applied optics

Breaking camouflage and detecting targets require optic flow and image structure information

Jing Samantha Pan,[1,*] Ned Bingham,[2] Chang Chen,[1] and Geoffrey P. Bingham[3]

[1]Department of Psychology, Sun Yat-sen University, Guangzhou, Guangdong 510006, China
[2]School of Electrical and Computer Engineering, Cornell University, Ithaca, New York 14850, USA
[3]Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47404, USA
*Corresponding author: panj27@mail.sysu.edu.cn

Received 24 January 2017; revised 30 June 2017; accepted 7 July 2017; posted 10 July 2017 (Doc. ID 285570); published 1 August 2017

Use of motion to break camouflage extends back to the Cambrian [In the Blink of an Eye: How Vision Sparked the Big Bang of Evolution (New York Basic Books, 2003)]. We investigated the ability to break camouflage and continue to see camouflaged targets after motion stops. This is crucial for the survival of hunting predators. With camouflage, visual targets and distracters cannot be distinguished using only static image structure (i.e., appearance). Motion generates another source of optical information, optic flow, which breaks camouflage and specifies target locations. Optic flow calibrates image structure with respect to spatial relations among targets and distracters, and calibrated image structure makes previously camouflaged targets perceptible in a temporally stable fashion after motion stops. We investigated this proposal using laboratory experiments and compared how many camouflaged targets were identified either with optic flow information alone or with combined optic flow and image structure information. Our results show that the combination of motion-generated optic flow and target-projected image structure information yielded efficient and stable perception of camouflaged targets.

OCIS codes: (330.5020) Perception psychology; (330.5000) Vision - patterns and recognition; (330.1880) Detection.

https://doi.org/10.1364/AO.56.006410

1. INTRODUCTION

In nature, animals evolve to adopt appearances that help them camouflage as a protection mechanism. According to zoologists Stevens and Merilaita's taxonomy [1], there are two classes of camouflage where animals alter their static appearances to disguise themselves [2]. In one way, visual targets look similar to their surroundings by matching the luminance, texture, and/or color, such that the detection of targets becomes hard for a distant observer with weak disparity cues [3]. For example, a draco lizard shows markings that are nearly identical to the bark of the tree it is on. This is the process of *crypsis*. Alternatively, targets can be disguised by resembling other uninteresting objects commonly found in a given environment, such that the recognition of targets becomes difficult. For example, some nocturnal fishes in the Amazon look like fallen leaves in streams during daylight [4]. This is the process of *masquerade*. Either by crypsis or by masquerade, many animals evolve to have appearances suitable for concealing themselves in their natural habitats. Thus, hunting predators often must break camouflage to distinguish targets from distracters, which is a crucial skill for their survival. In this study, we mainlyinvestigate how observers may perceive crypsised targets with two experiments and we briefly discuss how observers may perceive targets with masquerade.

In the case of crypsis, by definition, targets and distracters (or environmental backgrounds) are indistinguishable based on static appearances (equivalently, image structure information). However, targets and distracters are real objects that occupy unique spatial locations and hence their depth relations and relative motion properties with respect to a moving observer are different. For example, when a flounder is resting on the seabed, its appearance enables it to blend in with the surrounding environment so well that it is extremely difficult to spot the fish from above the water surface. However, when it moves, it becomes easily identifiable, even against a similar looking background. Thus, to enable camouflage, a target stays still and intricately manipulates its appearance to mask the spatial (depth) distinction and to blend in with the environment; to break camouflage, an observer therefore needs to undo the masking and regain veridical spatial relations of objects in the environment, probably via motion, which segregates the scene into figure and ground [5,6].

Although it is easier to detect camouflaged targets when there is relative motion, to increase the success rate of a targeted action, the perception must persist after motion stops because it is much easier to capture an unmoving target. In other words, detected targets needs be preserved in the absence of the information that originally specifies the targets (i.e., information that accompanies relative motion) and so, the locations must be remembered over the course of action. Thus, successful hunting relies on visual perception that is efficient (i.e., accurately detecting a large number of camouflaged targets with motion) and stable (i.e., to continue perceiving unmoving targets for acquisition). We propose that efficient and stable perception of crypsised targets relies on the interaction of two sources of optical information, which is available to a moving observer, namely, optic flow and image structure information.

In an environment populated by objects with opaque surfaces, light is reflected from and therefore deterministically structured by surfaces surrounding the observer. The structured light available to an observer is described as the optic array. Motions of the observer and/or surfaces in the environment yield continuous and lawful changes in the optic array known as optic flow [7]. Optic flow is structured by motions and corresponds to and specifies these motions. The speed of optic flow covaries with the distance and direction (angular displacement) between surfaces and the point of observation [8]. Therefore, detection of optic flow patterns allows the observer to become aware of the 3D spatial relations of surfaces in the environment. For example, continuous relative motion between an observer and world objects informs him/her about the object's coherent 3D forms or shapes (a process known as structure from motion) [9–12], the layout of surfaces in cluttered terrain [12,13], and the relative moving speeds and directions of objects [14]. All of these require minimum image-based information and sufficient optic flow. Therefore, perceiving 3D relations and separating figures from the ground should be possible with strong optic flow and indistinguishing image appearances, as in the context of camouflage.

Although optic flow contains strong and immediate information in specifying 3D spatial layout, this information is ephemeral. It varies in quality with the relative speeds of motion and becomes unavailable when motion stops. While an observer remains still and thus, without optic flow, must he or she retain all previously detected optic flow information about the surroundings strictly in the head? Perhaps not, given the availability of the other source of optical information, namely, image structure, and its relation to optic flow.

Image structure refers to static optical structures or patterns that are projected to the eye by surfaces [15,16] and characterized in terms of color, contour, contrast edges, and the like. In the case of crypsis, by definition, image structures for visual targets and surrounding objects are very similar, which renders target identification based on image structure alone extremely difficult. However, image structure is stable. It remains available to an observer as long as objects are present.

Optic flow and image structure are both available to a moving observer. They are two sources of optical information that are simultaneously projected to the observer from the same surfaces in the surroundings (They are said to exhibit an essential symmetry. "Symmetry" is a synonym for "sameness"; see Ref. [17]). Although image structure is weaker in specifying depth relations, it is stable. 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 (one image "flows" to the next), optic flow and image structure are intrinsically related and largely symmetric with respect to the layout of surfaces in the world (i.e., the surfaces that project image structure and generate optic flow when moving). In part, the relation could be cast as a calibration of image based information about 3D structure by the more powerful optic flow information. Optic flow specifies the changes in 3D spatial structure that, in turn, relates sequential images. Furthermore, once the optic flow has ceased, the static images remain and preserve the information provided by the optic flow. Offloading the information provided by transient optic flow to external stable image structure would allow observers 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 a storage system of perceived world properties that is outside of the mind (but interacts with it). It allows situated, active observers to get access to perceived world properties in real time.

The combined and interacting optic flow and image structure information has been shown to lead to the effective perception of object shapes [18], object locations [19,20], and more complex natural events [19]. In these studies, performance based on perception using both optic flow and image structure was always superior to performance based on either source of information in isolation.

Particularly, Pan, Bingham, and Bingham studied the perception of target locations, amid distracters, after targets had become progressively occluded [19]. The authors manipulated the optical information available (optic flow only, image structure only, or both), the number of targets, the delay between perceiving and identifying targets, and whether image structure was available during the delay or not. The important results were that when (and only when) both optic flow and image structure were available (1) a large number of hidden targets were accurately identified (≈10 out of 18 targets on average), and (2) identification performance was stable over long time delays (in their experiments, 25 s). In the latter case, when image structure was not available during the delay period, performance exhibited classic memory decay. The lack of decay (that is, the lack of a reduction in the number of targets successfully identified as the length of the delay increased) when image structure remained available suggested that something other than internal memory must have been used to facilitate stable perception. Given the symmetry between optic flow and image structure, the authors argued that spatial layout was perceived using optic flow and this was offloaded to and preserved in the external static image structure, which remained after the optic flow ceased. This embodied system, particularly the taking of information from the external world in real time (in contrast to relying on memory-in-the-head), allowed observers to perceive locations of many targets for an extended period of time after motion stopped.

In the above-mentioned study, targets and distracters were on the same depth layer but carried different image structures (i.e., targets were pink and distracters were gray). What if targets and distracters have the same image structure but occupy different spatial locations? In this case, the targets are crypsised and undistinguished from the surrounding distracters based on image structure alone. Hence, might the optic flow-image structure synergy enable target identification and, if so, how efficient and stable might this be?

To study the perception of camouflaged targets, we designed a paradigm that allowed us to manipulate the availability of optical information. In Experiment 1, there was only optic flow information, and in Experiment 2, there was both optic flow and image structure information, with image structure being identical between targets and distracters. We expected higher efficiency and stability of perception in Experiment 2, where optic flow and image structure were available and interacting.

2. EXPERIMENT 1

Experiment 1 was designed to test how many camouflaged target objects observers could identify when only optic flow was available without static image structure.

A. Methods

1. Participants

Ten adults (five males and five females, aged between 20 and 35) participated in Experiments 1 and 2. They took self-determined breaks of variable length between experiments. The order of the experiments was counterbalanced across the participants. All participants had normal or corrected-to-normal vision. Participants were paid $7 per hour for completing the experiments. A performance-based bonus was given in addition. All participants signed informed consent in accordance with the procedures approved by the Indiana University Institutional Review Board.

2. Apparatus

Participants sat in front of a 20” LCD computer monitor with a viewing distance of 50 cm. The refresh rate of the monitor was 60 Hz and the spatial resolution was 1680 × 1050.

3. Stimuli

A simulated 3D display was presented on the computer monitor. The display consisted of a randomly textured background (each pixel in the texture was assigned a random RGB value from 0 to 255 for each component), which extended well beyond the edges of the computer screen so that its edges never appeared on screen during the display. In front of the background (i.e., closer to the observer along his or her line of sight), there were small squares (side = 7 mm or approximately 0.5° visual angle) filled with identical random texture as the background. These squares were the targets and they were all on the same depth layer. The size of the visible part of the background was approximately 45° (W) × 30° (H) visual angle. The possible locations where a target could appear covered the central 2/3 of the background or approximately 30° (W) × 20° (H). Participants were clearly instructed that the squares closer to them at the beginning of each trial were targets to be perceived and identified. The number of targets varied as an experimental manipulation. Because there was no distinct image structure, when the display was stationary, the targets were not distinguishable from the background. In other words, participants saw a flat surface, filled with random texture, and perpendicular to their line of sight. However, when the surfaces rigidly rotated (with the relative spatial relations between the layers unchanged), structure from motion (SFM) occurred. Due to the depth differences, the flow was faster for the targets (the closer surfaces) and slower for the background (the farther surface), yielding progressive occlusion of portions of the rear surface by the front surfaces. This allowed participants to perceive the depth relations in the display and hence identify the spatially defined targets.

Next, the SFM display stopped and the two layers (that is, the front layer containing all targets and the background layer) were again oriented perpendicular to the line of sight. Then, targets on the front layer translated along the line of sight (i.e., moved farther away from the observer) and stopped when they matched the distance of the background layer (i.e., the targets were as far from the observer as the background). At the end of the translation, the display remained stationary and contained no visual information separating targets from the background; see Fig. 1(a). The task was to use the mouse and click on where the targets were in the field of random textures. (Readers may download and try the experimental display at http://www.indiana.edu/~palab/research.php, click to expand the tab “Perception and Embodied Memory,” and then “Experiment Demos.” The relevant demos are Demos 4 and 5, which correspond to Experiments 1 and 2 in this paper, and Demo 6, which corresponds to the masquerade experiment that is discussed in Section 4. The demos run on Mac OS only. Motion speeds may vary depending on the system settings. Non-Mac users may also find a video on the same page as a quick reference of the experimental display. Link to video: http://www.indiana.edu/~palab/Resources/Demos/camouflage_vid_demo_short2.mp4.

We encouraged participants to click accurately, instead of guessing, by introducing a point system: starting with 200 points, if they clicked on a target correctly (that is, a “hit”), they gained a point; if they clicked incorrectly (that is, the click fell on a pixel outside of targets yielding a “false alarm”), they lost a point; if they did not click, there would be no point change. At the end of the experiment, participants received a 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 two experiments, there were very few false alarms (i.e., participants rarely identified nontargets as targets). The median of false alarms in each blank-delay condition of the experiments was zero. This means that in all conditions tested, there were no false alarms in more than half of the trials. The mean number of false alarms in each condition was no more than 1, after removing outliers, which were defined as beyond mean ±2 standard deviations. (We removed the outliers before we calculated the means because the distribution of false alarms was skewed. For this reason, the medians were better measures of false alarms.) The extremely small number of false alarms in these experiments suggested that participants were careful and conservative when making responses. They might have mis-remembered, but they did not guess. Therefore, we analyzed the number of targets identified correctly (that is, hits) as a measure of perception performance in these experiments.

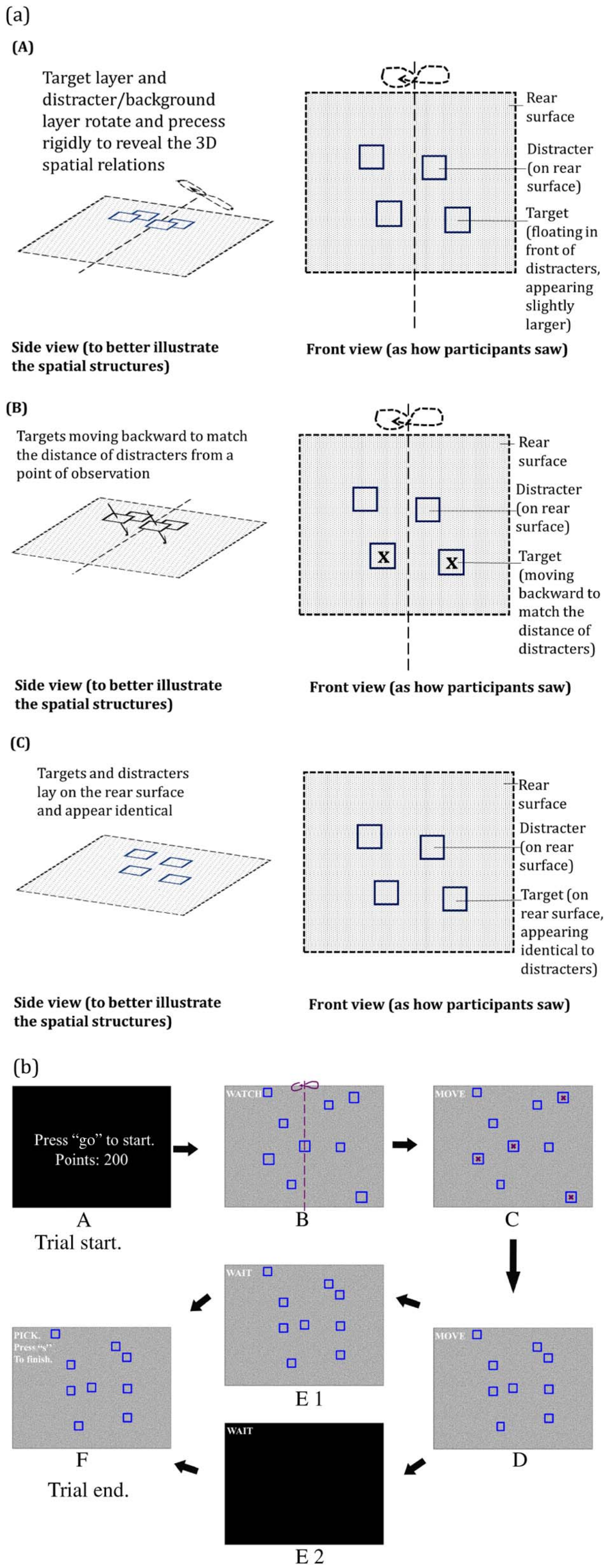
(a)
(A)
Target layer and distracter/background layer rotate and precess rigidly to reveal the 3D spatial relations
Rear surface
Distracter (on rear surface)
Target (floating in front of distracters, appearing slightly larger)
Side view (to better illustrate the spatial structures)
Front view (as how participants saw)
(B)
Targets moving backward to match the distance of distracters from a point of observation
Rear surface
Distracter (on rear surface)
Target (moving backward to match the distance of distracters)
X
X
Side view (to better illustrate the spatial structures)
Front view (as how participants saw)
(C)
Targets and distracters lay on the rear surface and appear identical
Rear surface
Distracter (on rear surface)
Target (on rear surface, appearing identical to distracters)
Side view (to better illustrate the spatial structures)
Front view (as how participants saw)
(b)
Press "go" to start.
Points: 200
A
Trial start.
WATCH
B
MOVE
C
MOVE
D
WAIT
E 1
WAIT
E 2
PICK.
Press "s"
To finish.
F
Trial end.

Fig. 1. (Continued)

4. Procedures

Participants read and signed consent forms and then sat in front of the test computer. They adjusted the seat height so that their eyes were aligned with the center of the computer screen. The experimenter explained the task and instructed the participants to click accurately on the small squares that appeared in the front (or closer to them). Then, participants attempted 3–10 practice trials in the presence of the experimenter to become familiar with the task.

At the beginning of each trial, participants saw a screen showing their current points and the instruction to press the "S" key to begin. Then the targets and the background rigidly rotated for 9.5 s, allowing participants to perceive the depth relation in the display through SFM and hence to study locations of targets. Afterwards, the front targets translated toward the background surface along the z direction. The translation took 1.5 s, after which the targets and background occupied the same depth layer. During translation, because targets were the only structures that moved, it was still possible to identify them. This made the total study time for remembering targets 11 s. After the translation phase, participants waited for either 1 or 5 s (the factor of "delay") before the mouse cursor appeared on screen for them to click on targets. During the delay, participants either continued to see the field of random textures or saw a black screen. These were the "No Blank" and "Blank" conditions, respectively; see Fig. 1(b). In each trial, there were 6, 9, 12, or 15 targets to be identified. The combination of the factors of delay, blank, and number of targets yielded 16 unique conditions, and all participants completed four repetitions per condition (or 64 trials in total) in one session.

Fig. 1. (a) An illustration of the experimental display (not drawn to scale). A front view (which is what participants actually saw) is shown on the right and a side view (which is for clarifying the 3D structure) is shown on the left. (A) SFM phase: two layers were separated in depth. Small textured squares floating in the front layers are targets. The background layer is textured identically and forms/contains distracters. The two surfaces rigidly rotate and precess in depth (or make figure-eight movement around a vertical axis in the frontoparallel plane). This motion reveals the 3D structure and allows spatially defined targets to be differentiated from the background. (B) Translation phase: targets move backward until they are at the same depth from an observer as distracters are. (C) Eventually, targets stop moving and stay on the same depth plane with the background/distracters. Participants click on targets from this display. (b) An illustration of experimental procedures (not drawn to scale). (A) A trial started with a black screen showing the current points and an instruction to begin. Then the background (containing distracters) and the target planes rotated and precessed rigidly around a center axis and revealed the 3D structures (SFM). (B) This allowed an observer to distinguish targets (defined as squares closer to the observer) from distracters and the background. (C) Then the front plane was translated backward until all targets coincided with the background. As a result of the translation in depth, target image sizes reduced to be equal to distracter image sizes, and intertarget spacing was reduced. (D) Consequently, the patterns formed by visible squares on screen, including both targets and distracters, were changed. A delay followed, during which the observer either continued to see (E1) the ending scene of the display or (E2) a black screen. (F) Finally, participants were prompted to click on targets given the ending scene of the display.

Participants also did a control condition after they completed the experimental condition, where they pointed at the targets when the two surfaces were moving, which lasted 11 s per trial. An experimenter was present to note down how many targets the participants correctly identified. The same four levels of targets were tested at four repetitions per target level (16 trials total). The delay or blank factors were irrelevant in this control condition because this was a test of whether observers could see the targets with optic flow information alone. It was a test of perception efficiency with ongoing motion-generated information; stability was not tested.

In Experiment 1, there were no visible borders around target or distracters squares, that is, no image structure information. In Experiment 2, there were blue borders outlining both targets and distracters (as shown in the drawing of this figure). In the thought experiment mentioned in Section 4, borders of targets and of distracters would be of different colors.

B. Results

In the control condition, with ongoing motion and the optic flow it generated, participants accurately and effortlessly identified all targets, up to the maximum of 15 that was tested. In the experimental condition, participants identified much fewer targets after motion stopped (mean = 0.95, SD = 0.84). To test the effects of blank and delay on identification performance, an omnibus repeated-measures ANOVA was performed on data collected from the experimental conditions. It showed that hits were significantly affected by blank [$F(1,9) = 27.7, p < 0.001, \eta^2 = 0.42$], delay [$F(1,9) = 28.4, p < 0.001, \eta^2 = 0.13$], and their interaction [$F(1,9) = 8.83, p < 0.02, \eta^2 = 0.03$]. More targets were identified in trials without the blank screen inserted between perceiving and recalling ($\text{mean}_{\text{NoBlank}} = 1.30$, $\text{SD}_{\text{NoBlank}} = 0.76$; $\text{mean}_{\text{Blank}} = 0.59$, $\text{SD}_{\text{Blank}} = 0.77$), and more hits occurred in trials with shorter delays than with longer delays ($\text{mean}_{\text{ShortDelay}} = 1.11$, $\text{SD}_{\text{ShortDelay}} = 0.82$; $\text{mean}_{\text{LongDelay}} = 0.78$, $\text{SD}_{\text{LongDelay}} = 0.82$).

Overall, when participants had to identify targets after motion stopped (experimental condition), the performance was poor regardless of blank and delay. In this case, the maximum number of targets identified was 4. This occurred in 4 out of 640 trials; see Fig. 2. Furthermore, hits were not affected by the number of targets available [$F(3,27) = 2.6, p = 0.08$]. In other words, no matter how many targets were present, hits were low and capped at four items. This number was even smaller than the suggested capacity for visual short-term memory (VSTM) [21–23]. We postulate 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. The contrast of performance between the experimental and control conditions suggested that perception based on optic flow information alone was possible but highly unstable. Thus, it is ineffective once motion ceased and target identification worsened with increased delay.

3. EXPERIMENT 2

In Experiment 2, we tested the stability of perception given both optic flow and image structure information. Optic flow, as in Experiment 1, was generated in the SFM and rigid translation phases. Image structure was provided by adding visible borders around targets. Identical borders were drawn on the background layer, forming outlined distracter squares with the same appearances as the targets.

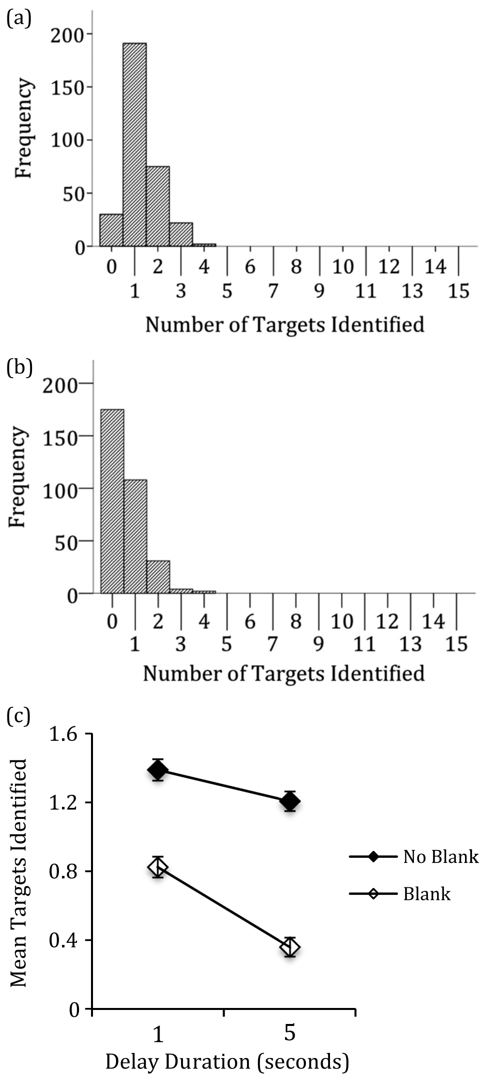

Fig. 2. Histograms displaying the frequency distribution of the number of targets identified in (a) no Blank and (b) blank conditions in Experiment 1. Hits were low regardless of the blank or delay. (c) However, there was a significant main effect of blank and interaction between blank and delay.

A. Methods

1. Participants

The ten participants who performed Experiment 1 also completed this experiment on the same day.

2. Apparatus

The same computer monitor used in Experiment 1 was used in this experiment.

3. Stimuli and Procedures

In this experiment, the display and procedures were similar to those in Experiment 1, except that blue square-shaped contours were drawn around targets in the front layer, and identical blue squares were also drawn on the back layer forming distracters; see Fig. 1. Targets and distracters were 0.5° × 0.5° visual angle in size. They were located on different depth layers, with distracters being farther away from an observer, and targets being closer to the observer. There were no depth discrepancies among targets or among distracters. At the beginning of the trials, because targets were located in a layer that is closer to the observer, they appeared slightly larger. The size difference was not noticeable because the squares were small to begin with and the distance between the two depth layers was small relative to the viewing distance. This was especially true due to foreshortening and perspective variation that occurred during motion. During the response phase, because the targets had moved back to be on the same depth plane as distracters, their sizes were identical.

Just as in Experiment 1, the SFM lasted for 9.5 s and the translation of targets lasted for 1.5 s. The translation in depth made both the target size and, more noticeably, intertarget spacing appear to shrink (from optical compression generated by retreating surfaces). Consequently, at the end of the translation, targets and distracters were of the same image size and the same distance in depth from the observer. Moreover, as the spacing between targets changed and distracters were now mixed in, the configuration previously formed by targets in the front layers was perturbed. This made it difficult to identify targets by remembering the overall patterns formed by them.

In this experiment, we manipulated the delay duration (5 s or 25 s), whether there was image structure continuously visible during delay (No Blank versus Blank), and the number of targets (9, 12, 15, or 18). The number of distracters was kept at 12. Each participant completed four repetitions per combination of conditions or 64 trials in total. The dependent measure was the number of targets correctly identified (i.e., hits).

B. Results

With both optic flow and image structure information available, participants identified more targets (mean = 8.23, SD = 2.78) as compared to in Experiment 1. A repeated-measures ANOVA testing the effects of blank (2 levels), delay (2 levels), and number of targets (4 levels) on hits showed that performance in this experiment was affected by the number of targets $[F(3, 27) = 32.0, p < 0.001, \eta^2 = 0.32]$, blank $[F(1, 9) = 10.7, p < 0.001, \eta^2 = 0.07]$, and blank-delay interaction $[F(1, 9) = 32.4, p < 0.001, \eta^2 = 0.04]$.

The perturbation of continuity of image structure information had an effect on hits. Specifically, more targets were identified in the No Blank condition than in the Blank condition ($\text{mean}_{\text{NoBlank}} = 8.68$, $\text{SD}_{\text{NoBlank}} = 2.58$; $\text{mean}_{\text{Blank}} = 7.77$, $\text{SD}_{\text{Blank}} = 2.90$). Performance was better with continuous, unperturbed image structure than without. More interestingly, hits dropped with extension of delay only in the Blank condition but not in the No Blank condition; see Fig. 3(a). This showed that the availability of persistent, calibrated image structure rendered performance independent of length of delay, while its absence led to the reliance on memory-in-the-head, which yielded a decrease in hits as delay extended. The significant interaction effect suggested that calibrated image structure contributed to stable perception performance.

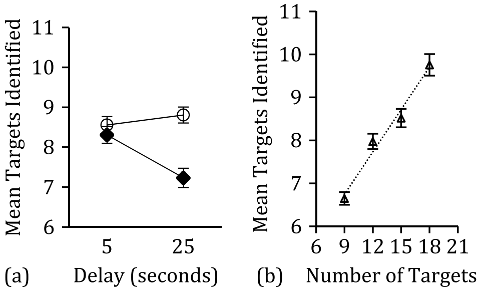

Fig. 3. (a) Interactive effect of blank and delay in Experiment 2. Open circles represent performance in the No Blank condition; filled diamonds represent performance in the Blank condition. (b) In Experiment 2, the number of targets identified (hits) increased linearly with number of targets available. In both graphs, error bars = 1 SE.

Collapsing across all other conditions, the mean number of targets identified increased linearly with the number of targets available, at the rate of 0.33 hits per target added, and up to 18 targets $[F(1, 2) = 87.9,\ p < 0.01,\ r^2 = 0.97,$ Fig. 3(b)]. Within the tested range, there was no observable asymptotic trend whether we looked at hits collectively or separately in each blank-delay condition (that is, Blank or No Blank paired with short or long delay). On average, 9.8 out of 18 targets were identified. Thus, performance in this experiment was much greater than the established upper bound of visual short-term memory capacity, about four items. At this point, the maximum number of targets that could be reliably recalled, as a measure of the efficiency capacity for the interactive system formed with optic flow and image structure, remained unknown.

4. GENERAL DISCUSSION

Typically, the goal of perception is not to detect and identify targets per se but to guide and control actions [7,24,25]. For a hunter, successfully detecting camouflaged targets is a means to the end of capturing them. To achieve this goal, it often requires the perception of target properties (such as spatial layout, location, and depth) to be temporally stable. Results from this study showed that accurate, efficient (in terms of identifying large numbers of targets), and stable (in terms of maintaining what has been perceived) perception of camouflaged objects in a 3D space requires optic flow and image structure information, both of which are available in a natural viewing environment.

In Experiment 1, when participants had to identify spatially defined targets, which were crypsised in an identically textured background, using optic flow information alone (no image structure information), they could perform the task but not very well. In the control group, participants identified all targets with ongoing optic flow, but in the experimental group, participants identified few targets after optic flow ceased (mean ≈ 1; max = 4, regardless of number of targets available). This contrast reflected a retention problem, not a detection problem. As commonly reported by participants, they were able to see all the targets when they were moving but were unable to see them after motion stopped. During post-test debriefing, we asked the participants whether they used any tactic to accomplish the task. They reported staring at a texture element and focusing on a few targets around it. In other words, participants used their VSTM, and this strategy was indeed reflected in the data: in the No Blank condition, where the textured screen was continuously available to the participants, hits were higher than in the Blank condition, where the texture elements were removed during delay. Relying on the extremely vulnerable VSTM is of course what participants had to do because, by design, there was no useful static image structure that they could otherwise use. The VSTM has an extremely small capacity of about four items [21–23], and like other forms of memory-in-the-head, it decayed with time. Hence, although optic flow enabled the perception of camouflaged targets, once it ceased, the target identification performance was poor in the absence of persisting image structure.

When both optic flow and image structure were available in Experiment 2, identifying multiple targets from distracters, despite their identical appearances, was quite efficient and stable. Regardless of delay duration and with up to 18 targets, the number of targets correctly identified increased at the rate of approximately 0.33 per target. There was no asymptotic trend, and in trials with 18 targets, the average number of targets that participants identified was almost 10. The mean number of targets identified in Experiment 2 well exceeded mean hits in Experiment 1 and the documented capacity of VSTM (that is, four items). Hence, target identification in Experiment 2 could not have been relying only on information stored in the VSTM. Instead, we argue that information yielded by optic flow was retained in the image structure. Being able to refer back to the visible layout made perception no longer subject to the capacity of memory-in-the-head because it was now offloaded from the head to the environment. Perception like so is an active, dynamical, and interactive process involving the observer, world objects, and the medium of structured light in between.

In Experiment 2, image structure did not differ between visual targets and distracters. The identical static image information could not inform an observer as to whether a given visible object was a target or a distracter. This required optic flow to specify the spatial relations of the visible objects and calibrate the image structure (or we can say optic flow assigns spatial meaning to the otherwise ambiguous image structure). This is the akin to broken crypsis. When prey blend into to a background by adjusting their image structure (such as color and texture), relative motion between the prey, the background, and the hunter/predator generates optic flow. While exposed to optic flow, the hunter is able to group the image structure on different depth planes and orientate them onto a depth map. Hence, he perceives the spatial relations of surfaces in the environment and organizes them into targets and distracters. The calibrated image structure continues to preserve the spatial relations after optic flow is gone. The relationship between optic flow and image structure is cyclic and mutually facilitative. They are not visual cues that may be weighted and summed to form a percept; they are components of a nonlinear and interactive system, which would not work without both.

It has been suggested that static binocular disparity is used to break camouflage [26], but disparity is only useful when observers view with both eyes and the distance between the targets and the observer is relatively small [3]. The distances at which binocular disparity would work to break camouflage depends on the anatomy of the eyes (e.g., interpupillary distance and amount of possible eye rotation, which differ among species) and typically in human observers disparity becomes weak beyond a few meters [27,28]. Quite obviously, disparity would not work if a hunter were aiming through the scope of a rifle with one eye open. In contrast, optic flow is monocularly detectable and robust over a much wider range of distances. Furthermore, once objects are segregated into targets and distracters, image structure enables an observer to perceive targets and maintain the perceived targets for a longer period of time than pure memory-in-the-head would allow. This has an important practical benefit because hunters are able to perceive targets both when they are moving and after they have stopped, so they can afford to wait and take action after the perceived targets have become stationary. The success of hunting is thereby enhanced.

A. Breaking Masquerade

Experiment 2 illustrated how combined optic flow and image structure information broke cripsis and yielded efficient and stable perception for action. This interactive system also accounts for how masquerade may be broken. In masquerade, a target is camouflaged by appearing like other uninteresting or undesired objects in the environment. In this case, the target contains distinguishing image structure from the background and/or distracters. It is not hard to detect the target, but it is hard to recognize it as a target.

We investigated the breaking of masquerade with an additional experiment using the current paradigm, where targets were on the front layer (closer to an observer) and distracters were on the back layer. Targets and distracters randomly possessed distinctive image structures, i.e., black or white borders, from trial to trial. When participants were looking at a stationary screen with black and white squares and asked to identify which squares were targets, they guessed and the performance was at chance. However, once motion occurred, the depth relations were immediately revealed, and thus, participants accurately identified all targets. Motion-generated optic flow information calibrated the distinctive image structures and informed the observer which colored squares were in the front (i.e., targets) and which colored squares were on the background. In this case, because image structure uniquely corresponded to the spatial relations specified by optic flow, perceived targets would be effectively preserved in image structures, yielding nearly 100% successful target identification, independent of the delay and blank conditions. Thus, perceiving masqueraded targets became trivial because uniquely coupled optic flow and image structure resulted in extremely efficient and stable perception and led to infallible target identification performance. The power of an interactive and nonlinear perceptual system was maximized when there was systematic one-to-one correspondence between optic flow and image structure information.

We may further alter the above experiment by adding a few white squares (that is, the same color as the targets) to the back layer to resemble a more complicated case of camouflage, where both cripsis and masquerade happen. This is like the nocturnal Amazonian fish masquerade into fallen leaves and cripsis into a river with a few real fallen leaves. In this case, optic flow would immediately reveal that all black squares are on the back (distracters) and most white squares are in the front (targets) plus a few white squares visibly (with optic flow) on the back layer (distracters). In such a trial, There exists approximate symmetry (that is, all black squares are distracters and most white are targets) that would allow an observer to ignore all the black items once (s)he sees that they are all distracters without exception and to concentrate on the white items. In this case, the task is essentially reverting to the conditions of Experiment 2, where targets and distracters of the same color. Therefore, the effect of combined optic flow and image structure information on breaking camouflage is nonlinear in that adding a different form of camouflage would not make target perception more difficult.

5. CONCLUSION

With these experiments, we demonstrated that combined optic flow and image structure information allowed observers to perceive large numbers of camouflaged visual targets over long time delays. With target-specifying optic flow and target-preserving image structure, participants were able to perceive more items stably than would be allowed by memory-in-the-head. These findings are consistent with those of Hayhoe and colleagues [29,30], who studied eye movement and found that when performing a sequence of actions, participants looked in real time to obtain information that was just-in-time for the next immediate action and generally avoided prestoring information in memory-in-the-head for future actions. For example, when making a peanut butter sandwich, participants did not first scan the whole visual scene, remember locations of the knife, the bread, and the peanut butter jar, and then make their sandwiches based on memory. Instead, they looked at the handle of the knife when they were about to pick it up and looked at the tip of the knife when they were about to dip it into the peanut butter jar. Therefore, when performing perceptually guided actions, individuals actively interact with the world and naturally seek stability and regularity in world structures to reduce load on the internal cognitive system. Use of image structure to protract the availability of perceived world properties (e.g., location and depth) is entirely consistent with the performance of a wide variety of everyday tasks.

The interaction between optic flow and image structure represents the fundamental structure and organization of animals' interactions with the environment in applications that have direct bearing on the continued existence over personal, historical, and evolutionary time. In particular, it is what allows a hunter, situated in a rich and complex physical environment, to break camouflage and perceive targets in a temporally stable manner until (s)he is ready to take action.

In this work, we outlined a perceptual mechanism that underlies accurate and stable perception of camouflaged targets and has great implications for machine vision and robotics. It has wide applications in industrial and/or military uses because an interactive system involving optic flow and image structure maximally takes advantage of the regularity of the environment and reduces demand on internalized memory.

Funding. National Natural Science Foundation of China (NSFC) (31571116); Sun Yat-sen University (SYSU) (16wkpy29).

REFERENCES AND NOTE

1. M. Stevens and S. Merilaita, "Animal camouflage: current issues and new perspectives," Philos. Trans. R. Soc. B **364**, 423–427 (2009).
2. Zoologists Stevens and Merilaita (2009) defined camouflage to include all forms of concealment with the goal of preventing target detection and recognition. On their list of camouflage methods, this goal could be achieved by crypsis (which prevents target detection), masquerade (which prevents target recognition), motion dazzle (which perturbs motion speed and trajectory estimation), and motion camouflage (which hinders motion detection). The first two types deal with camouflage on the static image level, whereas the other two types deal with disguising motion. In this work, we use the term "camouflage" to exclusively refer to image-based concealment, i.e., crypsis and masquerade, assuming effective and veridical motion detection.
3. D. Regan, "Orientation discrimination for objects defined by relative motion and objects defined by luminance contrast," Vis. Res. **29**, 1389–1400 (1989).
4. I. Sazima, L. N. Carvalho, F. P. Mendonça, and J. Zuanon, "Fallen leaves on the water-bed: diurnal camouflage of three night active fish species in an Amazonian streamlet," Neotrop. Ichthyol. **4**, 119–122 (2006).
5. A. T. Smith and R. J. Snowden, *Visual Detection of Motion* (Academic, 1994), pp. 1–2.
6. D. Regan, *Human Perception of Objects* (Sinauer Associates, 2000), p. 577.
7. J. J. Gibson, *The Ecological Approach to Visual Perception* (Houghton Mifflin, 1979/1986).
8. K. Nakayama and J. M. Loomis, "Optic velocity patterns: velocity-sensitive neurons and space perception," Perception **3**, 63–80 (1974).
9. M. L. Braunstein, "Structure from motion," in *Visual Detection of Motion* (Academic, 1994), pp. 367–396.
10. J. J. Koenderink and A. van Doorn, "Visual detection of spatial contrast; influence of location in the visual field, target extent and illuminance level," Biol. Cybern. **30**, 157–167 (1978).
11. J. T. Todd, "The perception of three-dimensional structure from rigid and non-rigid motion," Percept. Psychophys. **36**, 97–103 (1984).
12. J. T. Todd, "The perception of three-dimensional structure from motion," in *The Perception of Space and Motion* (Academic, 1995), pp. 202–221.
13. F. Domoni and C. Caudek, "3-D structure perceived from dynamic information: a new theory," Trends Cogn. Sci. **7**, 444–449 (2003).
14. W. H. Warren, "Optic flow," in *The Senses: A Comprehensive Reference* (Academic, 2008), Vol. **2**, pp. 219–230.
15. J. J. Koenderink, "The structure of images," Biol. Cybern. **50**, 363–370 (1984).
16. D. Marr and E. Hildreth, "Theory of edge detection," Proc. R. Soc. London B **207**, 187–217 (1980).
17. R. Rosen, *Fundamentals of Measurement and Representation of Natural Systems* (Elsevier, 1978).
18. Y. L. Lee and G. P. Bingham, "Large perspective changes (≥45°) yield perception of metric shape that allows accurate feedforward reaches-to-grasp and it persists after the optic flow has stopped," Exp. Brain Res. **204**, 559–573 (2010).
19. J. S. Pan, N. Bingham, and G. P. Bingham, "Embodied memory: effective and stable perception by combining optic flow and image structure," J. Exp. Psychol. **39**, 1638–1651 (2013).
20. J. S. Pan, N. Bingham, and G. P. Bingham, "Embodied memory allows accurate and stable perception of hidden objects despite orientation change," J. Exp. Psychol. **43**, 1343–1358 (2017).
21. S. J. Luck and E. K. Vogel, "The capacity of visual working memory for features and conjunctions," Nature **390**, 279–281 (1997).

22. E. K. Vogel, G. F. Woodman, and S. J. Luck, “Storage of features, conjunctions, and objects in visual working memory,” J. Exp. Psychol. **27**, 92–114 (2001).
23. M. E. Wheeler and A. M. Treisman, “Binding in short-term visual memory,” J. Exp. Psychol. **131**, 48–64 (2002).
24. A. D. Milner and M. A. Goodale, *The Visual Brain in Action*, 1st ed. (Oxford University, 1995).
25. A. D. Milner and M. A. Goodale, *The Visual Brain in Action*, 2nd ed. (Oxford University, 2006).
26. S. G. Wardle, J. Cass, K. R. Brooks, and D. Alais, “Breaking camouflage: binocular disparity reduces contrast masking in natural images,” J. Vis. **10**(14), 38 (2010).
27. J. E. Cutting and P. M. Vishton, “Perceiving layout and knowing distances: the integration, relative potency, and contextual use of different information about depth,” in *Handbook of Perception and Cognition: Perception of Space and Motion* (Academic, 1995), Vol. **5**, pp. 69–117.
28. R. L. Gregory, *Eye and Brain* (World University Library, 1966), p. 59.
29. M. M. Hayhoe, A. Shrivastava, R. Mruczek, and J. B. Pelz, “Visual memory and motor planning in a natural task,” J. Vis. **3**(1), 49–63 (2003).
30. M. F. Land and M. M. Hayhoe, “In what ways do eye movements contribute to everyday activities,” Vis. Res. **41**, 3559–3565 (2001).