targets recalled would continue to grow with an increase in the number of targets beyond 18. At this point, the size or even the existence of a maximum for the number of targets reliably recalled is unknown.

Figure 7 shows that the slope of the linear relation between hits and number of targets varied as a function of the blank/no blank manipulation and delay. We tested these trends using multiple regression to compare slopes and intercepts pairwise between conditions. First, we compared hits with respect to number of targets with 5-s and 25-s delays in the No Blank condition. Using trials from the No Blank condition and regressing mean hits onto the number of targets available, the linear trends for both levels of delay were significant (No Blank and 5-s Delay, $r^2 = 0.95$, $F[1, 2] = 37$, $p < .03$; No Blank and 25-s Delay, $r^2 = 0.99$, $F[1, 2] = 172$, $p < .01$); but comparing the two fitted lines, neither the slopes nor the intercepts were different. (We performed all these analyses both on the trial data and on the means and the results were all the same; we report analyses on the means.) The mean slope was 0.31. Then, we tested trials with the long delay and contrasted performance with no blank versus that with blank. When regressing mean hits on the number of targets in these conditions, the linear fits were both significant (No Blank and 25-s Delay, $r^2 = 0.99$, $F[1, 2] = 172$, $p < .01$; Blank and 25-s Delay, $r^2 = 0.94$, $F[1, 2] = 31.1$, $p < .05$). Additionally, although the intercepts were not different, the slopes were different, $F(1, 4) = 18.29$, $p < .02$. The slope for no blank and 25-s delay trials was 0.37, and that for blank and 25-s delay trials was 0.18. These implied that, with stable image structure, the embodied memory for perceived target locations was stable over long time delay. Thus, there was no difference between rates of increasing hits as a function of number of targets with short (5 s) or long (25 s) delay in the No Blank condition. However, a long delay of nearly half a minute did yield a significant decrease in this rate, cutting it in half, when continuous image structure was removed during the delay in the Blank condition. This difference was also reflected in the Blank/No Blank × Delay interaction in the ANOVA. Memory-in-the-head exhibits the classic memory decay or instability over time, whereas embodied memory exhibits remarkable stability over time.

General Discussion

Many of J. J. Gibson's insights about perception were developed in his analysis of the problem of progressive occlusion (Gibson, 1979/1986). As often noted (e.g., Chemero, 2009; Klatzky, Mac-

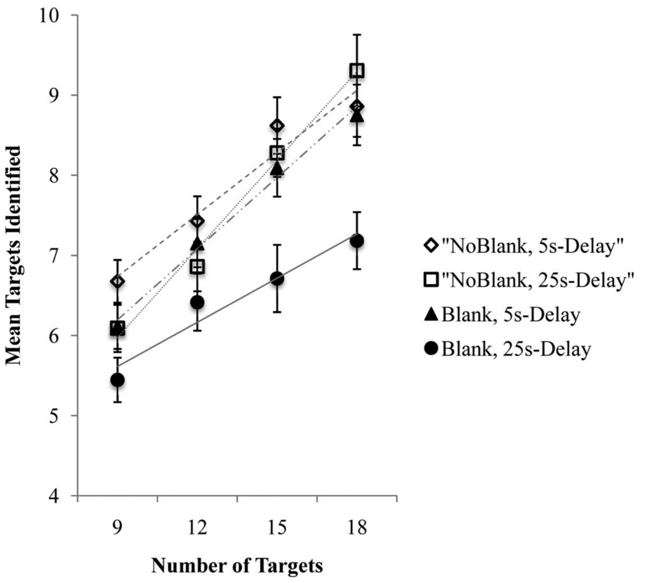


Figure 7. In Experiment 4, the number of hits increased with number of targets at a rate that varied as a function of the delay interval in the Blank condition, but not the No Blank condition. The rate fell by half (slope = 0.18) when delay increased from 5 to 25 s in the Blank condition, but remained stably larger in the No Blank condition (slope = 0.37). Error bar = ±1 SE.

Whiney, & Behrmann, 2008), Gibson's was an embodied approach. For Gibson, perception entailed a lawful relation between the perceiver and the environment, and thus the unit of analysis included structure in the surround and in the light, and the relevant structure was temporal as well as spatial. Optic flow was central to his analyses of information in light. However, time also played a role in a more memory-like way. His analysis of progressive occlusion entailed stability of perception. He argued that the perceiving of progressive occlusion logically entailed the perceiving of things that had become hidden from the observer's point of view. The logic of perceiving the progressive occlusion, as opposed to the progressive annihilation, of substantial surfaces required that the observer perceive the continued existence of surface elements that became hidden during the progressive occlusion event. Again, the perception of a hiding event, as such, requires that the elements that have become hidden from the observer's point of view be perceived as hidden, and thus continuing to exist beyond the occluding contour and thus behind the occluding surface. (Otherwise, when the relevant information—namely, deletion of optical texture along an edge—was detected, the perception would be of a different kind of event, an object annihilation event. But this does not happen.) Gibson did not address or pursue additional questions that arise naturally from his analysis, such as how long might the perceiver be expected to continue to perceive such hidden objects (just how stable might this perception be?) and can more than the mere existence of hidden objects be perceived, for instance, their location? A rigid motion taking surface elements out of view also provides potential information about where the hidden surface elements are located behind the occluding surface relative to the occluding contour.

In Gibson's analysis, the information for progressive occlusion is contained in optic flow (e.g., Gibson, 1979/1986). The question of stability becomes paramount when the optic flow ceases and only static image structure remains. Gibson argued that this was all perception, but the fact is that the stability entails an element of memory as well. It is indeed a kind of memory, but the unit of analysis has changed and the memory in question is not just in the head. The stability resides as much in the persistent image structure as in the perceiver. The stability is of information provided by the optic flow that calibrated the image structure (Lee & Bingham, 2010). So, the key is the combination of optic flow and image structure.

Results from these experiments demonstrated a powerful effect of this combination on identifying locations of multiple hidden objects. In these experiments, we showed that, without optic flow information, direction of translation could not be perceived, and hence the target identification task was impossible (Experiment 2, Instant Shift condition). When only optic flow information was available (Experiment 1), participants could do the task, but poorly. The small number of target locations participants could identify (less than two on average, and four or five at best) suggested that they were relying entirely on their visual STM, which, of course, they had to do because, by design, there was no useful static image structure that they could use. Visual STM has an extremely small capacity of about four items (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001; Wheeler & Treisman, 2002). Hence, although optic flow enabled the perception of 3D structure, the resulting perception was not stable once optic flow ceased in the absence of persisting image structure. When both image structure and optic flow were available, participants were able to identify more targets: 7.75 on average in the Continuous Shift condition of Experiment 2, and 7.26 on average in the No Blank condition of Experiment 3 (both with 5-s delays).

A comparison between performance in Version 2 of Experiment 1 to that in trials of Experiment 3 with a 5-s delay (the only difference between them being whether the windows had borders or, equivalently, the presence or absence of image structure) provided convincing evidence that combined optic flow and image structure information led to better performance than optic flow alone did. With a 5-s delay and equal number of targets and distracters in these experiments, the mean number of targets identified in Version 2 of Experiment 1 was 1.36 ($SD = 1.10$) and that in Experiment 3 was 7.09 ($SD = 2.85$). An omnibus ANOVA showed that the performance was significantly affected by the presence or absence of borders around the windows, $F(1, 14) = 197.00$, $p < .001$. Additionally, the number of targets and the interaction between "border" and number of targets were both significant, $F(3, 42) = 33.10$, $p < .001$; $F(3, 42) = 18.20$, $p < .001$, respectively. This reflected the fact that the number of targets identified only increased with number of targets available when the borders (i.e., image structure) were present.

In Experiment 4 of this study, we showed that when image structure was available, recalling multiple target locations in a 3D environment was relatively easy. The maximum number of targets that participants identified reliably was about 9 to 10 out of 18 targets. These results suggested that if 4 is the approximate number of items that can be held in visual STM, target identification in our study cannot be relying only on information stored in such memory. Hence, we argue that information yielded by optic flow is retained in image structure, which serves as reference for later recall. Being able to refer back to visible layout makes perception no longer subject to the capacity of memory-in-the-head, because it is now offloaded from the head to the environment. After being calibrated by optic flow, stable image structure allowed perception to surmount extended delays to remain effective after almost half a minute. This was seen in the significant interactive effect between blank and delay in Experiment 4, in which, given stable and continuous image structure (in the No Blank condition), performance did not deteriorate with an increase in delay to 25 s (as shown in Figure 6). Performance only deteriorated with extended delay when continuous image information was absent (in the Blank condition). This Blank × Delay interaction was a critical finding supporting our hypothesis that perceived spatial layout (that is, location of target objects) could be offloaded from memory and preserved in external image structures, making it accessible after long delays. Without such external image structure, perceived objects and their spatial relations had to be kept in memory and performance decreased with extended delay.

In Experiment 3, we did not find either an effect of delay or an interaction between blank and delay, although blank itself yielded a significant effect (and the data exhibited a trend for an effect of delay in the blank condition). The displays in that experiment allowed potentially greater use of configurality in performing the task. In addition, the delays (5 s, 10 s, and 15 s) were less than those tested in Experiment 4 (that is, 5 s and 25 s), which yielded a significant Blank × Delay interaction. It remains unclear whether configurality or the shorter delays, or both, were responsible for the lack of an effect of delay in Experiment 3. Ultimately, configurality is a natural part of embodied memory. Investigation of these aspects of the current results remains for future efforts.

Normally, perceiving objects per se is not the goal of perception. Instead, the goal is to guide and control actions (Milner & Goodale, 2006). Because animals frequently act upon information detected at some earlier time, control of action would be highly ineffectual and unreliable if it only relied on the capabilities of visual STM. Under representative conditions, this is rarely required. For example, when we are sorting a deck of cards into piles based on their suits, we usually do not need to remember exactly where the pile for each suit is located. Instead, we look at the piles each time when we are about to put down a card. In fact, past studies have suggested that in doing this kind of task, individuals constantly look back at what is around them for reliable "just-in-time" information to guide their ongoing actions (Droll & Hayhoe, 2007). The just-in-time information is what the environment can provide us through stable image structure.

To sum up, in an environment where motion-generated optic flow specifies the relations between objects, and subsequent image structure remains undisrupted, the 3D layout of the environment is readily perceived, and such perception is preserved externally by the stable image structure. This embodied memory is highly effective, allowing performance to exceed the limit otherwise imposed by human memory capacities.

Results from the current study and those from Lee and Bingham's (2010) study showed that maximally effective perception of objects occured when image structure and optic flow were both present and allowed to interact. This has significant implications for understanding visual system function and performance. Traditionally, vision has been studied using an image-based (constructivist) approach or using an optic-flow-based (ecological) approach. The original two visual systems hypothesis (Milner & Goodale, 1995) instantiated the two theories in an anatomical account (Norman, 2002). Specifically, the hypothesis suggested that neither image-based nor optic-flow-based vision can explain visual perception by itself. Instead, image-based vision and optic-flow-based vision were characterized as two streams in the visual system that yielded the functions of the ventral system and dorsal system, namely, object recognition and guidance of action. The current study, Lee and Bingham (2010), and Lee et al. (2012) showed that in both object identification and guidance of action, visual perception uses both image structure and optic flow. They are not functionally separate or independent. To the contrary, they are mutually facilitative, as the best object identification and the most accurate actions both occur when optic flow calibrates image structure that then preserves optic flow information. In fact, the two visual system hypothesis has been revised (Milner & Goodale, 2006) to bring together images and optic flow in a reconfigured set of anatomical streams, a revision that is especially appropriate given the results in the current studies.

Conclusion

In four experiments, using a paradigm that allowed us to control optic flow and image structure, we showed that effective perception is based on using both. Optic flow provides an immediate depth map of 3D layout, including the location of objects that become hidden, and this information is preserved in image structure that has been calibrated by the preceding optic flow. Offloading the transient optic flow information to external stable image structure allows individuals to access and act upon information provided by optic flow without having to hold it all in the head.

References

Anderson, J., & Bingham, G. P. (2010). A solution to the online guidance problem for targeted reaches: Proportional rate control using relative disparity. *Experimental Brain Research, 205,* 291–306. doi:10.1007/s00221-010-2361-9

Anderson, J., & Bingham, G. P. (2011). Locomoting-to-reach: Information variables and control strategies for nested actions. *Experimental Brain Research, 214,* 631–644. doi:10.1007/s00221-011-2865-y

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147. doi:10.1037/0033-295X.94.2.115

Bingham, G. P., & Lind, M. (2008). Large continuous perspective transformations are necessary and sufficient for accurate perception of metric shape. *Perception & Psychophysics, 70,* 524–540. doi:10.3758/PP.70.3.524

Brenner, E., & van Damme, W. J. M. (1999). Perceived distance, shape, and size. *Vision Research, 39,* 975–986. doi:10.1016/S0042-6989(98)00162-X

Carr, H. A. (1935). *An introduction to space perception*. Oxford, UK: Longmans, Green.

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, MA: MIT Press.

Domini & Caudek, C. (2003). 3-D structure perceived from dynamic information: A new theory. *Trends in Cognitive Sciences, 7,* 444–449.

Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1352–1365.

Edelman, S., & Bulthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research, 32,* 2385–2400. doi:10.1016/0042-6989(92)90102-O

Fajen, B. R. (2005a). Calibration, information, and control strategies for braking to avoid a collision. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 480–501.

Fajen, B. R. (2005b). The scaling of information to action in visually guided braking. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1107–1123.

Fajen, B. R. (2008). Learning novel mappings from optic flow to the control of action. *Journal of Vision, 8,* 1–12.

Fajen, B. R., & Warren, W. H. (2003). Behavioral dynamics of steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 343–362.

Gibson, J. J. (1950). *The perception of the visual world*. Oxford, UK: Houghton Mifflin.

Gibson, J. J. (1979/1986). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Hu, Y., Eagleson, R., & Goodale, M. A. (1999). The effects of delay on the kinematics of grasping. *Experimental Brain Research, 126,* 109–116. doi:10.1007/s002210050720

Hummel, J. E., & Biederman, I. (1992). Dynamical binding in a neural network for shape recognition. *Psychological Review, 99,* 480–517. doi:10.1037/0033-295X.99.3.480

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research, 31,* 1351–1360. doi:10.1016/0042-6989(91)90056-B

Kaplan, G. A. (1969). Kinetic disruption of optical texture: The perception of depth at an edge. *Perception & Psychophysics, 6,* 193–198. doi:10.3758/BF03207015

Klatzky, R. L., MacWhiney, B., & Behrmann, M. (2008). *Embodiment, ego-space, and action*. New York, NY: Psychology Press.

Lee, Y. L., & Bingham, G. P. (2010). Large perspective changes yield perception of metric shape that allows accurate feedforward reaches-to-grasp and it persists after the optic flow has stopped! *Experimental Brain Research, 204,* 559–573. doi:10.1007/s00221-010-2323-2

Lee, Y. L., Crabtree, C. E., Norman, J. F., & Bingham, G. P. (2008). Poor shape perception is the reason that reaches-to-grasp are visually guided online. *Perception & Psychophysics, 70,* 1032–1046. doi:10.3758/PP.70.6.1032

Lee, Y. L., Lind, M., Bingham, N., & Bingham, G. P. (2012). Object recognition using metric shape. *Vision Research, 69,* 23–31. doi:10.1016/j.visres.2012.07.013

Li, L., & Warren, W. H., Jr. (2000). Perception of heading during rotation: Sufficiency of dense motion parallax and reference objects. *Vision Research, 40,* 3873–3894. doi:10.1016/S0042-6989(00)00196-6

Li, L., & Warren, W. H., Jr. (2002). Retinal flow is sufficient for steering during observer rotation. *Psychological Science, 13,* 485–490. doi:10.1111/1467-9280.00486

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19,* 577–621. doi:10.1146/annurev.ne.19.030196.003045

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. doi:10.1038/36846

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action* (1st ed.). New York, NY: Oxford University Press.

Milner, A. D., & Goodale, M. A. (2006). *The visual brain in action* (2nd ed.). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780198524724.001.0001

Milner, A. D., Paulignan, Y., Dijkerman, H. C., Michel, F., & Jeannerod, M. (1999). A paradoxical improvement of misreaching in optic ataxia: New evidence for two separate neural systems for visual localization. *Proceedings of the Royal Society of London: Series B: Biological Sciences, 266,* 2225–2229. doi:10.1098/rspb.1999.0912

Nakayama, K., & Loomis, J. M. (1974). Optic velocity patterns: Velocity-sensitive neurons and space perception. *Perception, 3,* 63–80. doi:10.1068/p030063

Norman, J. (2002). Two visual systems and two theories of perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences, 25,* 73–96.

Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics, 53,* 279–291. doi:10.3758/BF03205183

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3,* 1199–1204. doi:10.1038/81479

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. doi:10.1016/S0960-9822(07)00492-7

Simpson, W. A. (1993). Optic flow and depth perception. *Spatial Vision, 7,* 35–75.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychological Bulletin & Review, 2,* 55–82. doi:10.3758/BF03214412

Tarr, M. J., & Bulthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition, 67,* 1–20. doi:10.1016/S0010-0277(98)00026-2

Tarr, M. J., & Vuong, Q. C. (2002). Visual object recognition. In H. Pashler & S. Yantis (Eds.), *Stevens' handbook of experimental psychology. Volume 1: Sensation and perception* (pp. 283–314). New York, NY: Wiley.

Tittle, J. S., & Braunstein, M. L. (1993). Recovery of 3-D shape from binocular disparity and structure from motion. *Perception & Psychophysics, 54,* 157–169. doi:10.3758/BF03211751

Todd, J. T. (1995). The visual perception of three-dimensional structure from motion. In W. Epstein & S. J. Rogers (Eds.), *Perception of space and motion* (pp. 201–226). New York, NY: Academic Press.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 92–114.

Wann, J., & Land, M. (2000). Steering with or without the flow: Is the retrieval of heading necessary? *Trends in Cognitive Sciences, 4,* 319–324. doi:10.1016/S1364-6613(00)01513-8

Warren, W. H. (2008). Optic flow. In A. I. Basbaum, A. Kaneko, G. M. Shepherd, & G. Westheimer (Eds.), *The senses: A comprehensive reference* (Vol. 2, pp. 219–230). San Diego: Academic Press.

Warren, W. H., & Hannon, D. J. (1990). Eye movement and optic flow. *Journal of the Optical Society of America A: Optics, Image & Science, 7,* 160–169. doi:10.1364/JOSAA.7.000160

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131,* 48–64.

Wilkie, R. M., & Wann, J. P. (2006). Judgments of path, not heading, guide locomotion. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 88–96.

Yilmaz, E. H., & Warren, W. H., Jr. (1995). Visual control of braking: A test of the tau-dot hypothesis. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 996–1014.

Received July 25, 2012
Revision received January 23, 2013
Accepted January 28, 2013 ■