Vision Research 158 (2019) 49–57

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Perceiving blurry scenes with translational optic flow, rotational optic flow or combined optic flow

Hongyuan Wu[a,1], Xiaoye Michael Wang[b,1], Jing Samantha Pan[a,*]

[a] Department of Psychology, Sun Yat-sen University, Guangzhou, China
[b] Department of Psychological and Brain Sciences, Indiana University, IN, USA

ARTICLE INFO

Keywords:
Translational optic flow
Rotational optic flow
Image structure
Scene perception
Low vision

ABSTRACT

Perceiving the spatial layout of objects is crucial in visual scene perception. Optic flow provides information about spatial layout. This information is not affected by image blur because motion detection uses low spatial frequencies in image structure. Therefore, perceiving scenes with blurry vision should be effective when optic flow is available. Furthermore, when blurry images and optic flow interact, optic flow specifies spatial relations and calibrates blurry images. Calibrated image structure then preserves spatial relations specified by optic flow after motion stops. Thus, perceiving blurry scenes should be stable when optic flow and blurry images are available. We investigated the types of optic flow that facilitate recognition of blurry scenes and evaluated the stability of performance. Participants identified scenes in blurry videos when viewing single frames and the entire videos that contained translational flow (Experiment 1), rotational flow (Experiment 2) or both (Experiment 3). When first viewing the blurry images, participants identified a few scenes. When viewing blurry video clips, their performance improved with translational flow, whether it was available alone or in combination with rotational flow. Participants were still able to perceive scenes from static blurry images one week later. Therefore, translational flow interacts with blurry image structures to yield effective and stable scene perception. These results imply that observers with blurry vision may be able to identify their surrounds when they locomote.

1. Introduction

Human observers are adept at perceiving and identifying visual scenes. Converging evidence suggests that perceiving real-world scenes requires separating and locating surfaces in space (Greene & Oliva, 2009; Kimchi, 1992; Navon, 1977; Oliva & Torralba, 2006). In other words, for scene perception, it is more important to perceive the 3D spatial layout than to identify individual objects. This is especially true when the stimuli are blurry (Peyrin, Chauvin, Chokron, & Marendaz, 2003; Schyns & Oliva, 1994).

3D spatial relations of surfaces and objects can be specified by motion-generated optic flow information. When an observer and the surrounding surfaces move relative to one another, the opaque surfaces project images to the observer, and motion continuously and lawfully transforms those images. The lawful transformation is called optic flow (Gibson, (1979/1986)), that, in part, is structured by the distances between the observer and the surfaces in the environment. Variation in distances also leads to motion-parallax in the flow, where the magnitude of flow is greater for surfaces that are closer (Nakayama & Loomis, 1974). As a result, progressive occlusion occurs when a near-surface passes in front of a far surface, the optical texture projected from the far surface is deleted along the contour projected from the relevant edge of the front surface. When part of the far surface comes back into view, the optical texture projected from it is accreted along the contour projected from the relevant edge of the surface to the front. In this way, optic flow provides immediate and powerful information about the 3D layout of surfaces in a cluttered terrain (Todd, 1995; Domini & Caudek, 2003).

A primary function of perception is to guide actions. When guiding locomotion, in particular, the scene must be accurately and stably perceived throughout the interaction. When scenes appear blurry, how might an observer perceive them in an effective and enduring fashion? We propose that in this case combined optic flow and image structure are required for two reasons. First, optic flow specifies the 3D spatial relations and motion detection is generally functional when there is a lack of image information, as in the case of blurry images (Barton &

* Corresponding author.
E-mail address: panj27@mail.sysu.edu.cn (J.S. Pan).
[1] The first and second author contributed equally to this work.

https://doi.org/10.1016/j.visres.2018.11.008
Received 4 July 2018; Received in revised form 28 November 2018; Accepted 29 November 2018
0042-6989/

Rizzo, 1994; Whitaker & Buckingham, 1987). Second, although optic flow is ephemeral and becomes unavailable when observer motion ceases, the image structure persists and, having been calibrated by the preceding optic flow, stably preserves the information that was provided by the flow. Observers with blurry vision have been shown to use motion-generated information to discriminate 3D shapes (Norman, Beers, Holmin, & Boswell, 2010) and to perceive depth (Jobling, Mansfield, Legge, & Menge, 1997).

Pan and colleagues investigated how interacting optic flow and image structure information led to effective and stable perception of visual events with blurry images in normal controls and in age-related macular degeneration (AMD) and amblyopic patients (Pan & Bingham, 2013; Pan et al., 2017). In these studies, participants attempted to identify events from blurry images (1) when single frames of blurry images from videos were presented; (2) when all blurry frames from a video were presented with motion masks; or (3) when all blurry frames were presented sequentially with detectable motion. Events were not reliably perceived in Conditions 1 or 2 with limited image structure, but were better perceived in Condition 3 with motion. Good performance persisted when participants viewed the same individual blurry images again, both immediately after the motion condition and after a delay of at least five days. The results from performance of AMD, amblyopic patients and normally sighted controls all replicated one another.

In these studies, optic flow was generated by moving objects visible to a stationary observer. This kind of optic flow is called local flow. Conversely, when an observer moves in an otherwise stationary environment, global optic flow is generated (Gibson, 1966). We now investigate what happens when an observer, who has access to only low spatial frequencies, moves while observing a stationary surround. Is she able to identify surrounding scenes using blurry image information and global optic flow? Answers to these questions have direct application to low vision rehabilitation, because, if this is effective, then an observer should be trained to take a proactive role in moving to create optical information for perception.

In a stationary environment, observer motions generate two types of global optic flow: locomotion of an observer includes translational motion that generates flow in a radial pattern while rolling of the head or eye generates rotational flow in a solenoidal pattern. Translational flow is effective in providing information regarding depth relations and spatial structures (Nakayama & Loomis, 1974; Warren, Morris, & Kalish, 1988). On the other hand, flow generated from rotation around an axis that passes through the point of observation does not specify relative depth and thus does not provide meaningful information regarding 3D spatial layouts (e.g. Koenderink, 1986; Warren & Hannon, 1990; Koenderink & Van Doorn, 1991; Lind, 1996). Because perceiving spatial layout of surfaces and objects is essential for scene identification, scene perception should be successful with translational flow but not with rotational flow. However, due to eye movement and/or postural change, translational and rotational flow typically occur together. The question, therefore, is whether the combination of translational and rotational global flow is as effective as pure translational flow in specifying scenes, or whether the presence of rotational flow perturbs the use of the entire optic flow field and prevents the observer from recovering 3D spatial layout in the environment for scene identification.

Similar investigations on the interactions between optic flow generated through translational and rotational motion have been performed in the context of the perception of heading. Heading direction is specified by the focus of expansion (FOE), which is readily detectable in the global flow pattern generated by pure translation. However, when rotational flow is added, the FOE is shifted and is no longer aligned with heading direction. This is the "rotation problem" in perception of self-motion (Warren, 1998). When translation and rotation are combined, the visual system separates translational flow from rotational flow and regains the FOE by using extra-retinal information (i.e. oculomotor information), retinal flow information (i.e. instantaneous flow velocities at depth edges) or both (Banks, Ehrlich, Backus, & Crowell, 1996; Crowell, Banks, Shenoy, & Andersen, 1998; Royden, Banks, & Crowell, 1992; Royden, Crowell, & Banks, 1994). Generally speaking, it is more challenging to perceive self-motion with retinal flow information alone, which indicates that the addition of rotational flow may perturb the entire flow field, which otherwise supports the detection of depth layout (Warren & Hannon, 1990; Crowell, Maxwell, Shenoy, & Andersen, 1998).

The task of perceiving blurry scenes is related to that of perceiving heading direction, because theoretically both are about extracting optical information from the global flow field that specifies 3D properties of the environment and observer movement relative to that environment. In the case of heading perception, the task is to locate the FOE. In the case of scene perception, the task is to perceive depth relations and spatial layout of environmental surfaces and objects. Given that heading perception encounters the rotation problem, similarly, perceiving blurry scenes may also become problematic when rotational flow is added.

The aforementioned heading studies were primarily concerned with recovering the FOE in flow fields composed of both translational and rotational flows for guiding self-motion. Scene recognition, on the other hand, focuses on the use of the overall flow patterns to recover 3D spatial layout. Even though the FOE is only one aspect of optic flow, to recover it still requires the entire flow field and relies on the presence of effective depth structures, as in the retinal flow solution to the rotation problem. Given that optic flow from rotation around an axis that passes through the point of observation does not generate relative motion in depth, adding it to pure translational flow could be disruptive for recovering 3D spatial layout. Therefore, it is possible that the combination of rotational and translational flow could affect scene identification with blurry images.

In the current study, we simulated observer motions of translation, rotation or both by moving a camera and filming the otherwise stationary environments. The filmed scenes were processed to make them black-and-white and blurry. Participants identified the scenes in three experiments. In Experiments 1 and 2, participants were exposed to translational flow and rotational flow respectively. These experiments addressed the questions of whether information generated by observer motion helped with perceiving blurry scenes. In Experiment 3, we explored the effect of combining translational and rotational flow on scene recognition. We compared performance across the experiments to determine how combined translational and rotational flow affected performance. Identification accuracy and performance stability were used to evaluate performance, i.e. whether one identified blurry scenes with motion and continued to do so after motion stopped.

2. Experiment 1: Identifying blurry scenes with translational flow

In Experiment 1, we moved a camera along straight paths in daily environments to film familiar scenes. We then applied Gaussian blur to the recorded materials and investigated whether translational flow yields effective scene perception despite image blur, and whether performance remains temporally stable.

2.1. Methods

Participants: Ten undergraduate students from Sun Yat-sen University completed this experiment. All participants reported having normal or corrected to normal vision. They were informed of the possible consequences of the study and signed informed consent before the experiment. Participants were compensated at the rate of ¥30/hour. The research was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Materials: The experimenter held a Nikon D7000 camera (18 mm lens, field of view = 66° × 47°) stably and shot ten videos of ten familiar scenes. These scenes included a wharf, a lush garden, a supermarket, a dormitory room, a fitness room with exercise bikes, a

Fig. 1. The frames of the original and edited videos of the ten scenes used in this study.

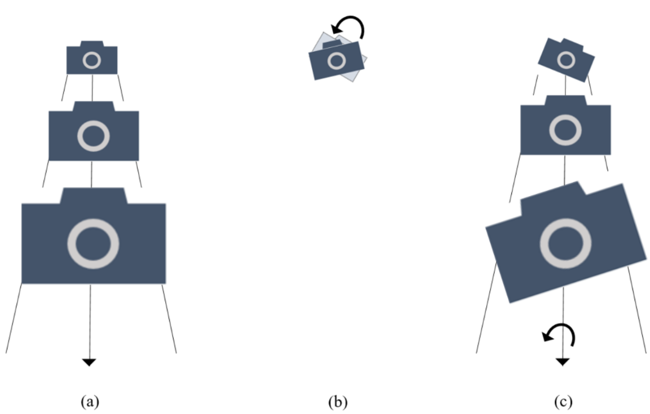

Fig. 2. An illustration of camera movement while recording videos that contained (a) translational flow in Experiment 1, (b) rotational flow in Experiment 2 and (c) combined translational and rotational flow in Experiment 3.

museum exhibition hall, a clothing store, a bicycle parking lot, a road with parked vehicles, and a boxing gym (Fig. 1). When recording, the camera translated forward in the scene with the lens aligned with the direction of motion (Fig. 2). In the selected scenes at the moment of recording, there was no noticeable motion of objects. This was to control for local optic flow associated with object motion.

Recorded videos (image quality 1920 × 1080 pixels) were processed using Adobe Premiere Elements 14. The videos were converted to black and white. We then applied Gaussian blur filters with equal radii in the horizontal and vertical directions. The Gaussian blur radii were between 10 and 15 pixels for the ten videos, and the cutoff frequencies were 0.02–0.03 cycles/pixel (with the actual display size of 43.6° × 25.4°, the cutoff frequencies converted to 0.88–1.32 cycles per degree visual angle). The Gaussian blur radii varied slightly to achieve a similar subjective experience of image blur[2].

The testing stimuli consisted of 20 still frames that were extracted from each processed video and were presented in a java applet. On each trial, the top part of the display showed the visual stimulus and the bottom part was a textbox, where responses were entered. There was also a "Next" button in the bottom right corner of the display. Clicking on the button changed the display to the next trial. The stimuli were presented on a 25-in. Dell monitor with a refresh rate of 60 Hz. The display size of the stimuli was 40 cm × 22.5 cm on screen (16:9), and the viewing distance was 50 cm (thus, the testing stimuli spanned over 43.6° × 25.4°).

Procedures: This experiment was conducted in a fully lit room (screen brightness = 51.7 cd/m^2). Participants completed the experiment in two sessions. During the first session, they signed the consent forms and completed four testing conditions. Participants were told that they would see some blurry static images or blurry animations of familiar daily scenes or environments, and they needed to describe what this place was or what scene was represented in the stimuli.

In Condition 1, three static frames from each blurry scene were randomly selected and presented, one at a time. Participants described the scene while the frame was presented on the screen. The responses were not timed. A naïve experimenter typed out the participant's description in the textbox at the bottom of the screen. The experimenter encouraged participants to give a description for each blurry frame, although participants often reported that they did not see anything

[2] After adding blur to each recorded scene, we extracted static frames from each video and ordered the frames from each video from least to most blurry. We presented these images in order to naïve observers, beginning with the least blurry. If three consecutive observers correctly identified the scenes (that is, correctly named or described the scenes), then we presented blurrier images to the next observer. When ≤ 1 in 5 consecutive observers was able to identify the scene, we selected this level of blur as the testing stimuli. We did this for all 10 scenes. These observers did not participate in the actual experiment.

during the first attempts. After participants finished describing a frame, the experimenter pressed the "next" button to move on.

Next, in Condition 2, the 20 blurry frames from each video were played in the order they appeared in the original videos. A blank screen with no image structure, serving as a motion mask, was inserted after each frame. The duration of each frame was 100 ms and the duration of each mask was 500 ms. On each trial, the frames and masks were played in a loop and the experimenter made sure that participants viewed the entire sequence at least once. Stimuli in this condition provided participants with full image information but no motion information.

In Condition 3, the motion masks were removed and the 20 frames from each video were played and looped in order. The frame duration was 100 ms and the final stimuli looked like blurry animations that contained both low spatial frequency image information and motion information.

After viewing scenes in the context of motion, in Condition 4, participants identified scenes from blurry static images again. Similar to that in Condition 1, the stimuli were three randomly selected images from each of the ten scenes.

One week later, participants returned for the second testing session. They first wrote down scenes that they recalled having seen in the previous testing session. Then, they completed Condition 5, which was similar to Conditions 1 and 4, where participants identified scenes from 30 blurry static frames (3 frames randomly selected from each of the ten scenes).

(To watch a video demo of this study, go to https://youtu.be/CtMbrM38U2A. To download the full experiment and test yourself, go to https://whypsy.github.io/material/.)

Data processing: Each of the ten participants completed 30 trials (10 scenes and 3 frames from each scene) in Conditions 1, 4 and 5; and 10 trials (for the 10 scenes) in Conditions 2 and 3, yielding 110 trials per participant and 1100 trials total.

Two raters, who were naïve to the purpose and procedures of the study, were recruited to code participants' responses. Before coding, they viewed both the original untreated videos and the actual experimental stimuli. Raters were instructed only that a correct answer should describe the essence of the scenes or the key objects in the scenes. (For example, one of the scenes was a fitness center with rows of exercise bikes. A correct answer could either be a name of the place, such as "gym", or contain important details, such as "a room with many bikes".) The two raters coded the responses independently. Inter-rater reliability was high, 91.7%, with the raters in agreement on 1009 out of 1100 trials (assessed using the "joint probability of agreement" method, Uebersax, 1987). For the 91 trials that the raters coded differently, we randomly picked one rater's coding as the final result.

2.2. Results and discussion

In Experiment 1, we tested if translational flow provided information to allow for scene identification from blurry images or animations, despite impoverished image structure information. We performed a repeated-measures ANOVA comparing the proportion of trials correctly identified in different conditions for different scenes. There was a significant main effect of Condition ($F(4, 36) = 28.80$, $p < 0.001$, $\eta_p^2 = 0.76$). As shown in Fig. 3, the correct rate of scene identification in pre-motion conditions (i.e. Conditions 1 and 2) was lower than in the motion (i.e. Condition 3) and post-motion conditions (i.e. Conditions 4 and 5). There was also a significant main effect of Scene ($F(9, 81) = 2.55$, $p = 0.012$, $\eta_p^2 = 0.22$), which suggested that some scenes were harder to identify than others. The interaction between Scene and Condition was not significant ($F(6.4, 57.6) = 0.965$, $p = 0.46$), suggesting that scenes did not affect performance differentially across the five conditions. Hence, in the subsequent analysis, we focused on studying how performance fluctuated among the conditions.

In Condition 1, scenes were correctly perceived in 86 out of 300 trials, or in 28.7% of the trials (95% CI = [23.7%, 33.6%]). In Condition 2, 32 out of 100 trials were identified, or in 32.0% of the trials (95% CI = [20.9%, 43.1%]). There was no improvement in performance from Condition 1 to Condition 2 ($t(9) = 0.67$, $p = 0.51$). Thus, regardless of the number of frames (or equivalently the amount of blurry static information), with only image-based information, the accuracy of scene identification was low. In Condition 3, when both static image information and motion information were available, scenes were identified in 63 out of 100 trials, that is, in 63% of the trials (95% CI = [54.0%, 72.0%]). Compared to Condition 2, with added motion information in Condition 3, performance improved significantly ($t(9) = 6.46$, $p < 0.001$).

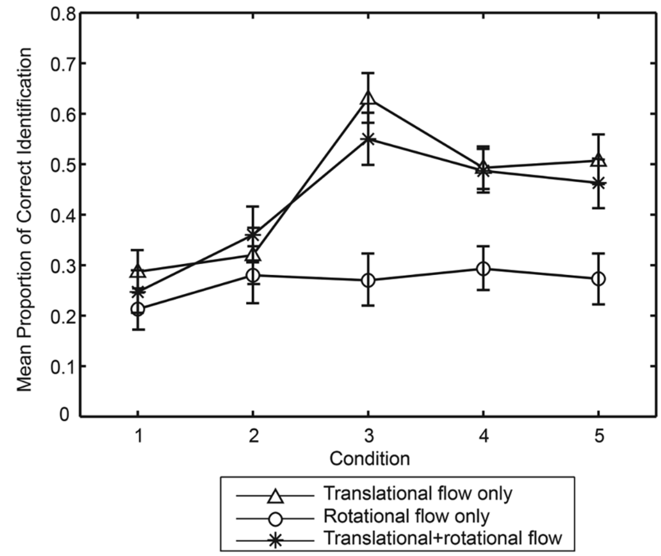

Fig. 3. Translational flow, whether it was available alone (△) or in combination with rotational flow (○), improved perception of blurry scenes. However, pure rotational flow (*) did not aid scene perception. Error bars = 1 SE.

After viewing the scenes with motion, good performance persisted when participants were presented with static blurry images again in Condition 4. In 300 trials, participants identified 148 trials correctly, with a correct rate of 49.3% (95% CI = [45.1%, 53.6%]). Although performance in Condition 4 was worse than that in Condition 3 ($t(9) = 5.07$, $p = 0.001$), possibly as an effect of removing motion information, it was significantly better than performance in Condition 1 ($t(9) = 6.47$, $p < 0.001$). Given that the only difference between Conditions 1 and 4 was whether the task was performed before or after the scenes were presented with motion, superior performance in Condition 4 suggested that the effect of motion was preserved temporally in the blurred image structure.

In Condition 5 (second testing session one week later), participants recalled what they had seen in the first testing session. Altogether, ten participants successfully recalled a total of 39 scenes out of 100 trials. Next, when looking at the static blurry frames again, participants identified scenes in 152 out of 300 trials, or 50.7% (95% CI = [46.1%, 55.3%]). Performance in this condition was significantly better than that in Condition 1($t(9) = 7.59$, $p < 0.001$). More importantly, performance did not drop from Condition 4 ($t(9) = 0.76$, $p = 0.44$). Thus, good performance persisted for a week. Furthermore, for scenes that were correctly identified in Condition 3, the rate of identification in Condition 5 was 78.83% (149 out of 189 trials); for scenes that were identified in Condition 3 but not recalled, the identification rate in Condition 5 was 68% (54 out of 75 trials). These results suggested that performance in Condition 5 was dependent on performance in Condition 3 (i.e., performance with motion information present), but not on free recall.

In sum, these results suggested that translational optic flow generated by locomotion enabled scene perception when static image structures contained only low spatial frequencies that were uninformative about the scene. Furthermore, scene identification persisted over long delays and was independent of free recall performance suggested that optic-flow-calibrated spatial relations were preserved in the image structure, not solely in the mind. The combination of translational flow and image structure improved scene perception performance in a temporally stable fashion despite poor static image quality.

3. Experiment 2: Identifying blurry scenes with rotational flow

In this experiment, we explored how rotational flow alone, typically occurring with eye rotation or head tilt, affects blurry scene perception. Rotational flow was created by rolling the camera while filming. During this process, the camera stayed at a fixed location, i.e. the distance between the camera and the surfaces in the scene did not change.

3.1. Methods

Participants: Ten undergraduate students from Sun Yat-sen University, who did not participate in the previous experiment, completed this experiment. (Another participant's data were excluded because of no show during retest.) All participants reported having normal or corrected to normal vision. Before the experiment, participants signed informed consent. Participants were compensated at the rate of ¥30/hour.

Materials: We used the same Nikon D7000 camera to film the same scenes as in Experiment 1 (see Fig. 2). However, the camera motion was different. Instead of translating, the camera stayed at a fixed location and was attached to a tripod and rolled from −50 degrees to 50 degrees (the tripod had a built-in protractor that allowed precise measurement of rolling angles). The rolling was slow and smooth with a speed of approximately 10 degrees per second. The resultant videos resemble the visual experience of tilting one's head when viewing.

We used the same methods as in Experiment 1 to edit the videos, extract the static frames and display them on the same Dell monitor. Stimuli display size was $43.6^{\circ} \times 25.4^{\circ}$.

Procedures: Experiment 2 followed the same procedures as Experiment 1 for all 5 conditions, with the one exception that Condition 3 included rotational motion instead of translational motion.

Data processing: The data were coded following the same coding protocols and by the same raters as in Experiment 1. Inter-rater reliability rate was 92.55%, with consistent ratings for 1018 out of 1100 trials (assessed using the joint probability of agreement method, Uebersax, 1987). For trials that were rated differently, we randomly selected one rater's coding as the final score.

3.2. Results and discussion

In Experiment 2, we investigated whether pure rotational flow aided blurry scene perception. A repeated measures ANOVA was conducted to compare the proportions of correct scene identification across different conditions and different scenes. Same as in Experiment 1, there was a significant main effect of Scene ($F(9, 81) = 5.95$, $p < 0.001$, $\eta_p^2 = 0.40$). This suggested that some scenes used in this experiment were harder to identify than others. However, neither the main effect of Condition ($F(4, 36) = 2.28$, $p = 0.08$), nor the interaction between Condition and Scene ($F(36, 324) = 1.06$, $p = 0.38$) was significant. Therefore, scenes did not affect performance in the five conditions differentially. Moreover, unlike in Experiment 1 where performance was better during and after the motion condition, performance in this experiment remained constant across the conditions. Participants performed rather poorly in all 5 conditions, with the rate of correct scene identification ranging between 14.7% and 28%. Thus, rotation in the frontoparallel plane did not facilitate the perception of blurry scenes.

We designed Experiment 2 to contain only rotational flow information. Nonetheless, there might have been residual translational flow patterns in these stimuli, such as movements of leaves in the wind. In terms of optical information, these flow patterns should have diminished the differences in optical information between Experiment1 and Experiment 2, if anything. That there were virtually no influences of the rotational flow patterns on scene recognition, however, suggesting that these influences were probably negligible.

With the first two experiments, we demonstrated in Experiment 1 that translational optic flow provided information about spatial layout that aided scene recognition even with low spatial frequency images. Results in Experiment 1 supported this claim, where blurry scenes were better perceived in Condition 3 (with motion) than in Condition 2 (without motion). However, in Experiment 2 we showed that rotational flow did not aid scene recognition (as performance in Condition 3 was not better than that in Condition 2), because rotational flow only occurs in the frontoparallel plane. Thus, it does not provide any useful information regarding the depth structures of a scene. The results from Experiments 1 and 2 show that the improvement in scene recognition in Experiment 1 was due to the spatial structural information provided by translational flow, not simply the addition of motion per se or repeated viewing.

4. Experiment 3: Identifying blurry scenes with translational and rotational flow

Heretofore, we have shown that translational optic flow enabled effective and stable perception of scenes with blurry image structure, but rotational optic flow did not. Nonetheless, in everyday life, observers are typically exposed to both flows simultaneously because eye/head movement accompanies locomotion. Are blurry scenes still perceptible when translational and rotational flow are combined with image structure? Specifically, is the spatial information from translational flow still detectible and veridical to specify 3D spatial relations, or does the presence of rotational flow interfere with translational flow, preventing effective recognition of visual scenes (analogous to the case of perception of self-motion)? In Experiment 3, we explore these questions by using motion that contains both translational and rotational flow.

4.1. Methods

Participants: Ten undergraduate students from Sun Yat-sen University, who did not participate in the previous experiments, completed this experiment. All participants reported having normal or corrected to normal vision. Before the experiment, they were informed of the possible consequences of the study and signed informed consent. Participants were compensated at the rate of ¥30/hour.

Materials: Ten videos were recorded using the same camera in the same ten environments. The camera was held and translated in the same heading direction as in Experiment 1, and during this motion, the camera was rolled from −50 degrees to 50 degrees, same as in Experiment 2. The camera moved in the transverse plane and in the frontoparallel plane simultaneously so that the videos contained both translational flow and rotational flow. See Fig. 2.

We used the same procedures as those in Experiments 1 and 2 to edit the videos, selected 20 frames from each scene and presented the stimuli. The difference in roll angle between each pair of consecutive frames was between 3 and 7 degrees. The display size of the stimuli was $43.6^{\circ} \times 25.4^{\circ}$.

Procedures: The task and procedures in this experiment were the same as those in Experiments 1 and 2, with the exception that Condition

3 contained both translational and rotational motion.

Data processing: The data was coded by the same two raters following the same coding protocol as in Experiments 1 and 2. Inter-rater reliability was 93.0%, with rater agreement on 1023 out of 1100 trials (assessed using the "joint probability of agreement" method, Uebersax, 1987). For trials that were rated differently, we randomly selected one rater's coding as the final score.

4.2. Results and discussion

In Experiment 3, we investigated if the combination of rotational flow and translational flow calibrated image structure information to allow scene identification despite limited image structure information.

First, a repeated measures ANOVA with Condition and Scene as the within-subject factors, and rate of scene identification as the dependent measure showed a significant main effect of Condition ($F(4, 36) = 13.09$, $p < 0.001$, $\eta_p^2 = 0.59$). As depicted in Fig. 3, the pattern of performance change across conditions was similar to Experiment 1. Specifically, the rate of scene identification was low in the pre-motion conditions (Conditions 1 and 2), but increased in Condition 3, with the addition of both rotational and translational flows. More importantly, performance in the post-motion conditions (Conditions 4 and 5) was better than in the pre-motion conditions. In addition, the main effect of Scene was significant ($F(9, 81) = 5.63$, $p < 0.001$, $\eta_p^2 = 0.39$), suggesting that some scenes were harder to identify than others. However, the interaction of Scene and Condition was not significant ($F(6.43, 57.88) = 1.65$, $p > 0.05$), so scenes did not differentially affect performance in the five conditions.

In Condition 1, 74 out of 300 trials were correctly identified (24.7%, 95% CI = [18.5%, 30.8%]). Scene identification improved significantly when all 20 frames were presented in Condition 2 ($t(9) = 3.77$, $p = 0.004$), where participants successfully identified 36 out of 100 trials, (36.0%, 95% CI = [21.6%, 50.4%]). In Condition 3, when translational flow, rotational flow, and image structure information were all available, participants performed significantly better than in Condition 2 ($t(9) = 5.43$, $p < 0.01$), identifying 53 out of 100 trials (or 53%, 95% CI = [41.4%, 68.6%]). After the images had been calibrated by motion, participants were still able to identify the scenes from static blurry images in Condition 4. In 300 trials, they correctly identified 48.7% (95% CI = [42.6%, 54.7%]). This was significantly better than in Condition 1($t(9) = 9.6$, $p < 0.001$).

During retest, participants first performed a recall test and the 10 participants, in total, recalled 42 out of 100 trials. Then, they viewed the blurry static images, as in Conditions 1 and 4. The percentage of scene identification in Condition 5 was 46.3% (139 out of 300 trials, 95% CI = [39.7%, 52.9%]), which was not different from performance in Condition 4 ($t(9) = 0.70$, $p = 0.48$) and was better than performance in Condition 1($t(9) = 5.43$, $p < 0.001$). Like in Experiment 1, this improvement was related to the performance in Condition 3: for the scenes that participants identified in Condition 3, the percent correct in Condition 5 was 77.0% (127 out of 165 trials); for scenes that they failed to identify in Condition 3, the percentage of correct identification was 8.9% (12 out of 135 trials) in Condition 5. Comparing the identification performance in Condition 5 to the free recall performance, we found that for scenes which were identified in Condition 3 but not recalled, the identification rate in Condition 5 was 69.84% (54 out of 63 trials), which was significantly higher than performance in Condition 1 ($t(20) = 2.67$, $p < 0.05$). These results suggest that performance in Condition 5 was tied to performance in the motion condition, not to recall, in other words, it was a function of perception of spatial structure not the memory of it.

These results suggested that the combination of translational flow and rotational flow information enabled scene perception when low spatial frequencies were present in the stimuli. Additionally, the high rate of scene identification persisted for a week and this performance superseded free recall performance in terms of stability. Together, translational flow, rotational flow, and image structure yielded effective and stable scene perception despite poor static image quality.

We subsequently tested the relationship among the three experiments with different types of flow, using a 3 (types of flow) × 5 (conditions) × 10 (scenes) mixed-design ANOVA. There were significant main effects of Type of flow ($F(2, 27) = 269.4$, $p < 0.001$, $\eta_p^2 = 0.91$), Condition ($F(4, 108) = 35.34$, $p < 0.001$, $\eta_p^2 = 0.567$), and Scene ($F(9, 243) = 5.21$, $p < 0.001$, $\eta_p^2 = 0.162$). Because the interaction between Scene and Condition was not significant ($F(36, 972) = 1.39$, $p = 0.07$), the main effect of Scene was interpreted as showing that some scenes were harder to identify than others. The main effect of condition was significant and performance varied across the conditions in two out of the three experiments (Experiments 1 and 3). The significant main effect of Type of flow suggested that the type of available optical information affected scene identification. Moreover, the only significant 2-way interaction was between Condition and Type of flow ($F(8, 108) = 6.44$, $p < 0.001$, $\eta_p^2 = 0.32$). There was no significant 3-way interaction ($F(72, 972) = 1.21$, $p = 0.124$). The 2-way interaction between Condition and Type of flow was important, because it revealed exactly how the optical information affected scene identification, which we now discuss.

In all three experiments, performance was equally poor in the premotion conditions. There was no difference in performance in Condition 1 ($F(2, 87) = 1.93$, $p = 0.15$) or in Condition 2 ($F(2, 87) = 0.54p = 0.59$) among the various types of optic flow information. From Condition 2 to Condition 3, image structure information remained the same, but optic flow information was added. In Experiments 1, 2 and 3, the optic flow information was, respectively, translational flow, rotational flow and combined translational and rotational flow. We performed an ANOVA on data collected in Conditions 2 and 3 of the three experiments and found a significant interaction between Condition and Type of optic flow ($F(2, 27) = 16.53$, $p < 0.001$, $\eta_p^2 = 0.55$). Post-hoc LSD tests showed that there was no difference between performance in Condition 3 of Experiments 1 and Condition 3 of Experiment 3 ($p = 0.26$), both were better than performance in Condition 3 of Experiment 2 ($p < 0.001$ in both cases). In other words, rotational flow did not improve the performance (Experiment 2), but translational flow did improve performance, regardless of whether translational flow was available alone (Experiment 1) or in combination with rotational flow (Experiment 3).

After participants identified scenes with motion, they again identified scenes from the blurry static images without optic flow in Conditions 4 and 5. Performance in Conditions 4 and 5 was better than that in Condition 1 for Experiments 1 and 3. However, in Experiment 2, scene identification was not better when presented post-motion in Conditions 4 or 5. This means that improved performance in post-motion conditions of Experiments 1 and 3 was not because of repeated exposure to the visual stimuli, but was an effect of calibrating the blurry images with using motion generated information. Therefore, when high spatial frequency signals were absent, static image structure information alone was unable to specify scenes. However, once viewed together with translational optic flow, blurry images of scenes became perceptible and they remained so over long time durations.

5. General discussion

In the current study, we explored how patterns of global optical flow affected the effectiveness and stability of perception of surrounding scenes when image-based information contained only low spatial frequencies. With three experiments, we showed that when paired with blurry images, translational flow yielded effective and stable scene perception, whether it was available alone or in combination with rotational flow. However, rotational flow did not aid performance. Moreover, translational flow evoked stable scene perception over long time delays, demonstrating that perception of spatial layout was temporally preserved in the image structure, not in memory (reflected in the relationship between performance in Condition 3 with motion information and performance in Condition 5 after delay).

Translational flow enabled blurry scene perception because surfaces at different depth project differential flow velocities. These result in motion parallax and progressive occlusion among surfaces that are separated in depth. Motion parallax and progressive occlusion inform observers about the depth layout of surfaces in scenes (Hildreth & Royden, 2011) and hence enable scene perception (Schyns & Oliva, 1994; Greene & Oliva, 2009). Additionally, rolling on the frontoparallel plane did not produce relative motion in depth and, thus, did not provide useful optic flow information that specified depth layout, or facilitated scene recognition.

Furthermore, scene perception was equally effective with translational flow alone and with the combined translational and rotational flow. There are two possible reasons for this. First, rotation only changes retinal images, but it does not alter the projected spatial relations among surfaces in the environment or between world surfaces and the observer. In other words, motion parallax and progressive occlusion are not affected by rolling in the frontoparallel plane (Li & Warren, 2000) and information in translational flow is robust enough to support effective scene perception. Second, similar to the perception of self-motion, rotational flow does incur extra difficulty to perceive depth order in scenes, but the visual system can resolve the rotation problem by picking up information from the blurry but distinguishable image structure information. This shows that perceiving self-motion with rotation was more accurate in complex and realistic visual environments than in displays containing only random dots (Cutting et al., 1997; Li & Warren, 2000). If this were true, then the role of image structure information in perception with blurry vision might be more prominent than previously thought. Previously, image structure information was considered as serving the function of external memory storage for keeping the spatial structures specified by optic flow (Pan et al., 2017). Given the current results and interpretations, image structure might directly facilitate perceiving scenes for a moving observer by providing landmarks for tracking and resolving the rotation problem.

As a preliminary attempt to test the above-proposed solutions, we extracted motion information from stimuli in the three experiments. We used the MatLab open implementation of Sun and colleague's *Classic + NL* optic flow estimation method (Sun, Roth, & Black, 2010, 2014) to demonstrate the effects of different types of flow on capturing depth variations of a scene. We extracted two consecutive frames from the fitness room scene (the room with many bikes) in each experiment to generate optic flow fields. Fig. 4 shows the resulting flow field in each experiment, plotted in a black and white magnitude scale, where the lighter the area, the stronger the flow. In Experiment 1, with only translational flow, the bikes' 3D depth structures were rather pronounced with only two-frame motion. (This is like flipping two pages back and forth very quickly. The actual testing stimuli with 20-frame motion would reveal 3D depth structures more rigorously.) Similarly, in Experiment 3, the flow field associated with 2-frame motion also captured the scene's depth structure, but it is noisier than that from Experiment 1, perhaps due to the presence of rotational flow, as that is the only difference between Experiments 1 and 3. Nonetheless, the bikes, which are now tilted clockwise, were still identifiable. However, the flow field corresponding to Experiment 2 provided little, if any, depth structure in the scene. These plots were based on pure optic flow information (no image-based analysis) and they seem to support the first explanation that rotational flow perturbs image structure but not the projected spatial relations (versus the alternative that rotational flow

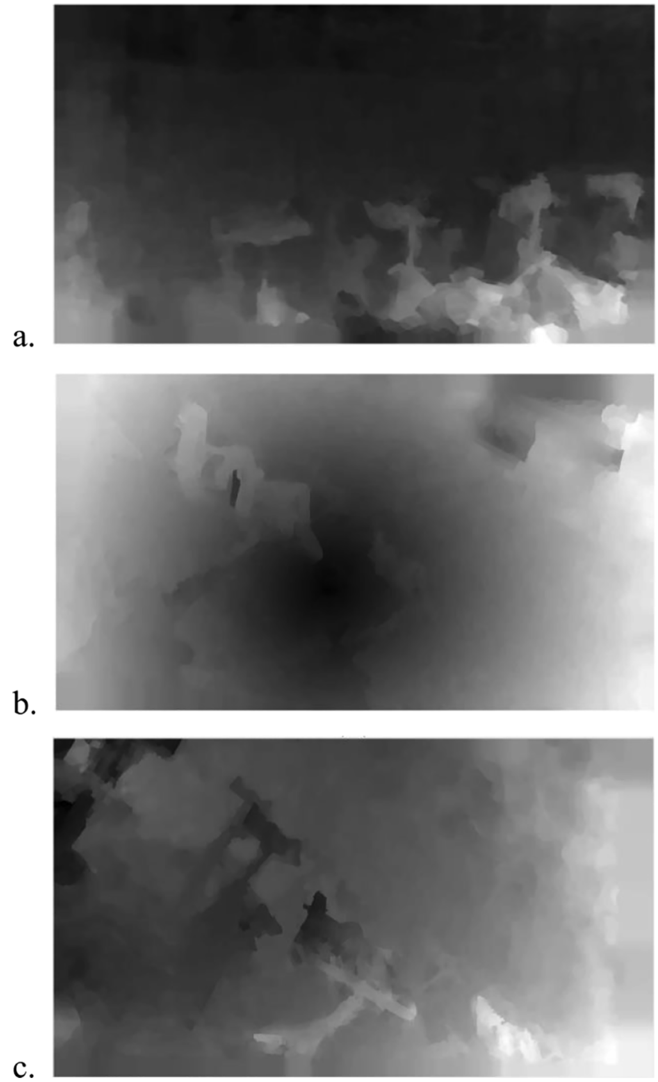

Fig. 4. Sample optic flow fields extracted from two consecutive frames of the bike scene in Experiments 1 (a), 2 (b), and 3 (c).

does perturb spatial layout information, but the rotation problem can be resolved by using image structures). Follow-up studies are being conducted to systematically and quantitatively analyze the flow patterns.

Blurry scenes were reliably perceived over time by virtue of the interaction between image structure and optic flow. When blurry images were presented as video clips, optic flow and image structure information were both available. Because motion measurement relies on low spatial frequencies, the detection of optic flow was unaffected by image blur. Although translational flow was strong in specifying spatial relations among world surfaces, it was transient. The information ceased to exist once motion stopped. Nonetheless, while optic flow was ongoing, it carried one structured image into the next structured image. In the context of motion, blurry images become spatiotemporally meaningful. In other words, translational flow specified spatial layout and, at the same time, calibrated the otherwise ambiguous image structure to specify scenes. After motion stopped, only image structure information was available. Unlike in the pre-motion condition, the calibrated blurry images now contained information from the optic flow, and continued to allow the perception of scenes so long as they remained in view. This interactive process utilized both dynamic optical information and static image information, each of which compensated the other's weakness. In this interaction, image structure information did not specify environmental surfaces and scenes directly, which would require high spatial frequencies in the image information. Instead, it held information across long time periods, making it pivotal for perceiving scenes in an effective and stable fashion.

Results of the current study have implications for low vision rehabilitation. Identifying one's surround is normally an easy task when images are clear (Boucart, Moroni, Thibaut, Szaffarczyk, & Greene, 2013). However, patients with many forms of eye diseases experience reduced visual acuity (VA) and/or contrast sensitivity (CS) (Jobling et al., 1997; Leat, Legge, & Bullimore, 1999). For these patients, it is traditionally believed that scene identification is challenging. However, VA and CS are not the sole determining factors of visual functioning because the human visual system naturally uses both static image-based information and motion-generated optic flow information. For instance, observers with blurry vision successfully identify objects in motion (Pan et al., 2017), and locomoting observers with blurry vision accurately perceive ground slant and step height (Bochsler, Kallie, Legge, & Gage, 2011; Bochsler, Legge, Kallie, & Gage, 2012). Results of the current study add to this literature, suggesting that active translation of a low vision observer may generate optical information that enables scene perception.

In a dynamic environment with nested observer and object motions, there is ample information that may lead to effective perception for observers who have blurry image vision and functional motion detection. Therefore, visual functioning should not be assessed using static test stimuli alone (e.g. visual acuity charts) and low vision rehabilitation, which aims to improve visual functioning, should include training that involves actively generating and detecting optical information during motion. For instance, low vision observers should be encouraged to locomote (walk, move in wheelchairs, or the like) to perceive. Additionally, given that rotational flow does not interfere with translational flow in its specification of spatial layout, a low vision observer need not execute stringent postural control while locomoting. For low vision observers who are reluctant to locomote, it is worth considering the use of locomoting devices like drones to help survey the environment and supply translational optic flow information. One of the authors used a drone (DJI Mavic Pro) to record a few large-scale blurry scenes (Wu, 2018). The blurry movies of scenes were also recognizable. In fact, flying generates very strong translational flow, because optic flow strength is proportional to locomoting speed. Of course, presently, drones are not designed to facilitate visual functioning and they can be complicated to operate. Nonetheless, it is a potential direction that manufacturers of electric vision aids may explore.

6. Conclusion

In three experiments, we showed that optic flow from translational motion, but not rotational motion, led to effective and stable visual scene perception, when there was blurry image-based information. In this case, translational flow may exist in isolation or together with rotational flow. The improvement was due to the spatial information contained in the optic flow generated by translation and not due to repeated viewing or motion per se. Continued stability of scene recognition is due to calibration of motion information (i.e., spatial layout information) onto the image structure. These results shed light on vision care and rehabilitation for patients with low vision, suggesting motion-generated optical information should be incorporated. For instance, coaching patients to actively locomote in the environment to acquire dynamic optical information is of key importance for regaining visual functions.

Acknowledgment

This study is supported by the National Natural Science Foundation of China (General Program, Grant 31571116) and by Sun Yat-sen University (Grant 16wkpy29) awarded to J. S. Pan. The funders have no role in study design, data collection, or preparation of this manuscript.

We thank Dr. Geoffrey P. Bingham and Ms. Rachel Herth for their contribution in preparing and editing the manuscript.

References

Banks, M. S., Ehrlich, S. M., Backus, B. T., & Crowell, J. A. (1996). Estimating heading during real and simulated eye movements. *Vision Research, 36*(3), 431.

Barton, J. J. S., & Rizzo, M. (1994). Motion perception in optic neuropathy. *Neurology, 44*(2), 273–278.

Bochsler, T. M., Kallie, C. S., Legge, G. E., & Gage, R. (2011). Recognition of ramps and steps by people with low vision. *Investigative Ophthalmology & Visual Science, 52*(14), 1910.

Bochsler, T. M., Legge, G. E., Kallie, C. S., & Gage, R. (2012). Seeing steps and ramps with simulated low acuity: Impact of texture and locomotion. *Optometry & Vision Science, 89*(9), 1299–1307.

Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. R. (2013). Scene categorization at large visual eccentricities. *Vision Research, 86*, 35–42.

Cutting, J. E., Vishton, P. M., Flückiger, M., Baumberger, B., & Gerndt, J. D. (1997). Heading and path information from retinal flow in naturalistic environments. *Perception & Psychophysics, 59*(3), 426–441.

Crowell, J. A., Banks, M. S., Shenoy, K. V., & Andersen, R. A. (1998). Visual self-motion perception during head turns. *Nature Neuroscience, 1*(8), 732–737.

Crowell, J. A., Maxwell, M. A., Shenoy, K. V., & Andersen, R. A. (1998). Retinal and extra-retinal motion signals both affect the extent of gaze-shift compensation. *Investigative Ophthalmology & Visual Science, 39*(4), S1093.

Domini, F., & Caudek, C. (2003). 3-D structure perceived from dynamic information: a new theory. *Trends in Cognitive Sciences, 7*(10), 444–449.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

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

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*(2), 137–176.

Hildreth, E. C., & Royden, C. S. (2011). Integrating multiple cues to depth order at object boundaries. *Attention Perception & Psychophysics, 73*(7), 2218–2235.

Jobling, J. T., Mansfield, J. S., Legge, G. E., & Menge, M. R. (1997). Motion parallax: Effects of blur, contrast, and field size in normal and low vision. *Perception, 26*(12), 1529–1538.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin, 112*(1), 24–38.

Koenderink, J. J. (1986). Optic flow. *Vision Research, 26*(1), 161–179.

Koenderink, J. J., & Van Doorn, A. J. (1991). Affine structure from motion. *Journal of The Optical Society of America A-optics Image Science and Vision, 8*(2), 377–385.

Leat, S. J., Legge, G. E., & Bullimore, M. A. (1999). What is low vision?. A re-evaluation of definitions. *Optometry and Vision Science, 76*(4), 198–211.

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.

Lind, M. (1996). *The perception of rigid 3-D structure from optic flow (PhD dissertation).* Uppsala: Acta Universitatis Upsaliensis Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-115529.

Nakayama, K., & Loomis, J. M. (1974). Optical velocity patterns, velocity-sensitive neurons, and space perception: A hypothesis. *Perception, 3*(1), 63.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*(3), 353–383.

Norman, J. F., Beers, A. M., Holmin, J., & Boswell, A. M. (2010). Effective 3-D shape discrimination survives retinal blur. *Attention Perception & Psychophysics, 72*(6), 1569–1575.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. In Martinez-Conde, Macknik, Martinez, Alonso, & Tse (Eds.). *Progress in brain research* (pp. 23–36). Elsevier, B. V.

Pan, J. S., & Bingham, G. P. (2013). With an eye to low vision: Optic flow enables perception despite image blur. *Optometry and Vision Science, 90*(10), 1119–1127.

Pan, J. S., Li, J., Chen, Z., Mangiaracina, E. A., Connell, C. S., Wu, H., et al. (2017). Motion-generated optical information allows event perception despite blurry vision in amd and amblyopic patients. *Journal of Vision, 17*(12), 13.

Peyrin, C., Chauvin, A., Chokron, S., & Marendaz, C. (2003). Hemispheric specialization for spatial frequency processing in the analysis of natural scenes. *Brain and Cognition, 53*(2), 278–282.

Royden, C. S., Banks, M. S., & Crowell, J. A. (1992). The perception of heading during eye movements. *Nature, 360*(6404), 583–585.

Royden, C. S., Crowell, J. A., & Banks, M. S. (1994). Estimating heading during eye movements. *Vision Research, 34*(23), 3197–3214.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dpendent scene recognition. *Psychological Science, 5*(4), 195–200.

Sun, D., Roth, S., & Black, M. J. (2010). *Secrets of optical flow estimation and their principles.*

IEEE international conference on computer vision and pattern recognition.
Sun, D., Roth, S., & Black, M. J. (2014). A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision, 106*(2), 115–137.
Todd, J. T. (1995). The visual perception of three-dimensional structure from motion. *Perception of Space and Motion,* 201–226.
Uebersax, J. S. (1987). Diversity of decision-making models and the measurement of interrater agreement. *Psychological Bulletin, 101*(101), 140–146.
Warren, W. H. (1998). The state of flow. In T. Watanabe (Ed.). *High-level motion processing* (pp. 315–358). Cambridge, MA: MIT Press.
Warren, W. H., & Hannon, D. J. (1990). Eye movements and optical flow. *Journal of The Optical Society of America A, 7*(1), 160–169.
Warren, W. H., Morris, M. W., & Kalish, M. (1988). Perception of translational heading from optical flow. *Journal of Experimental Psychology: Human Perception and Performance, 14*(4), 646.
Whitaker, D., & Buckingham, T. (1987). Oscillatory movement displacement thresholds: Resistance to optical image degradation. *Ophthalmic and Physiological Optics, 7*(2), 121–125.
Wu, H. (2018). *Optic flow and image structure information allow blurry scene perception and blurry event perception (Unpublished Master's thesis)*. Guangzhou, China: Sun Yat-Sen University.