Vision Research 98 (2014) 14–25

Contents lists available at ScienceDirect

Vision Research

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

Detecting moving objects in an optic flow field using direction- and speed-tuned operators

Constance S. Royden *, Michael A. Holloway

Department of Mathematics and Computer Science, College of the Holy Cross, United States

ARTICLE INFO

Article history:
Received 4 October 2013
Received in revised form 25 January 2014
Available online 4 March 2014

Keywords:
Optic flow
Heading
Motion
Moving object detection
Modeling

ABSTRACT

An observer moving through a scene must be able to identify moving objects. Psychophysical results have shown that people can identify moving objects based on the speed or direction of their movement relative to the optic flow field generated by the observer's motion. Here we show that a model that uses speed- and direction-tuned units, whose responses are based on the response properties of cells in the primate visual cortex, can successfully identify the borders of moving objects in a scene through which an observer is moving.

1. Introduction

When an observer moves through the world, the motion of images on the retina form a dynamic pattern known as the optic flow field. This motion contains a wealth of information about the world around the observer, including information about the observer's direction of motion, the relative distance to objects in the visual field and whether or not objects are moving relative to the rest of the scene (Clocksin, 1980; Gibson, 1950; Longuet-Higgins & Prazdny, 1980; Nakayama & Loomis, 1974; Royden & Picone, 2007; Thompson & Pong, 1990). In terms of recognizing moving objects, it is clear that moving observers can detect objects that are moving relative to the scene, but it is unclear how the visual system accomplishes this task. The difficulty arises because the observer's own motion creates motion throughout the visual image, even for images of stationary items, and thus the image motion of the self-moving object is not unique. While there have been several theoretical models proposed that could detect moving objects within the optic flow field (Hildreth, 1992; Thompson & Pong, 1990), there has been no analysis of how these mechanisms could be implemented by biological neurons. In the current study, we show how a model that is based on the response properties of neurons in the primate visual system, using operators with speed- and direction-tuned responses to motion, can be extended to detect moving objects in the visual scene.

1.1. Theoretical considerations

When an observer moves in a straight line through a stationary scene, the optic flow field forms a radial pattern (Fig. 1a). The center of this pattern, where the image motion is zero, is known as the focus of expansion (FOE) and corresponds to the observer's direction of motion, or heading. A moving object in the scene may introduce image velocities that are inconsistent with this pattern (Fig. 1b), and this inconsistency, in theory, can be used to detect the presence of a moving object. Thompson and Pong (1990) proposed a computational model in which one could identify a moving object if its motion differed in direction or speed from the expected optic flow field generated from a given camera motion. In their analysis they noted that knowledge of the relative depth of points in the scene, either from binocular stereo or monocular depth cues such as familiar size, can aid in detecting moving objects based on motion discontinuities.

When the observer is rotating as well as translating, such as when moving on a curved path or tracking an object with eye or head movements, the optic flow field becomes considerably more complex (Fig. 1c), making it more difficult to detect a moving object. Longuet-Higgins and Prazdny (1980) presented a mathematical analysis that showed how one could use local motion subtraction of image velocities to eliminate image motion due to rotation. The resulting difference vectors form a radial pattern with a center coinciding with the observer's translational direction of

* Corresponding author. Address: Department of Mathematics and Computer Science, P.O. Box 48A, College of the Holy Cross, Worcester, MA 01610, United States.
E-mail address: croyden@cs.holycross.edu (C.S. Royden).

http://dx.doi.org/10.1016/j.visres.2014.02.009
0042-6989/

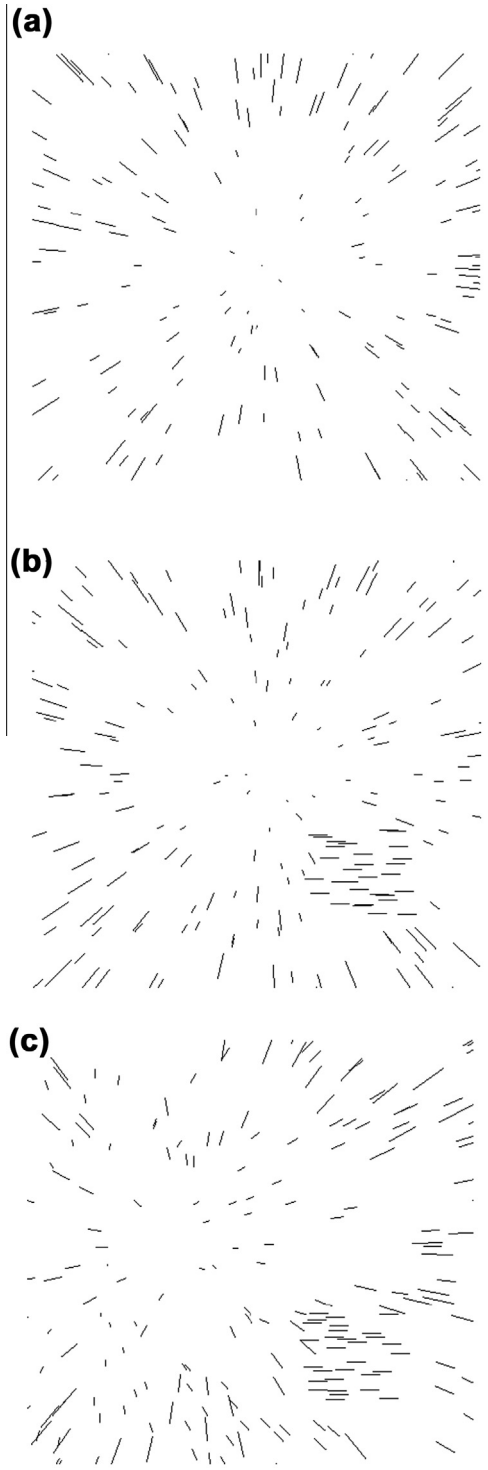

Fig. 1. Optic flow fields. (a) Radial optic flow field generated by an observer moving toward the center of a scene consisting of two planes at different depths. (b) Radial optic flow field with an object located in the lower right. (c) Optic flow field for an observer who is both translating and rotating, with a moving object in the lower right.

motion. Rieger and Lawton (1985) further showed that one could compute heading reasonably accurately using motion subtraction for two points separated on the image plane by a small amount. Hildreth (1992) showed that one could use local motion differences to determine an observer's direction of motion. Once the direction of motion is known, moving objects can be identified based on difference vectors whose angle does not fit the expected radial pattern for that observer heading.

The theoretical models discussed above rely on accurate measurements of both the direction and speed of 2D image velocities within a region of the visual field to compute the heading and the location of moving objects. This might lead one to assume that biological visual systems would also need to compute image velocities relatively accurately to accomplish these tasks. Because the initial neural processing of motion in primate visual cortex involves tuned responses to the direction and speed of motion (Maunsell & van Essen, 1983), an additional stage of processing would be required to compute the actual speed and direction of motion. For example, Priebe and Lisberger (2004) and Perrone (2012) have shown how this can be accomplished. Here, we ask whether one can accomplish these tasks without calculating speed and direction of image motion explicitly.

The model presented here is an extension of a model (Royden, 1997; Royden & Picone, 2007) that is based on the analysis of Longuet-Higgins and Prazdny (1980), but uses mechanisms that are based on the response properties of neurons in the primate visual system, such as speed- and direction-tuning, to compute heading. Here we extend the model to identify the borders of self-moving objects in the scene, by identifying locations where local motion differences lead to responses that differ from those expected for motion through a stationary scene.

1.2. Psychophysical evidence

It is clear from everyday experience that people can detect and interact with moving objects in a scene. Sports players can track moving balls and other players, and drivers can identify other moving vehicles or pedestrians in the scene. Royden and Connors (2010) showed that people can detect a moving object whose angle of motion differs from the radial flow pattern generated by an observer moving in a straight line. They reported a threshold detection angle of 10.3 deg deviation from the radial flow lines. They also noted that the global pattern was important for this detection, since people performed much more poorly when detecting an object deviating from a deformation pattern. Royden and Moore (2012) showed that people can also detect moving objects based on their speed. In their experiments, people detected objects whose speed was either 1.4 times faster or 0.6 times slower than it would be if it were part of the optic flow field (a 40% change in each case). From these two studies it is clear that people can detect an object for which the image motion varies in angle or speed from the optic flow field generated by an observer's straight line motion.

Warren and Rushton (2007, 2008, 2009; Rushton & Warren, 2005) have examined people's judgments of object trajectory when they are moving, and shown that the visual system appears to subtract out the optic flow due to the observer's own motion when computing the trajectory of the moving object. How this subtraction is accomplished by the visual system is an open question. Here we examine one potential model for motion subtraction to detect moving objects based on the response properties of neurons in the primate visual cortex.

2. The computational model

2.1. Mathematical derivation

The model for computing heading has been described in detail previously (Royden, 1997; Royden & Picone, 2007), but we provide a description of the main features here. To understand the mechanisms used by the model, we first review the mathematical underpinnings for determining heading for an observer undergoing

both translation and rotation, as first analyzed by Longuet-Higgins and Prazdny (1980). They showed that, for an observer moving with instantaneous translational velocity given by a vector (T_x, T_y, T_z) and rotational velocity given by the vector (R_x, R_y, R_z), one can calculate the image velocity for a point $P = (X, Y, Z)$, whose image projects onto an image plane at position $p = (x,y) = (X/Z, Y/Z)$, if the image plane is one unit of distance from the center of projection. Given these parameters, the velocity of image point, p, is given by

$$v_x = \frac{xT_z - T_x}{Z} + xyR_x - (1 + x^2)R_y + yR_z$$
$$v_y = \frac{yT_z - T_y}{Z} + (1 + y^2)R_x - xyR_y - xR_z \quad (1)$$

where v_x and v_y are the horizontal and vertical components of the image velocity, respectively.

Longuet-Higgins and Prazdny (1980) noted that these equations are separable into two components, one of which depends on observer translation but not rotation, and the other depends only on the rotation. The component that depends on translation is also inversely proportional to the depth, Z, of the image point, whereas the component dependent on rotation does not depend on depth. Therefore, if one could measure the image velocities for two points at different depths along a line of sight, e.g. on either side of a depth edge, then subtracting one of these velocity vectors from the other generates a difference vector given by

$$v_{xd} = (-T_x + xT_z)\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right)$$
$$v_{yd} = (-T_y + yT_z)\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right) \quad (2)$$

where v_{xd} and v_{yd} are the horizontal and vertical components of the difference vector and Z_1 and Z_2 are the depths of the two points in space. Note that the difference vectors depend only on the observer's translational motion. The rotational motion has been eliminated. Longuet-Higgins and Prazdny (1980) showed that these difference vectors form a radial pattern, the center of which corresponds to the observer's translational direction of motion. Thus, one can use the difference vectors to compute the observer's heading direction.

The analysis of Longuet-Higgins and Prazdny relies on an observer moving through a stationary scene. If there are self-moving objects present in the scene, then the difference vectors generated at the borders of those objects will differ from the radial pattern generated by the difference vectors. Our model, described below, uses neural-based mechanisms to estimate the difference vectors throughout the visual field and from these estimate the observer's direction of motion. In this study, we extend the model to identify locations where the difference vectors do not fit the radial pattern, and thus indicate the possible presence of a moving object.

2.2. Model using speed and direction tuned units

The Longuet-Higgins and Prazdny (1980) analysis cannot be used directly to compute observer heading, because in practice one cannot measure two different image velocities along the same line of sight. Rieger and Lawton (1985) showed that performing the vector subtraction for image velocities separated by a small amount on the image plane can also yield good estimates of observer heading. However, the visual system does not compute precise image velocity vectors, but rather has motion sensitive neurons that are tuned to the speed and direction of the 2-dimensional image motion in their receptive fields (Maunsell & van Essen, 1983). Thus, our model makes use of operators whose responses to motion are tuned to speed and direction, similar to the response properties of cells in the primate middle temporal visual area, MT (Maunsell & van Essen, 1983). In addition, the model operators have receptive fields that are divided into an excitatory region and an adjacent inhibitory region, based on the classical receptive field and adjacent inhibitory surrounds described for some MT cells (Allman, Miezin, & McGuiness, 1985; Raiguel et al., 1995; Xiao et al., 1995). The excitatory and inhibitory regions approximate the motion subtraction that eliminates the observer rotation component in the image velocities.

The model for computation of heading consists of two layers of cells (Fig. 2). In the first layer, each region of the visual field is processed by a group of cells whose receptive fields vary in terms of their direction tuning, their speed tuning and the orientation of the axis between the excitatory and inhibitory regions. The direction tuning is modeled as a cosine curve, with the response truncated at ±90 deg so the operator does not give negative responses. Thus the direction tuned response is given by the equation:

$$R_{dir} = v_{avg}\cos(\phi - \theta) \quad (3)$$

where v_{avg} and ϕ are the average speed and direction, respectively, of the image velocity within the excitatory (or inhibitory) region of the operator's receptive field, and θ is the preferred direction of the operator. While we chose a cosine tuning curve based on the average tuning properties of cells in MT (Maunsell & van Essen, 1983), previous studies have shown that the model also computes heading well with somewhat increased or decreased tuning widths (Royden, 1997). Speed tuning is modeled as a Gaussian centered on the preferred speed of the operator, with the speed tuned response given by

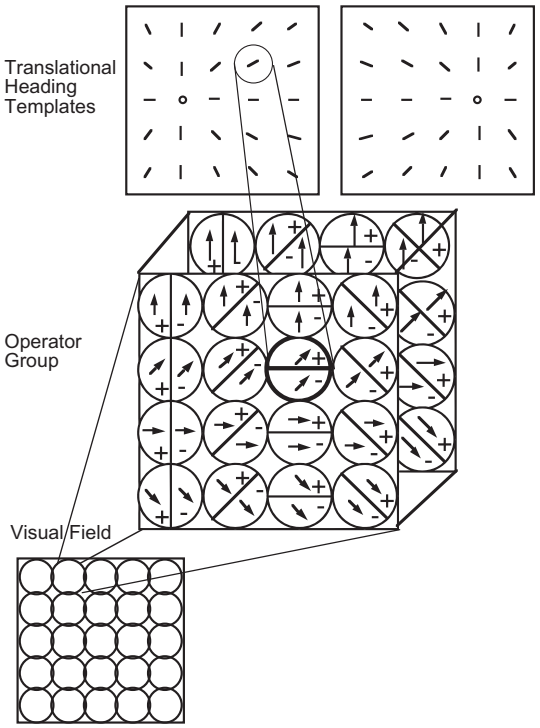

Fig. 2. Computational model. The bottom drawing shows the visual field divided into individual receptive field locations. The middle drawing shows the group of operators that process one location in the visual field. The operators vary in their preferred direction of motion, indicated by the direction of the arrows in the receptive fields, their preferred speeds, indicated by the length of the arrows, and the angle of the differencing axis between the excitatory and inhibitory regions. Only a subset of the operators used are shown here. The top image shows two receptive fields of template cells in the second layer that are tuned to radial patterns. These cells vary in the position of the center of their preferred radial pattern.

$$R_{speed} = e^{-0.5(\log_2(\frac{R_{dir}}{R_{pref}}))^2} \quad (4)$$

where R_{dir} is the direction-tuned response and R_{pref} is the preferred speed of the operator. This equation is based on the speed tuning properties of cells in MT (Maunsell & van Essen, 1983).

The speed and direction tuned response is computed separately for the excitatory region and the inhibitory region of the operator. The full operator response is computed by subtracting the inhibitory response from the excitatory response. If the result is less than zero, the response of the operator is set to zero. Thus the full operator response is given by

$$R_{op} = e^{-0.5\left(\log_2\left(\frac{v_{avg+}\cos(\phi_+ - \theta)}{R_{pref}}\right)\right)^2} - e^{-0.5\left(\log_2\left(\frac{v_{avg+}\cos(\phi_- - \theta)}{R_{pref}}\right)\right)^2} \quad (5)$$

where v_{avg+} and ϕ_+ are the speed and direction, respectively, of the average image velocity in the excitatory region of the receptive field and v_{avg-} and ϕ_- are the speed and direction of the average image velocity in the inhibitory region of the receptive field.

The operator with the largest response in each region of the visual field projects to a second layer of cells that act as templates for radial patterns of input, each with a different center of expansion. A cell in this layer receives support from the previous layer if the preferred direction of motion of that cell matches the direction of the radial pattern for that template at that position. The preferred direction of the input cell must be within some threshold of the template direction, determined by the operator spacing. Furthermore, cells with receptive fields that are closer to the template cell's preferred center of expansion contribute more to its excitation than those further away, as described in Royden (2002). These template cells have some properties in common with cells in the Medial Superior Temporal area (MSTd), such as a preference for radial patterns and for different locations of the center of the preferred pattern (Duffy & Wurtz, 1991; Duffy & Wurtz, 1995; Graziano, Andersen, & Snowden, 1994; Saito et al., 1986; Tanaka & Saito, 1989).

The model computes heading well under a variety of conditions, including observer rotations (Royden, 1997). It computes heading as well as humans when a moving object is present, showing a small bias when the object is over the focus of expansion (Royden, 2002). In addition it shows a bias in heading computation similar to that of humans when presented with an optic flow field consisting of a radial motion pattern superimposed on a field of laterally moving dots. In this case, both humans and the model show a bias in the direction of the laterally moving dots (Royden & Conti, 2003). Furthermore, the response magnitudes of the cells in the first layer can be used as an indication of the relative sizes of depth edges in the scene, allowing for the construction of a relative depth map of the visual field (Royden & Picone, 2007). Because the motion subtraction eliminates rotation effects, this depth map is stable even when the observer is rotating.

To extend the model to identify moving objects, we added a second stage to the computation. Once the heading is computed by identifying the maximally responding template cell in the second layer, the program examines the preferred direction and response magnitude of each of the maximally responding cells in the first layer. If any of these have a preferred direction that differs significantly from the direction expected in the radial pattern of the computed heading, that cell is identified as signaling a potential border for a moving object. Additionally, if any of the maximally responding first layer cells have a response magnitude that is significantly higher than the response magnitudes of other cells in the network, those cells are also identified as signaling potential moving object borders. The output of the model indicates the positions of potential moving object border by drawing a black circle at each location where it identifies a potential border. The diameter of the circle is proportional to the magnitude of the operator's response, which also serves as a measure of reliability, because operators with small responses tend to be more affected by noise in the system.

3. Model simulations

All simulations, unless otherwise noted, simulated observer motion toward two planes of 500 dots, half of which were at a distance of 400 cm and half were at 1000 cm from the observer. The background dots were randomly distributed within a 30 × 30 deg viewing window. The 6 × 6 deg object consisted of 50 dots and, unless otherwise noted, its center was located 7 degrees to the right and 7 degrees down from the center of the scene. The object was opaque, so no background dots overlapped the object's image position. The object was 400 cm from the observer. The observer's translational motion was simulated toward the specified heading at a speed of 200 cm/s. The object moved laterally with respect to the observer, with a speed of 52.6 cm/s (7.5 deg/s), unless otherwise noted. For each point in the scene, its image position and image velocity were calculated based on the observer or object motion parameters. The image velocities of the points were used as input to the model.

The viewing window was divided into a 13 × 13 set of circular regions (representing the receptive fields), each of which was 2 deg in radius, similar to the size of the classical receptive fields of MT cells near the center of the visual field (Felleman & Kaas, 1984). The centers of these regions were spaced every 2 deg, from −12 deg to 12 deg in both the horizontal and vertical directions, in order to adequately cover the full visual field. Each region was processed by a set of 2688 operators, which varied in preferred direction of motion, the angle of the differencing axis between the excitatory and inhibitory regions and the speed tuning. We used 24 preferred directions of motion, spaced every 15 deg from 0 to 360, 16 preferred differencing axes spaced every 22.5 deg, and 7 preferred speeds: 0.5, 1, 2, 4, 8, 16 and 32 deg/s. These parameters were chosen to cover the range of directions and speeds present in our stimuli. The response of each operator was calculated as described above. The second layer of cells consisted of 169 template cells, each tuned to a radial pattern. The centers of the patterns for this layer were spaced every 2 deg from −12 to 12 deg in both the horizontal and vertical directions.

3.1. Object detection using angle alone

We first tested the ability of the model to identify object borders based on the angle of the preferred direction of motion of the maximally responding operators. Potential moving object edges were identified for any maximally responding operator whose preferred direction of motion differed from the radial pattern of the computed heading by more than a given threshold angle. We tested thresholds of 15, 20, 25 and 30 deg. Fig. 3 shows an example of the flow field with the responses of operators identifying potential moving object edges for a threshold angle of 25 deg. It can be seen in the figure that most of the edges of the moving object are identified as potential moving object borders, with 11 of 12 possible edge positions yielding responses. However, it can also be seen that there are a number of scene positions, 17 out of a possible 153, that are also identified as potential moving object borders.

To quantify this result, we ran the simulation 5 times for each condition and found the average number of border positions and the average number of background positions identified as possible moving object borders. These are graphed in Fig. 4a as a percentage of the total possible operators (12 for the border and 153 for the background). The number of operators located at the border of

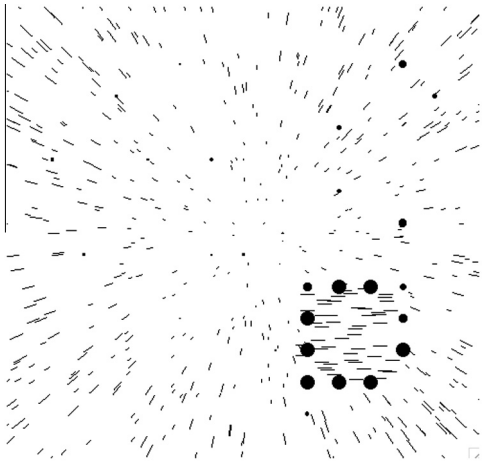

Fig. 3. Example of operator responses for a threshold angle of 25 deg. Each line indicates the image velocity of a point in the scene. The black circles show where the model has identified the potential border of a moving object. The radius of each circle is proportional to the response magnitude of the unit signaling the presence of a border.

the moving object that correctly responded ranged from 6.6 (55%) for the highest angle threshold of 30 deg to 9.6 (80%) for the 20 deg threshold. Generally the model had more difficulty identifying corners, presumably because the regions of the operator were not evenly distributed across the border in these cases. The number of background operators incorrectly indicating a moving object border ranged from 14.6 (9.5%) for a the 30 deg threshold to 37.6 (24.6%) for the 15 deg threshold. Unlike the operators on the border of the object, the operators that responded in the background tended to be scattered, and did not form lines or groups. Note that in all our simulations, operators on the interior of the object almost never indicated a potential border. This is likely because there is no depth variation in this region of the scene, so the operators' responses would tend to be zero.

It is clear that although lower angle thresholds give higher percentages of operators on the border of the object responding, they also lead to more noise, i.e. operators in the background part of the scene responding. The responding operators in the background had response magnitudes that were much smaller than those on the edges of the object, with a roughly 3.5-fold difference in the average response magnitudes between these two groups of operators. The average response magnitudes are shown in Fig. 4b.

Operators with small response magnitudes are more subject to noise generated by the random distribution of points within the operator's receptive field. We hypothesized that if the operators in the background were responding due to noise, then this noisy response would average out over multiple trials with different distributions of points, whereas the responses of operators on the actual borders of the object would be more consistent and therefore their responses would not average out. We therefore ran the simulation five times with different distributions of dots, and for each operator averaged the angle differences between the preferred direction of the maximally responding operator and the predicted radial flow field. We then used the averaged angle difference to identify potential borders of moving objects. We repeated this five times and the average numbers of locations on the border and in the background are shown in Fig. 4a, indicated by the dashed lines. This eliminated many of the background responses, but only decreased the number of border operators by a small amount. For example, for the 20 deg angle threshold, the number of border operators responding decreased by about 1, from 9.6 (80%) to 8.2

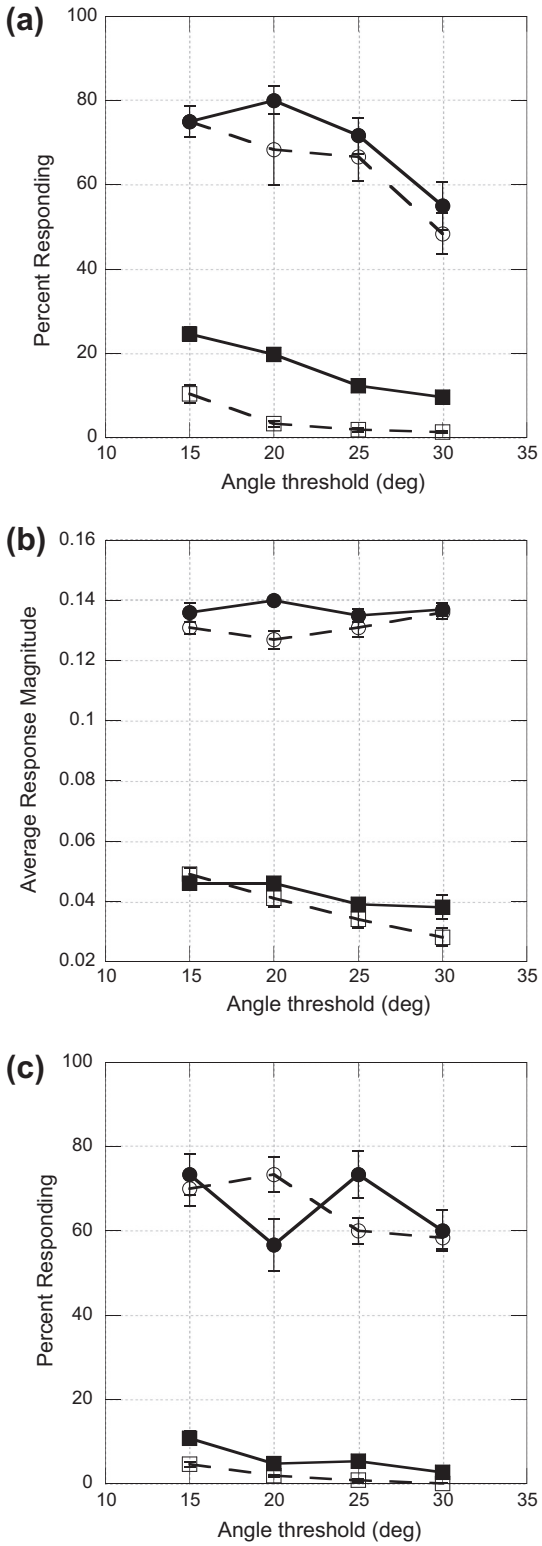

Fig. 4. Results of simulations testing angular thresholds. (a) Percentage of operators indicating a moving object border. Circles show the percentage of operators on the border correctly indicating a border. Squares show the percentage of operators in the background incorrectly indicating a border. Filled symbols show results when the calculations are from a single distribution of dots. Open symbols show results when the calculations are from the average of five distributions of dots. Error bars indicate ±1 standard error. (b) Response magnitudes of the operators from (a). All symbols are as in (a). (c) Results when only operators with response magnitudes above 0.05 are used to indicate moving object borders. All symbols are as in (a).

(68%), while the number of background operators decreased by 25, from 30.2 (19.7%) to 5 (3.3%).

One can also eliminate some of the noisy responses by setting a threshold level for the magnitude of the responses, with only operators whose responses exceed the threshold used for identification of moving objects. Fig. 4c shows the average percentage of operators responding on the border and in the background with the response threshold set to 0.05, which is approximately 80% of the median response magnitude for operators in the current conditions. The operators on the border of the object still identified between 6.8 (56%) and 8.8 (73%) positions, while the number of operators responding in the background was reduced to between 16.4 (10.7%) and 4 (2.6%). The dashed lines in Fig. 4c show the results using both noise reduction methods. We ran the simulation 5 times and averaged the angle differences over those 5 trials, then determined which operators exceeded both the angular and response magnitude thresholds. We repeated this 5 times and averaged the number of responding operators. The results show that the background noise is reduced to zero for the angular threshold of 30 deg. This decreased the number of correct responses by operators on the object border by a small amount, with the average number ranging between 7 (58%) for the 30 deg angular threshold and 8.8 (73.3%) for the 20 deg threshold. So there is a small tradeoff for reducing the number of noise responses in that 1 or 2 fewer operators on the border of the object give correct positive responses.

3.2. Object detection using speed alone

In theory, one can use speed differences to identify borders of moving objects. There are situations when this might be useful, for example when the optic flow vectors associated with the moving object are close to the same direction as the radial field generated by the observer. In this case, one can make use of speed differences to identify moving objects. Speed is a somewhat ambiguous cue, because speed can change due to a change in the depth of the scene or due to motion of the object. However, Royden and Moore (2012) showed that humans can use speed cues to identify moving objects in the scene, and Royden and Holloway (2007) showed that people perceive objects with image speeds faster than the background as moving. We first tested how well the model would identify the moving object by detecting locations where there is a large speed change between the background and the object, leading to a large operator response. For these simulations, we set the angular threshold to zero. Because the magnitude of the optic flow vectors increases with the distance from the focus of expansion, we normalized each response by dividing by the distance of the operator from the computed heading position (measured in degrees) and multiplying by 100. We then compared the normalized responses to a threshold magnitude. Only the normalized responses that were higher than the threshold were accepted as indicating potential moving object borders. As in the other simulations, we ran each simulation five times and recorded the average number of operators responding.

Fig. 5 shows the results for normalized response thresholds ranging from 0.6 to 1.8. The lowest is slightly lower than the average normalized response (which for these conditions was 0.7). As might be expected, for the lowest threshold, nearly all positions on the object border were detected, with an average of 11.2 (93.3%) out of 12 operators responding, but there is also a large response from the background operators, with an average of 79.2 (51.8%) responding. As the response threshold increases, the number of operators responding decreases for both the border and the background. At a threshold of 1.0, roughly 1.5 times the average normalized response, there appears to be a good balance between the number of border locations correctly identified, at 8.6 (71.7%),

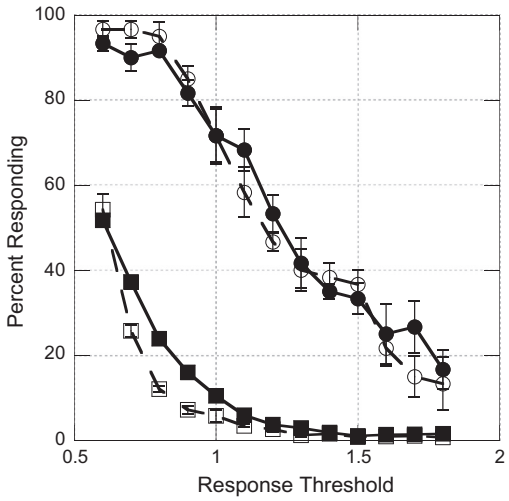

Fig. 5. Results of simulations testing response magnitude thresholds. Circles indicate the percentage of operators on the border correctly indicating a border. Squares indicate the percentage of operators in the background incorrectly indicating a border. Filled symbols show results when the calculations are from a single distribution of dots. Open symbols show results when the calculations are from the average of five distributions of dots. Error bars indicate ±1 standard error.

and the amount of background noise, at 16.0 (10.5%). The response magnitudes were much less variable than the preferred angles of the maximally responding operators. When we ran the simulations 5 times and averaged the response magnitudes for each operator before comparing to the threshold magnitude, the results were very similar to the individual runs, as can be seen by comparing the dashed lines (computed using the averaging technique) and the solid lines in Fig. 5, although there is some reduction in noise for the background for thresholds below 1.0.

3.3. Object detection using both angle and speed

We reasoned that we could combine angle and speed to maximize the number of operators on the border of the object that correctly indicated the presence of a moving object. In this set of simulations, we set an absolute response threshold of 0.05 (before normalization), and accepted any responses whose magnitude was

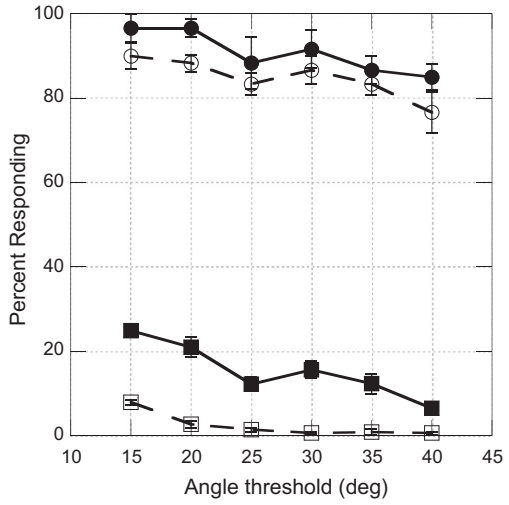

Fig. 6. Results of combining angle and response thresholds. The normalized response threshold was set to 1.0, with an absolute response floor of 0.05. The angle threshold was varied between 15 and 40 deg. All symbols are as in Fig. 5.

above that and either had an angle above a given angle threshold or had a normalized response above a given normalized response threshold. Fig. 6 shows the results with a normalized response threshold of 1.0 and for angle thresholds ranging between 15 and 40 deg, spaced every 5 deg. The bold lines show the results when the responses are computed based on a single trial. As before, the simulations were run 5 times and the results averaged. The dashed lines show results when the responses were computed based on the angular differences and normalized responses averaged over 5 simulations. These results were also computed five times and averaged. The border of the object is identified very well, with the average number of operators on the border correctly identifying the moving object border ranging between 10.2 (85%) for the 40 deg threshold and 11.6 (96.7%) for a 15 deg threshold for the single trial condition. Averaging the angles and speeds over 5 trials before comparing them with the thresholds yielded only slightly smaller numbers, with between 9.2 (76.7%) and 10.8 (90.0%) of the operators correctly responding for the 40 and 15 deg thresholds respectively. The number of background operators incorrectly indicating a moving object ranged from 9.8 (6.4%) to 38 (24.8%) for the single trial condition. When the angles and speeds were averaged over 5 trials, these numbers decreased, ranging between 0.8 (0.5%) and 12 (7.8%) for the 40 deg and 15 deg threshold angles, respectively. Thus, the model is able to detect the border of a moving object reasonably well while generating fairly few noise responses in the background.

For the remaining tests, we chose an angle threshold of 25 deg, as it appears to give a reasonable compromise between detecting the moving object borders while not generating many responses in the background. We tested the effect of changing the object's speed, the object's location, the observer's heading and the observer's rotation. For each of these we used an angle threshold of 25 deg, a normalized response threshold of 1.0 and an absolute response threshold of 0.05. In each condition we tested the results when the identification of borders was based on a single simulation, with the numbers averaged over five runs. We also tested results when the identification was based on the angle and magnitude of responses averaged over 5 trials. The numbers of operators responding were then averaged from the results of 5 repetitions.

3.3.1. Effect of object speed

We tested eight object speeds evenly spaced between 12.5 cm/s (1.8 deg/s) to 100 cm/s (14 deg/s) using the thresholds described above. Fig. 7a shows the results. As expected, few of the object border positions are detected for the lowest speeds. The number rises quickly and for speeds of 37.5 cm/s (5.4 deg/s) and above the model detects on average more than 8.8 (73.3%) locations, rising to nearly 100% for the fastest speed. The background response stays fairly constant, detecting a maximum of 28 (18.3%) for the 1 trial condition and a maximum of 3.4 (2.2%) for the averaged condition.

3.3.2. Effect of object location

We tested 9 object locations, with the object center located at all combinations of −7, −1 and +7 deg in the x and y directions. The results are shown in Fig. 7b. The results did not vary much for either the object border or the background. The number of operators correctly indicating the object border ranged from 10.2 (85%) to 12.0 (100%) in both the single trial and averaged trial results. For the background, the number of noise responses ranged from 19 (12.4%) to 25 (16.3%) in the single trial condition and from 1.0 (0.7%) to 4.2 (2.7%) in the averaged condition. Thus, the position of the object within the flow field does not appear to have a large effect on the model's ability to detect the object.

3.3.3. Effect of heading

We tested five heading directions, with the horizontal heading direction of −10, −5, 0, 5 and 10 deg from the center of the screen. The vertical direction was kept at zero. Fig. 7c shows the results. For the single trial condition, the number of object border locations identified increased from 6.8 (56.7%) to 11.8 (98.3%) between the leftmost heading of −10 deg to the rightmost heading of 10 deg. The results were quite similar for the 5 trial averaged condition. It seems likely that the decrease in the number of locations detected for the −10 deg heading is due to the fact that the flow field around the object has faster speeds for this heading and therefore the response magnitudes of the operators will be smaller and less likely to reach threshold. In addition, the angle of the flow field vectors near the object are more horizontal for this heading than for the other headings, and thus the angle of the difference vectors around the object are likely to be closer to the radial flow field pattern than for other headings tested.

The number of operators incorrectly identifying moving objects at non-object locations increases somewhat for the more peripheral headings, −10 and 10 deg. For the single trial condition, the number rises from 14.6 (9.5%) and 18.0 (11.8%) at the −5 deg and 5 deg headings, respectively, to 32 (20.9%) and 38.4 (25.1%) for the -10 and 10 deg headings, respectively. For the 5 trial averages, these numbers are 1.6 (1.0%), 2.8(1.8%), 8.0 (5.2%) and 8.2 (5.4%) for the −5, 5, −10 and 10 deg headings. The increase in noise at the more peripheral headings likely arises because the image velocities are higher at the edge of the scene opposite the peripheral heading, leading to higher response magnitudes, some of which exceed threshold.

3.3.4. Effect of rotations

One of the benefits of image motion subtraction for computing heading is that it eliminates the effects of observer rotation when computing heading direction. Thus, one would expect that rotation would not have a large effect on the identification of moving objects. We tested the model for a heading of (0, 0) and rotations about a vertical axis of 0, ±2.5 and ±5 deg/s. As can be seen in Fig. 7d, the model detected the borders of the object well for all rotations. For the single trial condition, the number of border operators correctly detecting an object border rose from 10.0 (83.3%) to 11.8 (98.3%) for the −5 and +5 deg/s rotations, respectively. The results were similar for the simulations that averaged over 5 trials. As with the results for heading, there was somewhat more noise for the higher rotation rates, with a maximum of 38.6 (25.2%) of scene operators detecting object borders in the case of the −5 deg/s rotation rate, and a minimum of 16.8 (11.0%) for the rotation rate of −2.5 deg/s. For the averaged trials, the maximum was 17.2 (11.2%) for a rotation rate of −5 deg/s and the minimum was 2.0 (1.3%) for a rate of 2.5 deg/s. As with the heading conditions, the increase is likely due to the faster image speeds at the edges of the viewing window that occur with the higher rotation rates.

3.4. Performance with deteriorated flow fields

When people view simulations of translational motion, i.e. without rotations, their ability to judge heading remains fairly robust in the presence of angular noise in the velocity field or when the density of points in the flow field is small (Foulkes, Rushton, & Warren, 2013a; Warren et al., 1991; Warren, Morris, & Kalish, 1988). In addition, perception of the 2D trajectory of an identified moving object within a radial flow field remains fairly consistent at low dot densities and deteriorates slowly as angular noise is added to the velocity field (Foulkes, Rushton, & Warren, 2013b). Finally, Warren, Rushton, and Foulkes (2012) showed that the radial flow field affects perception of object trajectory even when the flow field is only presented in half the visual field and the object is pres-

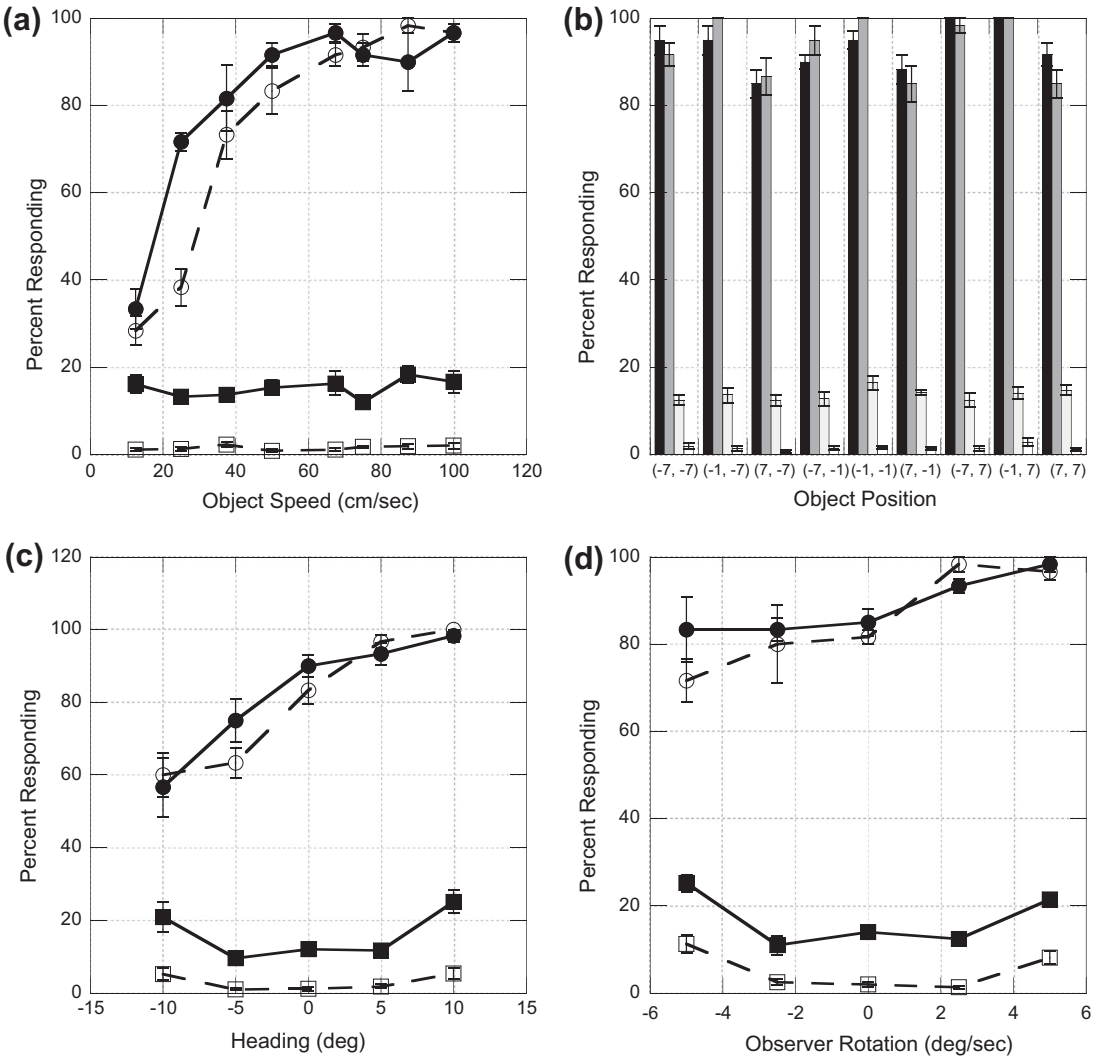

Fig. 7. Results of varying parameters. (a) Results of changing the object's speed. (b) Results of changing the object's position. (c) Results of changing the observer's heading. (d) Results of changing the rate of rotation about a vertical axis. For a, b and d, all symbols are as in Fig. 5. For (b), black and dark gray bars show the percentage of operators on the border correctly indicating a border, with black showing the single trial condition and dark gray showing the 5 trial condition. Light gray and white bars show the percentage of operators incorrectly indicating a border on the background, with light gray bars showing the single trial and white bars showing the 5 trial conditions. Error bars indicate ±1 standard error.

ent in the other half. While Royden (1997) showed that the model's accuracy in detecting heading is fairly robust in the presence of angular noise, the other conditions were not tested. Therefore we tested how well the current model performed under similar conditions that deteriorate the quality of the flow field. In addition, we tested the model's accuracy for two different scene configurations: a 3D cloud of points and a ground plane. For each of these we used the same parameters used in Section 3.3, i.e., an angle threshold of 25 deg, a normalized response threshold of 1.0 and an absolute response threshold of 0.05.

3.4.1. Addition of angular noise

In this set of simulations, we added angular noise to the flow field. The direction of each velocity vector was perturbed by an additive angle chosen based on a zero mean Gaussian distribution. We tested the model's performance for standard deviations of 7.5 deg (low noise) and 15.0 deg (high noise), matching the parameters used in Foulkes, Rushton, and Warren (2013a). Fig. 8a shows the results. For the single trial condition, the correct detection of the object border remained highly accurate at all noise conditions tested, averaging 11.2 (93.3%) for the zero and low noise conditions, and 11.0 (91.7%) for the highest noise condition. Results were similar for the simulations averaged over 5 trials, with the correct identification of 10.8 (90.0%), 11.2 (93.3%) and 10.2 (85.0%) operators on the object border for the zero, low and high noise conditions, respectively.

The number of operators incorrectly indicating an object border where there was none increased gradually with added noise, but the model did not fail catastrophically at any noise condition. For the single trial condition, the number of incorrectly responding operators was 20.8 (13.6%) , 26.2 (17.1%) and 51.0 (33.3%) for the zero, low and high noise conditions, respectively. For the conditions averaged over five trials, these numbers all decreased, as in other simulations, to 0.6 (0.4%), 4.2 (2.7%) and 15.2 (9.9%) for the three increasing noise conditions. So, as with previous studies of heading detection and judgment of object trajectory, the model's accuracy decreases slowly with added noise, with the overall accuracy remaining reasonably high even at the highest noise level.

3.4.2. Decreased flow field density

We tested the model's ability to identify moving objects as we decreased the density of the dot distribution in the simulated

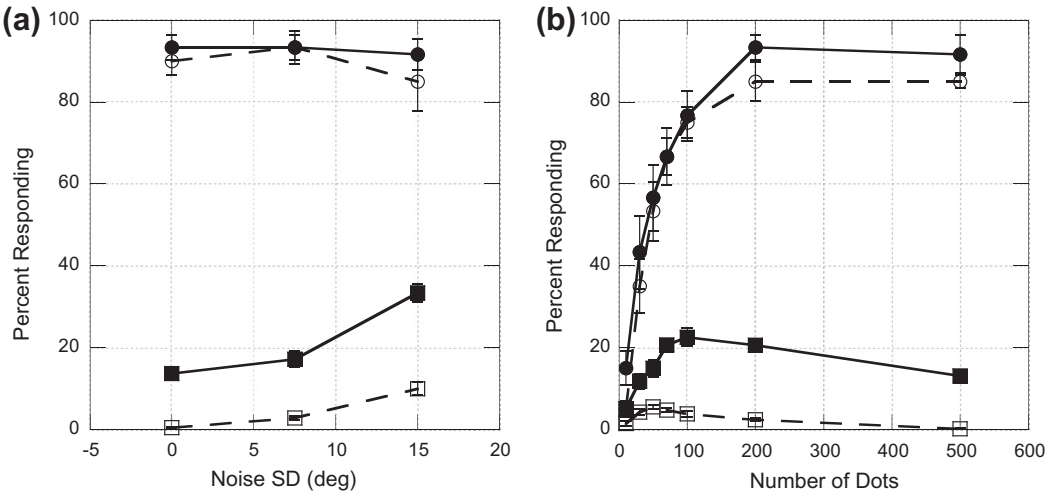

Fig. 8. Results of flow field deterioration. (a) Results of adding angular noise. (b) Results of decreasing dot density. All symbols are as in Fig. 5. Error bars indicate ±1 standard error.

scenes. To match the range of conditions used in (Foulkes et al., 2013a, 2013b) we tested the model with 10, 30, 50, 70, 100, 200 and 500 background dots. (Note that Foulkes et al. actually used 40% more dots in their model simulations than in the tests of human observers, so our 70 dot condition matches their 50 dot condition). The number of dots in the moving object was always one tenth of the number of background dots, e.g. there was 1 object dot for the 10 background dot condition and 50 object dots for the 500 background dot condition.

The results of these simulations are diagrammed in Fig. 8b. For the single trial condition, the number of operators correctly detecting the border of the object remains high for the 500 and 200 dot conditions (11 and 11.2 operators responding), and steadily decreases for the lower dot densities. Much of the decrease is due to the fact that as dot density decreases, more operators will have zero dots within their receptive fields, and thus they will give no response. Interestingly, as the dot density decreases, we start to see some of the operators responding to the interior of the object, which almost always generates a zero response for high densities. This is because, for the lower dot densities, some of these operators will be stimulated only in the excitatory half of their receptive field, because no dots fall in the inhibitory region, and thus the excitatory response is not offset by an equal inhibitory response. These operators are often correctly identified as being part of a moving object. If these interior operators are counted, the number of operators correctly indicating a moving object stays fairly constant at 11 down to the 70 dot condition. This decreases to 10 for 50 dots and 8 for the 30 dot condition. Results were qualitatively similar for the results averaged over five trials, with the number of operators detecting moving object borders decreased slightly over the single trial conditions, as can be seen in Fig. 8b. Thus, to the extent that the operators within the moving object have information available, they continue to respond correctly in the low density conditions.

For the single trial simulations, the number of operators incorrectly detecting moving object borders where none exist increases slightly between the 500 dot and the 100 dot conditions, from 20 (13.1%) to 34.4 (22.5%) operators, respectively. The numbers then decrease as the dot density decreases further. This decrease reflects the fact that fewer operators have any dots at all falling within their receptive fields for the lower dot densities. The results for the five trial averages behave similarly, but the total number of operators incorrectly responding is greatly decreased over the single trial conditions. In this case the number of operators rises from 0.2 (0.1%) for 500 dots to 8.4 (5.5%) for 50 dots, and then decreases as the dot density is further reduced. Thus, the model responds similarly to humans as dot density decreases, with accuracy declining only slightly down to 50 dots in the flow field.

3.4.3. Object not overlapping the flow field

In this simulation, we tested how well the object could be detected when velocity vectors were present in only half the visual field and the object was present in the other half, such that the object and flow field did not overlap. The model simulation limited the positions of the background dots to the left half of the visual field and the object was positioned in the lower right, at (7, −7), as in the other simulations. As with the other simulations, the angle threshold was 25 deg, and the normalized response threshold was 1.0 with an absolute response magnitude of 0.05. For the single trial condition, all 12 (100%) border operators correctly detected a moving object every time. An average of 12.4 (8.1%) operators incorrectly detected an object border in the background. For the conditions that averaged five trials, again all 12 border operators correctly detected a moving object border, and the number of background operators incorrectly detecting a border decreased to 7.4 (4.8%). Thus the model performs very well in this condition.

3.4.4. Response to a 3D cloud or a ground plane

Royden (1997) showed that the model computed heading well when presented with a 3D cloud of dots or a ground plane. This suggests that it will also perform well when detecting moving objects within these scenes. We tested model performance for a 3D cloud of points, where the depth of the 500 background points ranged between 400 and 1000 cm, and for a ground plane, positioned at -160 cm below eye-height and extending to 10,000 cm in the distance. For the 3D cloud of dots, an average of 11 (91.7%) operators correctly detected the object borders for both the single trial and five trial conditions. An average of 7.2 (4.7%) and 0.8 (0.5%) incorrectly signaled a moving object in the background for the single and five trial conditions, respectively. In the case of the ground plane, 11 (91.7%) operators correctly detected the object borders in the single trial condition and 11.4 (95%) detected them in the five trial condition. None of the background operators incorrectly detected moving object borders in either the single or the five trial condition. Thus, the model appears to

perform at least as well for these scene conditions as with the two transparent planes, possibly exhibiting less background noise than for the two planes.

4. Discussion

We have shown that a model that uses operators with motion processing properties based on those of neurons in the primate visual cortex can accurately identify the borders of moving objects in a scene through which an observer is moving. The results are remarkably robust over changes in heading, object position, added rotations, added angular noise and decreased dot density. In addition, a moving object can be detected when it is spatially separated from the background, or when the scene consists of a 3D cloud of points or a ground plane.

The model uses speed and direction differences within the image plane to identify moving object borders. The importance of this demonstration is primarily in the fact that the model was able to identify the location of the object using operators that are tuned to direction and speed, but does not explicitly compute the image velocity at any stage. None of the neurons in the input layer signal the exact direction of motion within their receptive fields. Rather, the response is tuned to a preferred direction as are neurons in the Middle Temporal area (MT) of primate visual cortex. Similarly, none of the neurons in this model explicitly signal the precise speed of the image velocity within their receptive fields. Each neuron is tuned to a specific speed, and its response is modulated based on the tuning curve. Nonetheless, the model is able to identify the borders of moving objects in two steps, by first computing the heading and then comparing the preferred directions of the maximally responding cells in the first layer to the expected radial pattern.

The results of the simulations are not 100% accurate. Not every border position on the object is identified, and there is some residual noise in the background region where locations are incorrectly identified as belonging to moving objects. Some of the noise is due to the fact that the computations are carried out using instantaneous measurements of image velocity. No doubt the temporal integration of responses accomplished by neurons would average out much of the noise in the background. We have shown that averaging over multiple measurements does in fact reduce the background noise. In addition, the visual system has lateral interactions not modeled here that can enhance responses when there are neighboring regions with strong responses, and can inhibit responses of isolated cells. Thus, we feel the small amount of inaccuracy generated by our model is at an acceptable level for detection of moving objects.

Note that for our final sets of simulations, the normalized response threshold was set to about 1.5 times the mean normalized response for the operators in layer 1. This is similar to the threshold factor of 1.4 needed by humans to detect moving objects based on speed (Royden & Moore, 2012). The chosen angle threshold of 25 deg was somewhat higher than the human angle threshold of 13 deg needed to detect moving objects (Royden & Connors, 2010), since the lower thresholds generated more background noise. It seems likely, however, that if one incorporated noise reduction techniques such as temporal integration or lateral inhibition, or if there were more preferred directions of motion represented in the first layer of cells, one could use an angular threshold closer to 13 deg without generating much background noise. Thus, the model's ability to detect moving objects is consistent with that of humans.

The model presented here uses motion-subtraction operators that are loosely based on cells found in primate visual area MT, which have adjacent excitatory and inhibitory receptive fields (Allman, Miezin, & McGuiness, 1985; Raiguel et al., 1995; Xiao et al., 1995) that could, in principle, perform this subtraction. This model has the advantage that it eliminates observer rotations and makes it easier to compute observer translational heading. However, the motion subtraction is not a requirement for the moving object detection. For example, the model presented by Perrone and Stone (1994) that uses more complex template cells to compute observer motion parameters, could likely also use a comparison of the preferred motion directions in the first layer of cells with the directional pattern within their more complicated template cells to identify locations that differ from the expected pattern.

We have not attempted to model exactly how the comparison between the radial pattern template cells and the first layer cells might be accomplished biologically. However, it is easy to imagine that the cells in the first layer could project to a separate area in addition to the template layer shown here (e.g. the area MSTl, which may be involved in identifying moving objects (Eifuku & Wurtz, 1998)). The radial template cells in the second layer of the model could inhibit the input of cells projecting to this second area if their preferred directions match the radial pattern. In this scenario, the only cells in this area that would respond would have input from regions in which the maximally responding cell (in the first layer) had a preferred direction that differed from that of the maximally responding template cell (in the second layer) that corresponds to the heading. There are undoubtedly other mechanisms that could also accomplish this computation.

(Warren & Rushton, 2007; Warren & Rushton, 2008; Warren & Rushton, 2009; Rushton & Warren, 2005) have advanced a theory for computing the trajectory of moving objects within an optic flow field that they have termed "flow parsing." According to this theory, the visual system somehow subtracts out the image motion generated by the observer's motion, and whatever image motion remains must be due to moving objects. While they have presented considerable psychophysical evidence consistent with this theory, they have not presented a specific model for how this parsing might be accomplished by neurons. The model presented here does not specifically subtract out the optic flow generated by observer motion, however the comparison required in the second stage of the object identification could be accomplished by a type of subtraction, or inhibition, as described above. Foulkes, Rushton, and Warren (2013b) have demonstrated that human ability to judge object trajectory decays in a manner similar to human ability to judge heading when challenged with flow fields with angular noise and reduced dot density. While we cannot compare our model responses directly with their results, because we are testing object detection and not trajectory, we have shown that our model's accuracy is similarly robust in the face of these manipulations of the flow field. In these cases our model is consistent with the idea of flow parsing.

One difference between the model mechanism presented here and the mechanism suggested for computing object trajectory is suggested by the results of a study by Warren, Rushton, and Foulkes (2012) in which they tested how people perceived object trajectory in the presence of overlapping radial and lateral optic flow fields. With this stimulus, people perceive an illusory shift of the focus of expansion in the direction of the lateral optic flow. The heading computed by our model in these conditions is consistent with that perceived by humans (Royden & Conti, 2003). In contrast, Warren, Rushton, and Foulkes (2012) found that the perception of object trajectory was more consistent with a vector averaging model of computing the radial flow field. They concluded that object trajectory does not depend on a prior estimate of heading, whereas our current model requires an initial estimate of heading. It is difficult to relate our current results to theirs, since they were examining perception of the trajectory of a moving object that is clearly identified by color, whereas we are modeling the detection of a moving object when only velocity information is

available. It is possible that the two tasks have a different dependence on heading computation. In addition, as pointed out above, the object detection performed by our model would likely work equally well if the heading were computed using a simple local vector averaging, consistent with Warren, Rushton, and Foulkes's (2012) results, rather than the motion subtraction used here. The differences would arise only in the presence of rotations.

Royden, Banks, and Crowell (1992) and Royden, Crowell, and Banks (1994) presented evidence that the visual system makes use of extra-retinal signals to compute heading in the presence of eye-movements. While modeling the influence of extra-retinal signals is beyond the scope of this paper, we note that the responses of cells in MSTd are modulated by smooth-pursuit eye-movements (Komatsu & Wurtz, 1988) and this modulation may aid in heading judgments in the presence of eye movements (Bradley et al., 1996). Because the responses of the operators in the second layer of our model resemble those of MSTd cells, this layer would be a good candidate for incorporating the effects of extra-retinal information. Incorporating extra-retinal information into the model is a goal of future research in our lab.

One potential problem for any system that uses speed differences to identify moving objects arises from motion parallax, i.e. the dependence of image speed on the distance of a point from an observer (Rogers & Graham, 1979). An object that has a faster image motion than that of the rest of the scene could be a moving object, or it could simply be closer to the observer. If the distance is large enough, the current model would pick up a stationary object and identify it as moving. This problem could be eliminated if one had depth information about the points in the scene, which could be obtained from stereo information. Indeed, it has been shown that the addition of stereo aids in the perception of heading in the presence of rotations (van den Berg & Brenner, 1994). Rushton, Bradshaw, and Warren (2007) showed that observers detect a moving object more quickly when stereo information is available, and Warren and Rushton (2009) have shown that the addition of depth cues aids in perception of object trajectory. Cutting and Readinger (2002) have shown that, under some conditions, observers appear to use pairwise comparisons of the motions of objects at different depths to identify a moving object. Their analysis requires information about the relative depth of the objects, and thus does not apply to conditions where depth cues are not available, such as those tested here. The use of stereo to disambiguate stationary from moving objects is the subject of ongoing research in our lab. We hope to be able to incorporate biologically-based stereo responses into the model in the future.

Acknowledgments

We thank Laura Webber and Sean Sannicandro for their help running simulations. This work was supported by NSF Grants #IBN-0343825 and #IOS-0818286.

References

Allman, J., Miezin, F., & McGuiness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception, 14*, 105–126.

Bradley, D. C., Maxwell, M., Andersen, R. A., Banks, M. S., & Shenoy, K. V. (1996). Mechanisms of heading perception in primate visual cortex. *Science, 273*, 1544–1547.

Clocksin, W. R. (1980). Perception of surface slant and edge labels from optical flow: A computational approach. *Perception, 9*, 253–269.

Cutting, J. E., & Readinger, W. O. (2002). Perceiving motion while moving: How pairwise nominal invariants make optical flow cohere. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 731–747.

Duffy, C. J., & Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large field stimuli. *Journal of Neurophysiology, 65*, 1329–1345.

Duffy, C. J., & Wurtz, R. H. (1995). Response of Monkey MST neurons to optic flow stimuli with shifted centers of motion. *Journal of Neuroscience, 15*, 5192–5208.

Eifuku, S., & Wurtz, R. H. (1998). Response to motion in extrastriate area MSTl: Center-surround interactions. *Journal of Neurophysiology, 80*, 282–296.

Felleman, D. J., & Kaas, J. H. (1984). Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. *Journal of Neurophysiology, 52*, 488–513.

Foulkes, A. J., Rushton, S. K., & Warren, P. A. (2013a). Heading recovery from optic flow: Comparing performance of humans and computational models. *Frontiers in Behavioral Neuroscience, 7*, 53. http://dx.doi.org/10.3389/fnbeh.2013.00053.

Foulkes, A. J., Rushton, S. K., & Warren, P. A. (2013b). Flow parsing and heading perception show similar dependence on quality and quantity of optic flow. *Frontiers in Behavioral Neuroscience, 7*, 49. http://dx.doi.org/10.3389/fnbeh.2013.00049.

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

Graziano, M. S. A., Andersen, R. A., & Snowden, R. (1994). Tuning of MST neurons to spiral motions. *Journal of Neuroscience, 14*, 54–67.

Hildreth, E. C. (1992). Recovering heading for visually-guided navigation. *Vision Research, 32*, 1177–1192.

Komatsu, H., & Wurtz, R. H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. I. Localization and visual properties of neurons. *Journal of Neurophysiology, 60*, 580–603.

Longuet-Higgins, H. C., & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B, 208*, 385–397.

Maunsell, J. H. R., & van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed and orientation. *Journal of Neurophysiology, 49*, 1127–1147.

Nakayama, K., & Loomis, J. M. (1974). Optical velocity patterns, velocity sensitive neurons, and space perception: A hypothesis. *Perception, 3*, 63–80.

Perrone, J. A., & Stone, L. S. (1994). A model of self-motion estimation within primate extrastriate visual cortex. *Vision Research, 34*, 2917–2938.

Perrone, J. A. (2012). A neural-based code for computing image velocity from small sets of middle temporal (MT/V5) neuron inputs. *Journal of Vision, 12*, 1–31.

Priebe, N. J., & Lisberger, S. G. (2004). Estimating target speed from the population response in visual area MT. *Journal of Neuroscience, 24*, 1907–1916.

Raiguel, S., Van Hulle, M. M., Xiao, D. K., Marcar, V. L., & Orban, G. A. (1995). Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque. *European Journal of Neuroscience, 7*, 2064–2082.

Rieger, J. H., & Lawton, D. T. (1985). Processing differential image motion. *Journal of the Optical Society of America A, 2*, 354–360.

Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8*, 125–132.

Royden, C. S. (1997). Mathematical analysis of motion-opponent mechanisms used in the determination of heading and depth. *Journal of the Optical Society of America A, 14*, 2128–2143.

Royden, C. S. (2002). Computing heading in the presence of moving objects: A model that uses motion-opponent operators. *Vision Research, 42*, 3043–3058.

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

Royden, C. S., & Connors, E. M. (2010). The detection of moving objects by moving observers. *Vision Research, 50*, 1014–1024.

Royden, C. S., & Conti, D. M. (2003). A model using MT-like motion-opponent operators explains an illusory transformation in the optic flow field. *Vision Research, 43*, 2811–2826.

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

Royden, C. S., & Holloway, M. A. (2007) The effect of object speed and angle on the perceived rigidity of an optic flow field. *Journal of Vision, 7*(9), 100, 100a [Abstract] <http://journalofvision.org/7/9/100/>, http://dx.doi.org/10.1167/7.9.100.

Royden, C. S., & Moore, K. D. (2012). Use of speed cues in the detection of moving objects by moving observers. *Vision Research, 59*, 17–24.

Royden, C. S., & Picone, L. J. (2007). A physiologically based model for simultaneous computation of heading and depth in the presence of rotations. *Vision Research, 47*, 3025–3040.

Rushton, S. K., Bradshaw, M. F., & Warren, P. A. (2007). The pop out of scene-relative object movement against retinal motion due to self-movement. *Cognition, 105*, 237–245.

Rushton, S. K., & Warren, P. A. (2005). Moving observers, relative retinal motion and the detection of object movement. *Current Biology, 15*, R542–R543.

Saito, H., Yukie, M., Tanaka, K., Hikosaka, K., Fukada, Y., & Iwai, E. (1986). Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *Journal of Neuroscience, 6*, 145–157.

Tanaka, K., & Saito, H. (1989). Analysis of motion in the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology, 62*, 626–641.

Thompson, W. T., & Pong, T. C. (1990). Detecting moving objects. *International Journal of Computer Vision, 4*, 39–57.

van den Berg, A. V., & Brenner, E. (1994). Why two eyes are better than one for judgements of heading. *Nature, 371*, 700–702.

Warren, P. A., & Rushton, S. K. (2007). Perception of object trajectory: Parsing retinal motion into self and object movement components. *Journal of Vision, 7*, 1–11.

Warren, P. A., & Rushton, S. K. (2008). Evidence for flow-parsing in radial flow displays. *Vision Research, 48*, 655–663.
Warren, P. A., & Rushton, S. K. (2009). Perception of scene-relative object movement: Optic flow parsing and the contribution of monocular depth cues. *Vision Research, 49*, 1406–1419.
Warren, P. A., Rushton, S. K., & Foulkes, A. J. (2012). Does optic flow parsing depend on prior estimation of heading? *Journal of Vision, 12*, 1–14.
Warren, W. H., Blackwell, A. W., Kurtz, K. J., Hatsopoulos, N. G., & Kalish, M. L. (1991). On the sufficiency of the velocity field for the perception of heading. *Biological Cybernetics, 65*, 311–320.
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*, 646–660.
Xiao, D. K., Raiguel, S., Marcar, V., Koenderink, J., & Orban, G. A. (1995). Spatial heterogeneity of inhibitory surrounds in the middle temporal visual area. *Proceedings of the National Academy of Sciences of the United States of America, 92*, 11303–11306.