Detection of moving objects using motion- and stereo-tuned operators

Constance S. Royden
Department of Mathematics and Computer Science,
College of the Holy Cross, Worcester, MA, USA

Sean E. Sannicandro
Department of Mathematics and Computer Science,
College of the Holy Cross, Worcester, MA, USA

Laura M. Webber
Department of Mathematics and Computer Science,
College of the Holy Cross, Worcester, MA, USA

A person moving through the world must be able to identify moving objects in order to interact with them and successfully navigate. While image motion alone is sufficient to identify moving objects under many conditions, there may be some ambiguity as to whether an object is stationary or moving, depending on the object's angle of motion and distance from the observer. Adding a measure of depth from stereo cues can eliminate this ambiguity. Here we show that a model using operators tuned to image motion and stereo disparity can accurately locate moving objects and distinguish between stationary and moving objects in a scene through which an observer is moving.

Introduction

When moving through a scene, we must locate and identify moving objects in order to interact with them in appropriate ways. For example, a driver on a busy street must distinguish between moving and stationary cars to successfully navigate through the scene. The human visual system may use many different cues to identify moving objects, such as the motion (2-D image velocity) on the retina or monocular and binocular cues to depth, but it is not known precisely how the brain processes the information from these cues to determine whether or not an object in the scene is moving. Previously, we have shown that a computational model that computes heading using differences in 2-D retinal image velocities can also identify the borders of moving objects in the scene. It does so by identifying locations where the changes in the speed or angle of image motion across space do not match those expected from the observer motion (Royden & Holloway, 2014). The model uses operators that are tuned to the speed and direction of motion within their receptive fields, similar to the motion tuning of neurons in the primate visual cortex. We showed that one does not have to calculate the original 2-D image velocities from these tuned responses to compute heading and identify the locations of borders of the moving objects with reasonable accuracy.

One problem with the model as presented was that it could not distinguish between a moving object and a stationary object that differed significantly in depth from the surrounding scene. This problem arises because an image-velocity difference at the border of a stationary object can be generated by motion parallax—that is, for an observer moving in a straight line and not moving his or her eyes, the images of nearby objects move faster than those of more distant objects (Rogers & Graham, 1979). If one had access to nonmotion information about the relative depth of objects in the scene, for example from binocular stereopsis, then one could potentially resolve this ambiguity. In this study, we added binocular stereo information to the computation of moving-object location. We show that a model using operators that are tuned to both motion and stereo disparity can locate the border of moving objects and can distinguish between moving and stationary objects.

Theory of heading detection

To understand how a moving observer might identify moving objects, we must examine the image motion generated by the observer's motion through the scene. When an observer moves through a scene, the

Citation: Royden, C. S., Sannicandro, S. E., & Webber, L. M. (2015). Detection of moving objects using motion- and stereo-tuned operators. *Journal of Vision, 15*(8):21, 1–17, doi:10.1167/15.8.21.

doi: 10.1167/15.8.21 Received September 25, 2014; published June 30, 2015

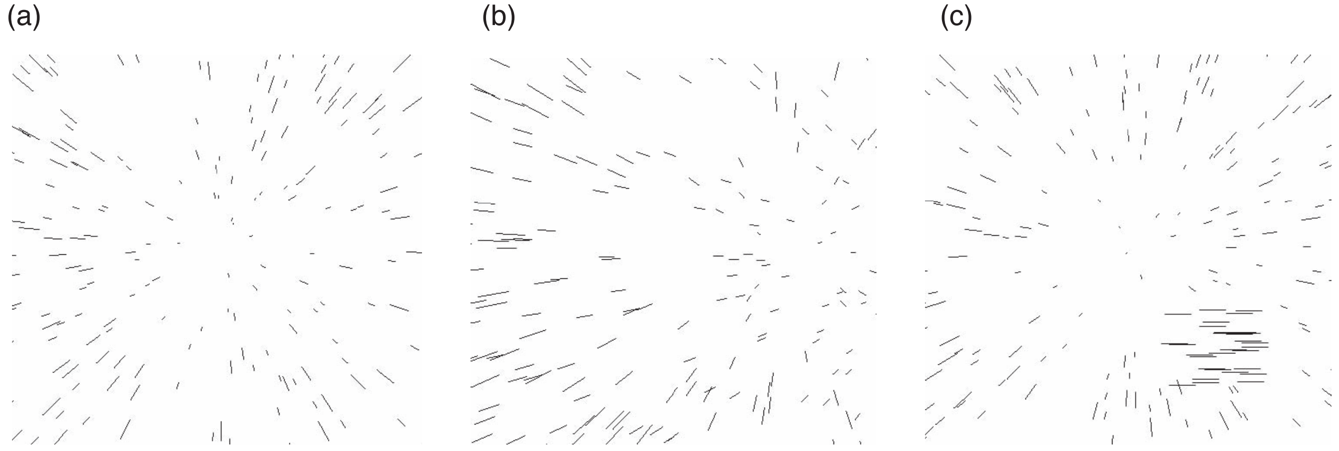

Figure 1. Optic flow fields for an observer moving through a scene. (a) The observer is moving in a straight line toward the center of two planes at different depths. (b) The observer is both rotating and translating toward the same scene as in (a). (c) The observer is translating toward a scene containing a moving object in the lower right corner.

image of each point in the scene moves across the retina. The motion of each point at an instant in time can be described by a velocity vector giving its speed and direction. The pattern of these vectors across the retinal image is known as the optic flow field (Gibson, 1950). If the observer is moving along a straight path, these velocity vectors form a radial pattern, so that lines extending through all of the vectors intersect in a central location known as the focus of expansion (FOE). This can be seen in Figure 1a. The FOE directly corresponds to the observer's heading, or direction of motion through the scene. If the observer is also rotating while moving through the scene, the flow pattern is more complex, which is seen in Figure 1b. Elimination of rotation by local subtraction of neighboring image velocities allows us to regain a radial pattern and therefore compute heading (Longuet-Higgins & Prazdny, 1980). Figure 1c displays a scene with a moving object present and the rotation eliminated. The moving object introduces velocities that disrupt the radial pattern, making it more difficult to compute heading.

It has been shown in psychophysical studies that when the rotation is generated by an eye movement, the visual system may use extraretinal eye-movement signals to eliminate the rotational motion (Royden, Banks, & Crowell, 1992; Royden, Crowell, & Banks, 1994; W. H. Warren & Hannon, 1988, 1990). However, as Royden (1994) pointed out, one still needs to compute both translational and rotational components of observer motion when the observer is moving on a curved path and not moving his or her eyes. It has been shown that even when the eyes are fixed, people who are presented with a scene containing both translation and rotation can judge the translational component of their heading if they are told they are moving along a straight line (Li & Warren, 2004; Royden, Cahill, & Conti, 2006). This means that the visual system is able to eliminate the rotational component of motion when the eyes are fixed, eliminating extraretinal signals. In the current study, we are assuming that the eyes are fixed, so there is no need to incorporate eye-movement signals into the model at this stage.

The velocities of image points are described by two equations representing their horizontal and vertical components, v_x and v_y, respectively. The instantaneous components of motion for the observer can be represented by six components: three translational components (T_X, T_Y, T_Z) along the X-, Y-, and Z-axes and three rotational components (R_X, R_Y, R_Z) for rotations about the X-, Y-, and Z-axes, respectively. A point $P = (X, Y, Z)$ in space projects onto an image plane located 1 unit of distance from the observer at position $p = (x, y)$, where $x = X/Z$ and $y = Y/Z$. Given these parameters, the image velocity for a stationary point in the scene is given by

$$v_x = \frac{xT_Z - T_X}{Z} + xyR_X - (1 + x^2)R_Y + yR_Z$$
$$v_y = \frac{yT_Z - T_Y}{Z} + (1 + y^2)R_X - xyR_Y - xR_Z. \tag{1}$$

These equations can be separated into two components—one that depends only on observer translation through the scene and one that depends only on observer rotation. The rotational components do not depend on depth Z, while the translational components do. Therefore, if one can measure the image velocity at two different depths along a line of sight—i.e., on either side of a depth edge—the equations will differ only in their translational components, not in their rotational components. If one subtracts one of these image velocities from the other, one can eliminate the rotational component (Longuet-Higgins & Prazdny, 1980). We refer to the vector resulting from this subtraction as a difference vector. All of the difference vectors form a radial pattern, in which each vector points towards or away from a central location which corresponds to the observer's direction of motion (Longuet-Higgins & Prazdny, 1980). The equations of these difference vectors are

$$\begin{aligned} v_{xd} &= (-T_X + xT_Z)\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right) \\ v_{yd} &= (-T_Y + yT_Z)\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right), \end{aligned} \tag{2}$$

where Z_1 and Z_2 are the depths of two neighboring surfaces.

The fact that one can use vector subtraction to compute heading using the mathematical abstraction of velocity vectors does not ensure that neurons would be able to use a motion-subtraction mechanism to do the same thing. Because the receptive fields of neurons in the motion-sensitive regions of the primate visual system are spatially extended and are tuned to the speed and direction of the stimuli within their receptive fields, they cannot directly perform a point-to-point vector subtraction. Increasing the spatial extent over which the subtraction takes place introduces noise through both the separation of the locations of the velocities that are being subtracted (Rieger & Lawton, 1985) and the possibility of averaging the velocities of multiple features within the receptive fields. Furthermore, the response of a single neuron cannot indicate the exact 2-D image velocity of the visual stimulus at the location of the receptive field, because the neurons will respond to a variety of speeds and directions of the stimulus, varying the intensity of the response as the speed and direction deviate from the preferred values. So before testing this idea with a computational model, we could not be sure that the subtraction of the tuned response magnitudes of neighboring receptive-field regions would allow us to compute heading. By building a computational model that uses the speed- and direction-tuned responses of operators with spatially extended receptive fields, we have shown that one can compute heading fairly accurately under a large number of conditions (Royden, 1997). We have further shown that this heading computation responds similarly to humans in the presence of moving objects (Royden, 2002), stimuli that generate visual illusions of heading (Royden & Conti, 2003), and added noise (Royden, 1997; Royden & Picone, 2007).

Once one has computed heading, one can theoretically detect moving objects by examining the image for velocities that differ in speed or direction from what is expected given the observer's motion parameters (Thompson & Pong, 1990). We have shown that humans can detect moving objects when their angle (Royden & Connors, 2010) or speed (Royden & Moore, 2012) differs from the expected radial pattern generated by a moving observer. In addition, we have shown that a model using operators tuned to speed and direction of motion, similar to the tuning found in the primate visual cortex, can use speed and direction deviations to identify the borders of moving objects in the scene (Royden & Holloway, 2014).

Although angular deviation can unambiguously identify a moving object, speed deviations can indicate the border of an object but do not distinguish between a moving and a stationary object, due to the speed variations generated from motion parallax. This would be a problem for the Royden and Holloway (2014) model, because it would cause the identification of stationary objects of varying depths as potentially moving. One would need additional information about the relative depth of the surfaces, e.g., from binocular stereopsis, to unambiguously determine whether the edge belongs to an object that is moving relative to the scene. Adding information about the relative depth of objects in the scene would thus increase the robustness of this model for detecting moving objects. The goal of the current project is to model how the brain might use both motion and stereo cues to detect moving objects. Specifically, we extended the model described earlier to include stereo responses that are tuned to stereo disparity. We asked whether a model that uses operators with spatially extended receptive fields tuned to speed and direction of motion and to stereo disparity is capable of distinguishing between moving and stationary objects in a scene through which an observer is moving.

Methods

The computational model

The motion computation of the model has been described in detail elsewhere (Royden, 1997; Royden & Holloway, 2014; Royden & Picone, 2007), so we will describe it briefly here. The model uses operators that have receptive fields that are tuned to speed and direction of motion based on the tuning properties of cells in the primate middle temporal (MT) visual area (Allman, Miezin, & McGuiness, 1985; Maunsell & van Essen, 1983a; Raiguel, Van Hulle, Xiao, Marcar, & Orban, 1995; Xiao, Raiguel, Marcar, Koenderink, & Orban, 1995). The operators have circular receptive fields divided into two halves, where each half is tuned to speed and direction of motion. One half, the classical receptive field, gives a positive or excitatory response to motion, while the other half, the inhibitory surround, gives a negative response. Thus, the response of each operator is given by the response of the excitatory region minus the response of the inhibitory region. In our implementation, we compute the average image velocity for points falling within each region and compute the direction-tuned response by taking the cosine of the angle between this average velocity and the preferred direction of motion of the operator. The response is truncated at ±90°, so responses less than 0 are set to 0. The direction-tuned response is

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

where v_{avg} is the average speed of motion in the receptive field and ϕ is the direction of the average motion. The preferred direction of the operator is given by θ.

The speed-tuned responses are calculated through a Gaussian tuning curve, $e^{-0.5x^2}$, where $x = \log_2(r/R_{\mathrm{pref}})$, r is the direction-tuned speed computed by the operator, and R_{pref} is the preferred speed of the operator. Thus the complete response of each operator is given by

$$R_{\mathrm{op}} = e^{-0.5\left(\log_2\left(\frac{v_{avg+}\cos(\phi_+ - \theta)}{R_{\mathrm{pref}}}\right)\right)^2} - e^{-0.5\left(\log_2\left(\frac{v_{avg-}\cos(\phi_- - \theta)}{R_{\mathrm{pref}}}\right)\right)^2}, \tag{4}$$

where $v_{\mathrm{avg+}}$ and ϕ_+ are the speed and direction, respectively, of the average image velocity in the excitatory region of the receptive field and $v_{\mathrm{avg-}}$ and ϕ_- are the speed and direction of the average image velocity in the inhibitory region of the receptive field (Royden & Holloway, 2014).

In the model, the visual field is divided into separate regions representing the receptive fields of the operators. Each region is processed by a set of operators whose receptive-field responses differ in terms of their preferred direction and speed of motion as well as the orientation of the axis dividing the positive and negative regions of the receptive field (Figure 2). The operator in each region that responds most strongly to a given pattern of image motion projects to a second layer of cells tuned to radial patterns of input from the first layer, with each cell having a different preferred center of expansion. The cell in the second layer that responds most strongly to a given input has a preferred center of expansion that corresponds to the observer's direction of motion, or heading (Royden, 1997). These second-layer cells have response properties similar to those found in the dorsal part of the medial superior temporal visual area in primates (Duffy & Wurtz, 1991, 1995; Saito et al., 1986; Tanaka & Saito, 1989). This model computes heading well for a variety of conditions (Royden, 1997) and shows biases similar to those of humans in the presence of moving objects or with stimuli that generate heading illusions (Royden, 2002; Royden & Conti, 2003).

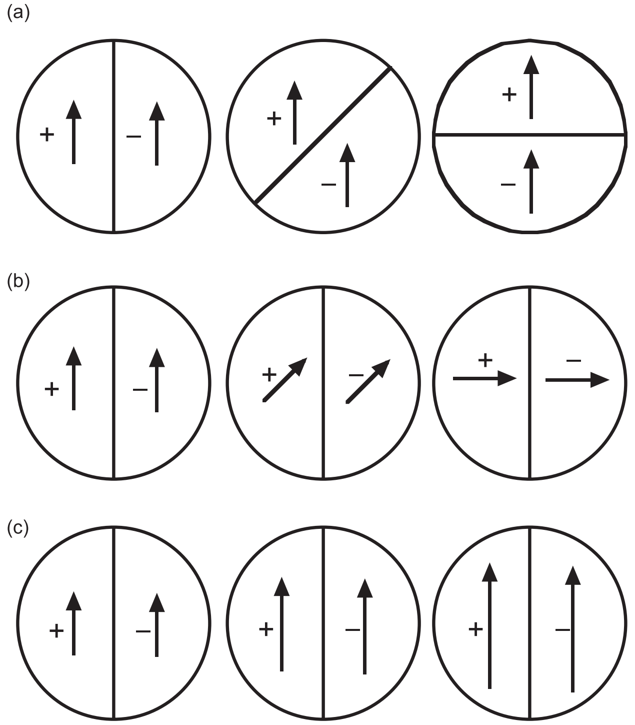

Figure 2. Motion subtraction operators used in the model. Each operator is divided into an excitatory (+) half and an inhibitory (–) half. The operators vary in terms of (a) the angle of the differencing axis, (b) the preferred direction of motion, as shown by the direction of the arrows in the receptive fields, and (c) the preferred speed, as indicated by the length of the arrows.

To compute the presence of moving objects, Royden and Holloway (2014) added a step to the model to detect regions of the visual field in which the speed or direction of image motion differs from what is expected from the computed heading. It identifies regions where the preferred direction of the maximally responding operator in the first layer of cells differs in direction from the radial pattern preferred by the maximally responding second-layer cell. In addition, the model identifies operators for which magnitude of response differs significantly from that of the neighboring operators. This comparison is made after normalization of the magnitude of the responses by dividing the response by the distance of the operator from the computed FOE to eliminate the effect of this distance on the response magnitude (see Equation 2). We showed that this model locates the edges of moving objects well, with little noise, under a variety of object and observer motion conditions, including degraded or noisy flow fields (Royden & Holloway, 2014).

Addition of stereo to the model

To address the problem of the ambiguity of speed responses in detecting moving objects, in the current study we added a stereo computation to compare the estimated change in depth based on motion and stereo measurements. In theory, when the observer moves through a scene, a stationary object will generate a difference in image speed that is proportional to the difference in inverse distance between the object and the background and the distance of the image point from the computed FOE. For an observer moving straight ahead (T_X and $T_Y = 0$), if we normalize the response by dividing by the distance between the image point and the FOE, we obtain

$$v_{\text{diff}} = T_{\text{obs}}\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right), \tag{5}$$

where v_{diff} is the normalized difference in image speeds, T_{obs} is the translational velocity of the observer (T_Z in Equation 2), and Z_1 and Z_2 are the distances from the observer to the object and background, respectively. The responses of the operators in the model already described are proportional to this value.

For stereo disparity, the horizontal angular disparity for a point located at position (X, Y, Z) is given by

$$\delta = 2\tan^{-1}\left(\frac{i}{2Z_{\text{fix}}}\right) - \tan^{-1}\left(\frac{\frac{i}{2} + X}{Z}\right) - \tan^{-1}\left(\frac{\frac{i}{2} - X}{Z}\right), \tag{6}$$

where δ is the angular disparity, i is the interocular distance, and Z_{fix} is the fixation distance, i.e., the distance from the midpoint between the two eyes to the point at which the observer is looking. The disparity difference between two points at positions (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2) is given by

$$\delta = \tan^{-1}\left(\frac{\frac{i}{2} + X_1}{Z_1}\right) + \tan^{-1}\left(\frac{\frac{i}{2} - X_1}{Z_1}\right) - \tan^{-1}\left(\frac{\frac{i}{2} + X_2}{Z_2}\right) - \tan^{-1}\left(\frac{\frac{i}{2} - X_2}{Z_2}\right). \tag{7}$$

If the angles are small, then $\tan(\theta) \approx \theta$, so Equation 7 becomes

$$\delta = i\left(\frac{1}{Z_1} - \frac{1}{Z_2}\right). \tag{8}$$

It is clear from examining Equations 5 and 8 that for small angles there should be an approximately linear relationship between stereo-disparity differences and motion differences generated from motion parallax; however, as the distance between the image points and the fixation point increases, Equation 7 will increasingly deviate from linearity. We therefore performed some computational analyses to determine whether the relationship remains sufficiently linear under the conditions that have been previously tested to justify using a linear approximation in our model. Figure 3 shows a graph of disparity difference versus speed difference for a background at a distance of 1000 cm and an object at distances ranging from 250 to 1000 cm from the observer. These are conditions that have been used for testing observer heading and object detection for a moving observer (Royden & Connors, 2010; Royden & Hildreth, 1996; Royden & Moore, 2012; Royden & Picone, 2007). The graph shows the disparity differences versus speed differences for conditions when the border of the object is positioned 10° to the right of center. The interocular distance was set at 6 cm and the observer speed was given as 200 cm/s. The fixation distance was set at 700 cm. For this fixation distance, the maximum absolute value of the disparity for the given object distances is 0.88°, which is within the range that humans can fuse the images (Grigo & Lappe, 1998). It can be seen that the graph is very close to linear. Given this result, we can conclude that at the border of a stationary object in the scene, defined by a depth discontinuity, the disparity change and the speed change will define a point on or near the line shown in Figure 3. Any border for which the point defined by the disparity difference and the speed difference falls at a distance from this line must therefore belong to a moving object.

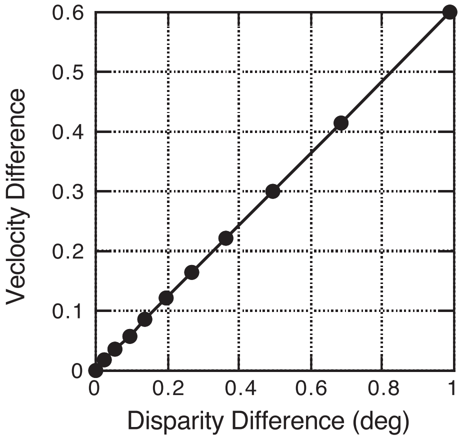

Figure 3. Graph of the calculated disparity differences versus the calculated speed differences at the border of an object in front of a background plane at a distance of 1000 cm from an observer. The calculations were made using Equations 5 and 7. The speed of the observer was 200 cm/s. Points are for object distances between 250 and 1000 cm, spaced every 75 cm.

This analysis is based on exact calculations of the image velocity and disparity of points in the scene. Neurons in early stages of the primate visual system do not compute exact image velocity and disparity, but rather are tuned to motion and disparity. Neurons that are tuned to a particular feature have a preferred value for that feature at which they give a maximum response. The response of the neuron decreases as the value of the feature deviates from that preferred value. Thus direction-tuned neurons will give a maximum response for motion in the preferred direction within the receptive field, and the response decreases as the direction of motion deviates from that preferred direction. This is modeled with a truncated cosine curve already described. In this study, we tested whether a model in which the operator responses were tuned to motion (i.e., speed and direction) and disparity, similar to the responses of neurons in the primate visual system, can distinguish between moving and stationary objects. We ran two sets of simulations. In the first set, we used the motion-tuned responses of the operators in the model described as the motion response. We graphed these responses against the calculated difference in average disparity for all the points lying in the two halves of the maximally responding operator at the border of a moving object. In other words, the response of each side of the operator was given by the average disparity of points in the operator. In the second set of simulations, we used disparity-tuned responses instead of the exact calculated responses. In this case, each operator had a preferred disparity for which it gave a maximum response. As with other tuned responses, the response of the operator decreased as the average disparity within the operator's receptive field deviated from the preferred disparity.

Simulations

The parameters used for the scene and the model were the same as in previous simulations (Royden & Holloway, 2014). All simulations, unless otherwise noted, simulated observer motion toward a background plane consisting of 500 dots located at $Z = 1000$ cm from the observer. The background dots were randomly distributed within a 30° × 30° viewing window. The 6° × 6° objects each contained 50 dots, and unless otherwise noted, the center was located 7° to the right and either 7° down or 7° up from the center of the scene. The objects were opaque, so no background dots overlapped the object's image position. Object distances of 250, 400, 550, 700, and 850 cm from the observer were used to compute the line representing the velocity versus disparity differences for a stationary object. The observer's translational motion was simulated toward the center of the scene at a speed of 200 cm/s. The image position and image velocity of each point in the scene were calculated based on the observer and object motion parameters and the 3-D position of the point. These computed image velocities were used to describe the retinal image that was presented to the model. The responses of the operators in the first layer to this retinal image were modeled on the speed and direction tuning of the cells in MT. We did not model the neural mechanisms used to generate these responses, because in this study we want to know what can be computed from these tuned responses, as opposed to how they are generated. Thus the model starts by computing what an electrophysiologist might measure if he or she were recording from MT cells given the retinal image presented. In this first simulation, the stereo disparity was computed based on the given fixation distance and a distance between the two eyes of 6 cm.

The receptive fields of the operators were circular, 2° 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 visual field was divided into circular regions, spaced every 2°, from −12° to 12° in both the horizontal and vertical directions, giving a 13 × 13 array of regions. Each region was processed by a set of operators which varied in preferred direction of motion, angle of the differencing axis between the excitatory and inhibitory regions, and preferred speed. The 24 preferred directions of motion were spaced every 15° from 0° to 360°; the 16 preferred differencing axes were spaced every 22.5°; and seven preferred speeds were given by 0.5°/s, 1°/s, 2°/s, 4°/s, 8°/s, 16°/s, and 32°/s, which covers the range of speeds in our stimuli. The motion response of each operator was calculated from Equation 4. The maximally responding operator in each region projected to the second layer of cells, consisting of 169 template cells each tuned to a radial pattern. The centers of the patterns for this layer were spaced every 2° from −12° to 12° in both the horizontal and vertical directions. The center position of the maximally responding operator in this second layer was used as the estimate of the observer's heading.

Results

Simulation 1: Tuned motion, calculated disparity

In the first simulation, we sought to verify a linear relationship between the velocity responses of our tuned motion operators and the calculated disparity differences between the two sides of the operators. The maximum disparity response was computed separately from the velocity response. Each region in the visual field was processed by a set of 16 disparity operators, consisting of circular regions 2° in radius, divided into a positive half and negative half by a differencing axis similar to that used in the velocity response calculation. The operators varied in the angle of the differencing

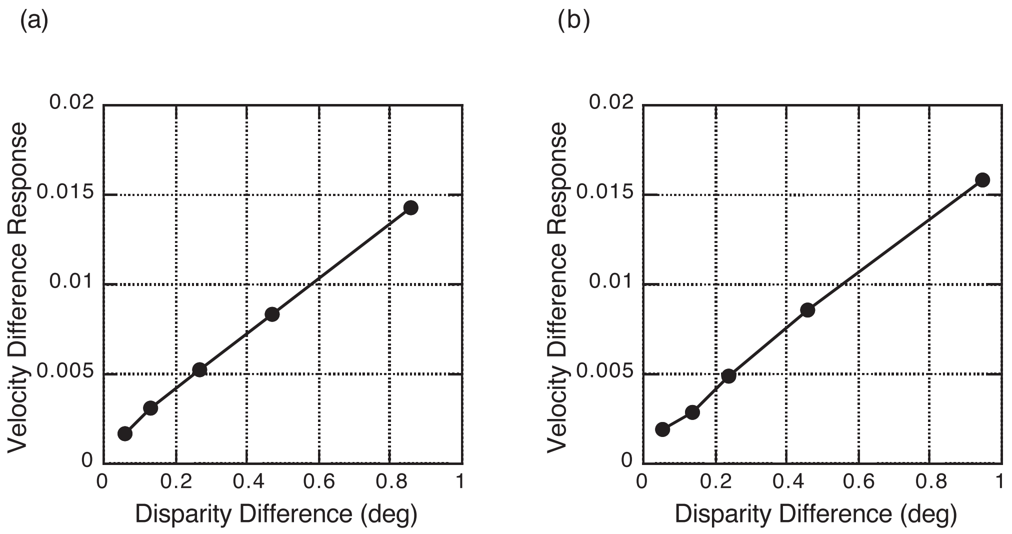

Figure 4. Graphs of calculated disparity differences versus the response magnitude of the tuned motion operators used in the model. The background plane was at 1000 cm. The stationary object distances were at 250, 400, 550, 700, and 850 cm. (a) Graph for a fixation distance of 250 cm. (b) Graph for a fixation distance of 850 cm.

axis, spaced every 22.5° between 0° and 360°. The response of each operator was calculated as the average disparity of points in the excitatory region minus the average disparity of points in the inhibitory region. The response of the operator with the largest response in each region was used to analyze the relationship between the velocity response and the disparity response.

Figure 4 shows the relationship between the average disparity response and the average motion response for the maximally responding operators at the borders of the stationary object. The two graphs show the results for two different fixation distances: 250 and 850 cm from the observer. Note that both graphs are fairly linear. We also ran simulations for fixation distances of 400, 550, and 700 cm and fitted lines to those points as well. The lines all have very good fit ($R^2 > 0.99$), and the equations for the lines are all very similar. The slopes range from 0.0153 to 0.0158 (average = 0.0155), and the intercepts range from 0.00090 to 0.00109 (average = 0.00101). Thus, one can determine an average line that can be used to determine whether or not an object in the scene is moving or stationary:

$$R_V = 0.0155\ R_D + 0.001,$$

where R_V and R_D are the velocity and disparity responses, respectively. If the point described by the disparity and velocity differences falls off this line, then that point should indicate the border of a moving object.

We tested whether the addition of this calculation of stereo disparity could allow the model, described earlier, to distinguish between stationary and moving objects. As before, the model identifies possible moving objects by locating regions where the speed and/or direction of the motion response differs from the radial pattern expected from the computed heading of the observer. As described in our previous work (Royden & Holloway, 2014), we used an angular threshold of 45° and a normalized response threshold of 1.0 to give a robust response to edges where the operator response differed by speed or angle from the expected optic flow field. The normalized responses are computed by dividing the operator response by its distance from the FOE and multiplying by 100. To eliminate noise from operators with little to no response, we applied an absolute motion-response threshold of 0.05. To distinguish between moving objects and stationary objects, we added a module that identifies locations where the velocity difference response and the disparity difference fall more than a small threshold distance from the calculated line. We chose a threshold of 0.002 empirically to eliminate background noise (false positives) while still identifying object borders. We tested a scene containing two objects, a stationary object located at (7°, 7°) and a moving object located at (7°, −7°) from the middle of the optic flow field. The background plane was 1000 cm from the observer, the object was 400 cm from the observer, and the fixation distance was 550 cm from the observer.

Figure 5 shows the results. The black circles indicate locations of operators that signal the border of a moving object. The radius of each circle is proportional to the response magnitude of the operator, which can be used as a measure of confidence. In Figure 5a and b, the object is moving laterally at 52.6 cm/s (7.5°/sec). Figure 5a shows the response of the motion model without the disparity module added. Figure 5b shows the response when the disparity module is added. As can be seen, when there is no test for disparity, the

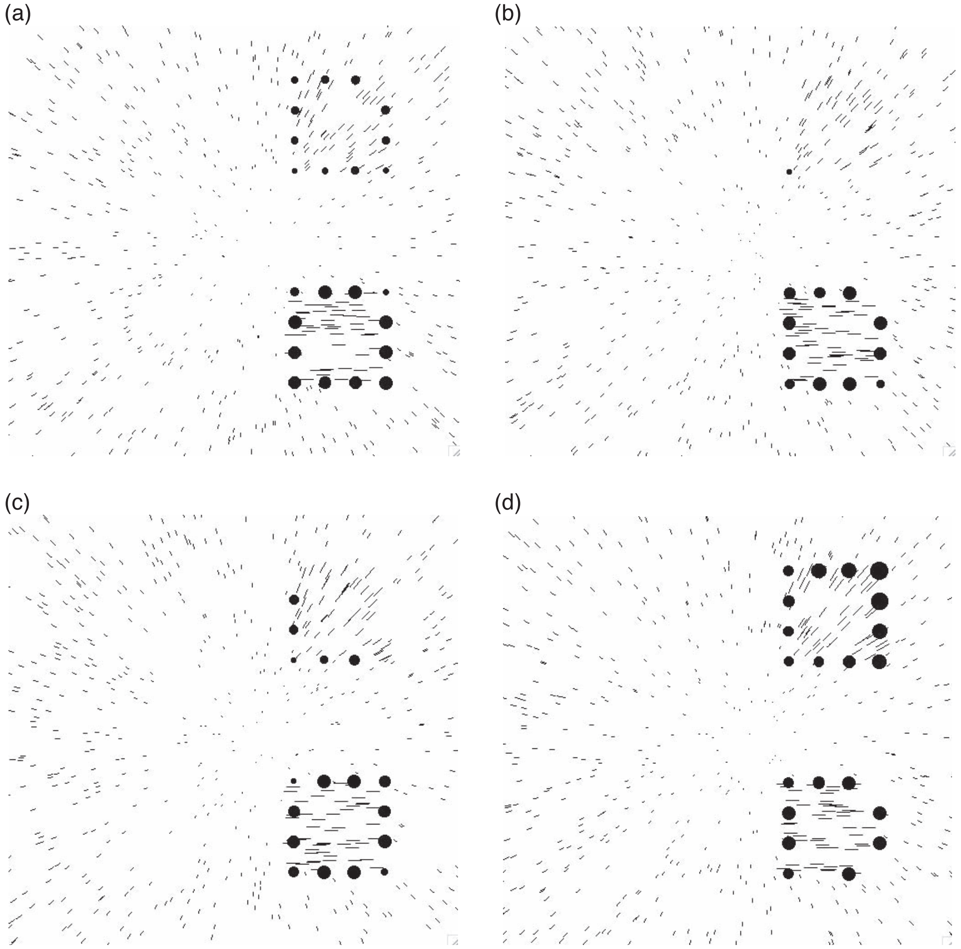

Figure 5. Flow field for an observer moving toward a scene consisting of a background plane, a stationary object, and a moving object. The moving object is in the lower right quadrant, and the stationary object is in the upper right quadrant. Operators that detect a moving object are indicated by a black filled circle at the location of the operator's receptive field. The radius of the circle is proportional to the operator's response magnitude. (a) Model response without the check for disparity. (b) Model response with the check for disparity. (c) Model response with the upper object moving along optic flow lines with speed factor of 1.2. (d) Same as (c) except the speed factor is 1.5.

model identifies both the stationary and moving objects as potentially moving. When the test for disparity is added, most of the operators on the borders of the moving object correctly signal a moving object, while nearly all of those on the borders of the stationary object do not. Figure 5c and d shows that the object is still identified if the angles of the velocity vectors within the object are consistent with the radial flow field but the speeds are not. In these simulations, the distance of the top object was kept at 400 cm, but the image speed was multiplied by a factor of 1.2 and 1.5 in the two different simulations, thus keeping the direction of the image velocities consistent with the radial flow field. These higher image speeds are inconsistent with a stationary object in the scene at that distance, and so the object is now moving relative to the background. It can be seen that some of the edges are identified for a factor of 1.2, and all of the edges are identified for a factor of 1.5, consistent with results of Royden and Connors (2010).

Simulation 2: Operators tuned to disparity

While the previous simulations demonstrated that one can use disparity differences to distinguish moving from stationary objects, the model does so by calculating the disparity of each point exactly and using that calculation to determine the average disparity differences. Cells sensitive to stereo disparity have been described in several regions in the primate visual cortex, including V1, V2, V3, V3a, V4, MT, and MST (Cumming & DeAngelis, 2001). Generally these cells are tuned to disparity and thus do not signal the exact disparity by their responses. Cells in MT have responses to disparity that fall into four categories as

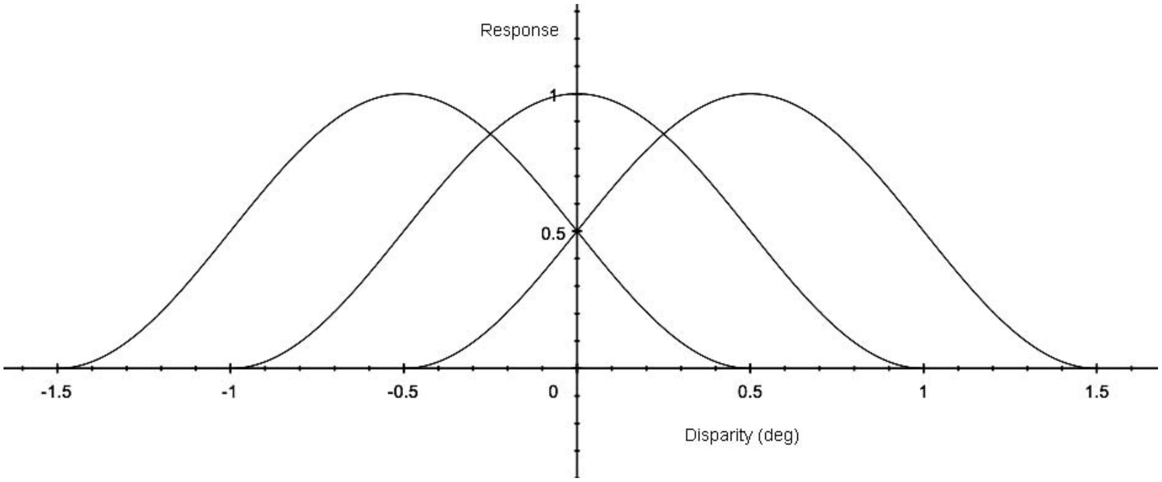

Figure 6. Tuning curves used for the disparity operators. The horizontal axis shows disparity in degrees. The vertical axis shows the tuned response of the operator. Curves are shown for the three operators used in the model.

described by Maunsell and van Essen (1983b): tuned excitatory, tuned inhibitory, near, and far. The tuned excitatory and tuned inhibitory cells have a peak response (either excitatory or inhibitory, respectively) near zero disparity, while the near and far cells have broad tuning for negative or positive disparities, respectively. More recently, cells with narrower tuning to near or far disparities, known as tuned near and tuned far, have also been described (DeAngelis & Uka, 2003). We modeled the tuning curves for the tuned excitatory, near, and far cells using the following function:

$$R = \cos^2\left(\frac{\pi}{2}(\partial + \phi)\right), \tag{9}$$

where R is the response of the operator, ∂ is the average disparity (in degrees) of points within the receptive field, and ϕ is the shift of the tuning curve, given as –0.5°, 0°, and 0.5° for the near, tuned excitatory, and far cells, respectively. The tuning curves are shown in Figure 6.

As with the velocity differences, we computed differences in disparity using operators with an excitatory center and an inhibitory surround, similar to the disparity sensitivity of the centers and surrounds of cells in MT (Bradley & Andersen, 1998). We used a set of operators with circular receptive fields, divided in half into an excitatory and an inhibitory region. Each operator computed a disparity difference by using Equation 9 to compute the response to the average disparity in the excitatory half of the operator's receptive field and subtracting the response to the average disparity in the inhibitory half. The maximum disparity difference response was computed for each location of the visual field by comparing the responses of operators with varying differencing axes and disparity tuning. As with the velocity differences, we used 16 differencing axes, spaced every 22.5° between 0° and 360°. We used the three disparity tuning curves already described, for a total of 48 operators analyzing disparity at each location. If the velocity operator at a given location signaled a possible moving-object border, the response magnitude of the maximally responding disparity operator was then compared to the response magnitude of the maximally responding velocity operator. If the point described by these values did not fall within a threshold distance of the expected linear relationship between the two, the location was identified as a moving-object border. Otherwise, it was assumed to belong to a stationary object that was at a different distance from the observer than was the background.

To determine the expected linear equation for velocity responses versus disparity responses for stationary objects, we repeated the same process of comparing the average disparity difference responses on the borders of a stationary object to the velocity difference responses. We tested objects at distances of 250, 400, 550, 700, and 850 cm at a heading of (0, 0) and a fixation distance of 400 cm. The relationship was close to linear, as shown in Figure 7a. Fitting a line to the points gives an R^2 value of 0.992. We repeated this process for fixation distances of 250, 400, 550, 700, and 850 cm, all of which gave similar linear fits with $R^2 >$ 0.92. Therefore, we surmised that we could use the average linear relationship of these points to distinguish between moving and stationary objects. We averaged the slopes and intercepts of these lines for the different fixation distances and found an average line of $y = -0.0006 + 0.011x$, where y is the velocity response and x is the disparity response. As in Simulation 1, we added a test in the model to determine whether or not the relationship of the disparity and velocity differences for the object being considered was within 0.01 of this line.

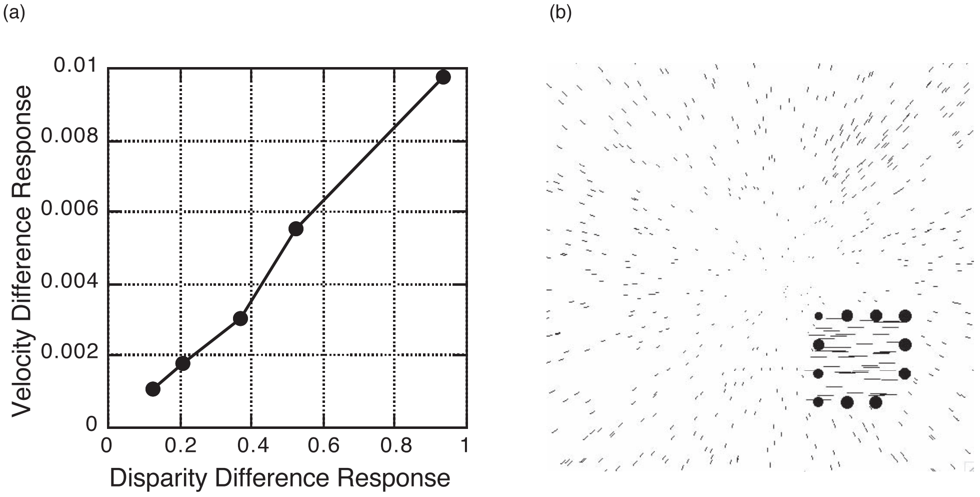

Figure 7. (a) Graph of motion-tuned responses versus disparity-tuned responses at the border of a stationary object in front of a background plane. The observer motion was 200 cm/s toward the center of the scene, and the fixation distance was set at 400 cm from the observer. The distance of the stationary object was 250, 400, 550, 700, or 850 cm from the observer. (b) Flow field showing the operator responses for a scene with a background plane at 1000 cm and stationary and moving objects at 550 cm from the observer. The stationary object is in the upper right quadrant of the flow field, and the moving object is in the lower right quadrant. The fixation distance was 550 cm from the observer. Operators that detect a moving object are indicated by a black filled circle at the location of the operator's receptive field. The radius of the circle is proportional to the operator's response magnitude.

If it was, then the object was identified as stationary, and if it was not, then the object was identified as moving.

To determine how well this version of the model could distinguish stationary from moving objects, a stationary object was placed in the top right quadrant of the scene, centered at (7°, 7°), and a moving object was placed in the lower right quadrant of the scene, centered at (7°, –7°). As in Simulation 1, there were 500 background dots and 50 dots for each object present in the scene. The image velocity of the object was 10°/sec for each simulation. Thus, the simulated speed of the moving object in the world was proportional to its distance from the observer. Figure 7b shows the results of running the program at a fixation distance of 550 cm and an object distance of 550 cm. As in Simulation 1, each circle indicates the location where an operator signals the border of a moving object, with the radius of the circle proportional to the magnitude of the operator's velocity response. All but one of the operators located on the edge of the moving object indicate the presence of a moving object. None of the operators on the edges of the stationary object indicate a moving object. Thus, in this condition, the model performs very well.

To quantify the results of the model simulations, we averaged data from five trials for each condition. We averaged the total number of operators on the border of the moving object, on the border of the stationary object, and in the background that indicated a moving object. The model should be identifying operators on the edge of the moving object but not on the stationary object or the background. There were a total of 169 operator positions in the scene: 12 on the edges of the moving object, 12 on the edges of the stationary object, and the remaining 145 in the background. We found these averages for object distances of 250, 550, and 850 cm at fixation distances of 250, 550, and 850 cm. For the stationary object at a distance of 250 cm, there were always on average fewer than 0.600 out of 12 possible operators that responded incorrectly as moving. For the 550- and 850-cm object distances, there were never any operators responding on the stationary object. None of the background operators incorrectly indicated a moving object for any condition. For the moving object, a possible total of 12 responding operators would correctly identify all the edges of the moving object. Figure 8a shows the average number of operators correctly identifying a moving-object border for the different object and fixation distances and using a distance threshold of 0.01 from the line computed previously. For the object distances of 550 and 850 cm from the observer (i.e., closer to the background plane), each condition had greater than 8.6 (72%) operators responding on the border of the moving object on average over the five trials. Over all the trials for these conditions, the model detected an average of 80% of the edge positions belonging to the moving object. Thus, using disparity-tuned operators is sufficient to distinguish moving from stationary items in the scene for these conditions.

When the object was closer to the observer, i.e., further from the background plane, at a distance of 250 cm, fewer operators detected the borders of the moving

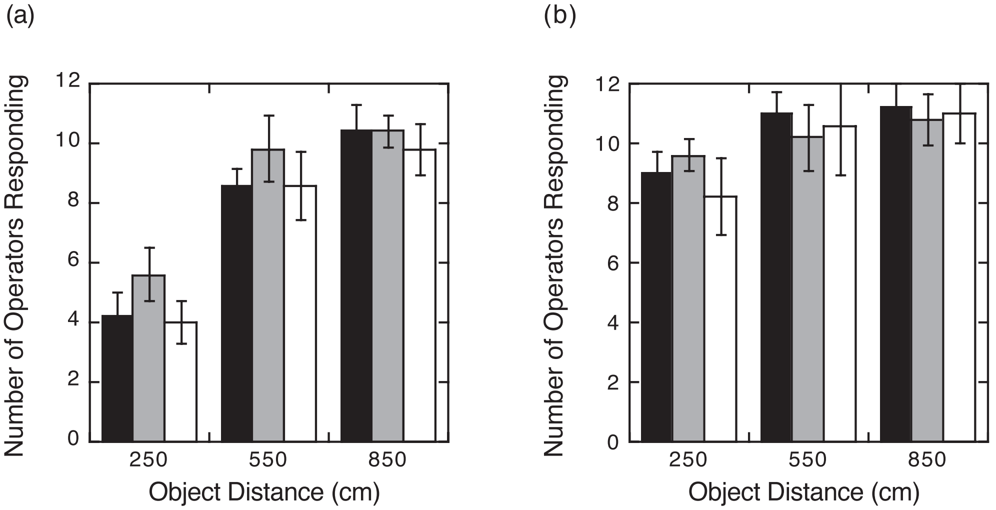

Figure 8. Number of operators correctly detecting moving objects. The vertical axis gives the number of operators on the border of the moving object that accurately detected the moving object, averaged over five trials. The horizontal axis indicates the object's distance from the observer. Black, gray, and white bars indicate the results for fixation distances of 250, 550, and 850 cm from the observer, respectively. (a) Results for a threshold of 0.01. (b) Results for a threshold of 0.005. Error bars show ±1 standard deviation.

object. On average across the three fixation distances, only 4.6 (38%) of the operators detected the moving object. We can increase this number by lowering the threshold. Figure 8b shows the number of operators on the border of the moving object responding when the threshold is lowered to 0.005. For this threshold, the number of operators responding on the border of the moving object averaged 8.9 (74%), 10.6 (88%), and 11 (92%), respectively, for the object distances of 250, 550 and 850 cm from the observer. For the object distances of 550 and 850 cm, as with the higher threshold, none of the operators on the border of the stationary object or in the background responded. Unfortunately, for the object distance of 250 cm from the observer, on average 9.5 (79%) of the operators on the border of the stationary object incorrectly indicated a moving object. It seems likely that the larger disparity and velocity differences for this object distance introduce a larger variability in the disparity and velocity responses, and thus a larger threshold (e.g., 0.01) is required to eliminate false positives. This larger threshold also eliminates some of the responses that correctly indicate a moving object (false negatives). It would be interesting to test human observers to see whether they have more difficulty distinguishing moving from stationary objects when the relative distance between the object and background is large, leading to large disparity and velocity differences between the two. One should note that with the higher threshold value there are almost no false positives, even in the condition with the object position at 250 cm from the observer. With between four and five operators responding along the borders of this object, one might expect that this would be enough to identify the location of the object. It would be easy for the human visual system to extrapolate this cluster of responding operators along the borders of the object to mark the entire thing as moving. Therefore, even in this case, the model responses would be very helpful in identifying moving objects.

Simulation 3: Operators with different sizes

One way in which our model differs from neurons in the primate visual cortex is that all of the operators have the same size receptive fields. A reviewer pointed out that in MT, the receptive-field size increases with the preferred speed of the operator, and that this is an unavoidable aspect of these neurons that arises from the circuitry that leads to speed tuning. We therefore tested our model with the sizes of the receptive fields proportional to the preferred speed of the operator. All parameters were kept the same as for the previous simulation, using the threshold of 0.01. We varied the sizes of the operators, with radii of 0.1°, 0.2°, 0.4°, 0.8°, 1.6°, 3.2°, and 6.4° for the preferred speeds of 0.5°/s, 1.0°/s, 2.0°/s, 4.0°/s, 8.0°/s, 16.0°/s, and 32.0°/s, respectively. The spacing of the centers of the operators' receptive fields was kept the same as in the previous simulation, regardless of the size of the operator. Thus the maximum response for a set of operators was computed for all the operators that had the same center position.

Figure 9a shows the operator responses for a fixation distance of 550 cm and an object distance of 550 cm. In this particular trial, nine out of the 12 operators on the boundary of the moving object and three of the four

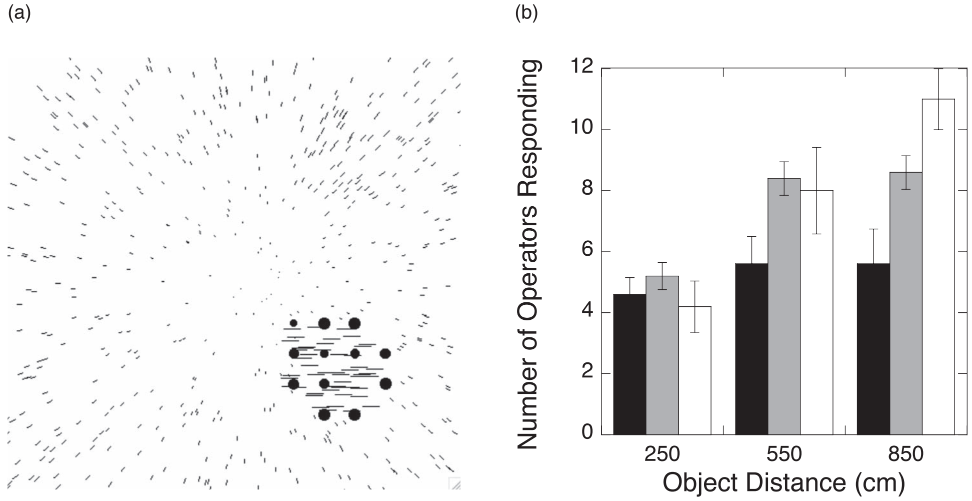

Figure 9. (a) Flow field showing the operator responses for a scene with a background plane at 1000 cm and stationary and moving objects at 550 cm from the observer for the simulation using varied operator sizes. The stationary object is in the upper right quadrant of the flow field, and the moving object is in the lower right quadrant. The fixation distance was 550 cm from the observer. Operators that detect a moving object are indicated by a black filled circle at the location of the operator's receptive field. The radius of the circle is proportional to the operator's response magnitude. (b) Results for condition using varied operator size. The vertical axis gives the number of operators on the border of the moving object that accurately detected the moving object, averaged over five trials. The horizontal axis indicates the object's distance from the observer. Black, gray, and white bars indicate the results for fixation distances of 250, 550, and 850 cm from the observer, respectively. Error bars show ±1 standard deviation.

operators internal to the object indicated the presence of a moving object, while none of the other operators indicated a moving object. In general, the model performed reasonably well with this configuration, although the number of operators responding at the borders of the object decreased somewhat for some of the conditions. As with the previous simulations, for the object distance of 250 cm there were always fewer than 0.6 operators, on average, responding to the edge of the stationary object. For the object distances of 550 and 850 cm, there were none. On average, there was never more than one operator responding in the background. Frequently, this background operator that responded was immediately next to the border of the moving object. The numbers of operators on the border of the object are graphed in Figure 9b. For an object distance of 250 cm, on average 4.6 (38%), 5.2 (43%), and 4.2 (35%) operators on the border of the moving object responded for the fixation distances of 250, 550, and 850 cm, respectively. This is similar to the results of the original simulation. For the object distance of 550 cm, on average 5.6 (47%), 8.4 (70%), and 8.0 (67%) operators responded on the border of the moving object for the three fixation distances. For the object distance of 850 cm, the average number of operators responding on the border of the moving object was 5.6 (47%), 8.6 (72%), and 11.0 (92%) for the three fixation distances. For a fixation distance of 850 cm (white bars), the results were similar to those of the simulation using constant receptive-field size. For the fixation distance of 550 cm (gray bars), the number of border operators responding decreased a small amount for the object distances of 550 and 850 cm compared to the previous simulation, but always by less than two operator responses on average. For the fixation distance of 250 cm (black bars), there was a larger decrease in the number of operators responding for the 550- and 850-cm object distances when compared to the previous simulation, but in both these cases nearly half of the operators on the border of the moving object responded, on average. As discussed earlier, this is a large enough number of operators to identify the moving object, given that the number of operators responding to the background or the stationary object was, on average, less than one. Furthermore, in this condition with the varied receptive-field sizes, there were often additional operators located on the interior of the moving object that responded. This is likely due to the increased receptive-field sizes for the larger speeds, which would make localization of an edge less precise. These added responses would aid in locating the position of the moving object, so in this way this condition may be an improvement over the condition that uses a single receptive-field size.

Discussion

We have shown that a model using speed- and direction-tuned motion operators and disparity-tuned stereo operators can accurately identify the borders of moving objects without also misidentifying stationary objects at different distances from the background. This is an important result, because it was not clear that the responses of speed- and direction-tuned motion operators and disparity-tuned stereo operators would relate to one another linearly. Our analysis shows that under many conditions in which people are moving through a scene, the relationship between the tuned motion and disparity operators is approximately linear; therefore, one can use this fact to distinguish stationary from moving objects. The results of our simulations show that a model using tuned speed and disparity responses can use this relationship effectively to distinguish between stationary and moving objects. We emphasize that the model does not calculate exact image velocity or disparity at any stage following the tuned responses given by the first-layer operators. This simplifies the calculations that must be done to identify the moving objects, as the calculation of exact velocity and disparity would require considerable extra processing (see Priebe & Lisberger, 2004; Qian, 1994). The computation of 2-D image velocity occurs at a later stage in visual processing than MT (Priebe & Lisberger, 2004), and could be useful for other computations such as estimating the time to contact of an object in the scene. However, for the purpose of judging heading and identifying moving objects in the scene, this extra processing to compute 2-D velocities would cause the moving-object detection to be slower. Identifying moving objects quickly would be important for an animal moving through the world, so the fact that it can be done directly from the tuned responses is an important result.

Relationship to physiology

The tuning properties of the operators in our model are based on the motion and disparity tuning properties of cells in MT of the primate visual cortex for high-contrast stimuli similar to the ones that have been used in psychophysical experiments. These cells are tuned for direction and speed (Maunsell & van Essen, 1983a) as well as for disparity (Maunsell & van Essen, 1983b), and we have used the shape of these tuning curves, given a particular retinal image velocity, as a model for the tuning curves used in our model. The motion subtraction is based on the description of the inhibitory surround for motion in MT-cell receptive fields (Allman et al., 1985; Raiguel et al., 1995; Xiao et al., 1995). MT cells also show a surround inhibition of the response to disparity (Bradley & Andersen, 1998), which could perform the disparity differencing implemented in this model. The current model uses two different populations of cells to detect the moving object. One population is tuned to motion and the other is tuned to stereo disparity. While many (34%) of the MT cells have surrounds tuned to both motion and stereo, there are a sizable number (34%) for which the surround effects are driven only by motion, and some (9%) in which the surround effects are driven only by disparity (Bradley & Andersen, 1998). The latter two populations of cells could be used in the computation modeled here.

We have not attempted to model the responses of MT cells in detail in terms of their tuned responses to motion and stereo disparity, as our goal was to show that the identification of moving objects was possible using the motion- and stereo-tuned responses, without calculating the exact image velocities or disparities. For example, MT-cell responses vary at low contrasts, decreasing their firing rate, showing a small shift in preferred speeds, and showing a decrease in surround suppression (Krekelberg, van Wezel, & Albright, 2006; Pack, Hunter, & Born, 2005; Priebe & Lisberger, 2004). But these contrast effects are small for contrasts above 20% (Krekelberg et al., 2006) and thus should not affect the results for high-contrast stimuli. In all of the simulations here, we are assuming high-contrast stimuli similar to those used in psychophysical studies so that we can compare the model results to those obtained with human subjects. It is not known how well people can judge heading or detect moving objects in conditions of low contrast. Consequently, modeling these effects would not be informative, because the results cannot be compared to human abilities, and therefore we have not incorporated these contrast effects in our model. Furthermore, if contrast decreases uniformly across the visual field, such as might occur with dim lighting, the pattern of firing across the population of MT cells remains the same (Priebe & Lisberger, 2004). Because our results depend on relative motion and disparities, the model should still work in the case of a uniform lighting change. The only modification necessary would be a simple normalization of dividing each response by the average population response, as suggested by Priebe and Lisberger (2004). Incorporating more detailed responses similar to those of cells in MT will be the subject of future research.

The important result here is that the subtraction of tuned motion and stereo responses can be used directly to compute heading and unambiguously detect moving objects. This may be accomplished in MT, or there may be other mechanisms. For example, the motion subtraction could be accomplished by V1 cells or by presynaptic inhibition between neural fibers projecting from MT to the medial superior temporal area, as we have previously suggested (Royden, 2004; Royden & Picone, 2007). The result here could apply to these other mechanisms as well, with a different configuration of the model, and thus is much more general than would be possible if we had modeled the detailed properties of MT cells.

We have also not tried to model how the motion responses and the stereo responses are compared physiologically. However, given the small value of the intercept in our computed lines, it seems likely that one could simply divide the stereo response by the motion response to get a constant factor. This could be accomplished electrophysiologically with a divisive inhibitory mechanism. A threshold could be used to identify inputs for which the divided responses are significantly above this constant factor. For example, the populations of cells in MT computing motion and stereo differences could project to cells in the lateral portion of the medial superior temporal area, which may be involved in segmentation of objects within a scene (Eifuku & Wurtz, 1998). These cells are known to be affected by both motion and stereo disparity input, and some of the cells may be modulated by relative disparity differences between the center and surround regions of their receptive fields (Eifuku & Wurtz, 1999). We suggest that these cells not only act to segment the scene into separate objects based on motion and stereo disparity differences, but also may act to distinguish between moving and stationary objects in the scene.

Limitations of the model

In order to compare the performance of the model with results from psychophysical experiments, the model has only been tested on simulated random-dot fields in which all dots have the same contrast and a relatively constant density across the visual field. It is therefore unclear how well it would perform given a variety of different real scenes. As we have pointed out in previous publications (see Royden & Picone, 2007), it is clear that there are some conditions for which the performance of the model will deteriorate. First, the model requires texture in the scene, particularly on either side of the borders of objects, to compute the motion responses. In real scenes there is usually texture on either side of an edge, and it seems likely that human ability to judge heading and detect moving objects would deteriorate somewhat in scenes where there are large regions with no texture. The same argument applies for regions of the scene that contain long, extended edges with no texture nearby. This leads to errors in measuring the motion at the edge due to the aperture problem, which states that one can only measure the perpendicular component of motion at an edge. Again, most real scenes have considerable texture on either side of edges, and because of the integration of information in the second layer of our model, small regions with extended edges should not cause the model performance to deteriorate substantially. It seems likely that humans would have difficulty judging heading given stimuli consisting of large regions with extended edges in only one direction, but this has not been tested. To fully examine these ideas, one needs to perform psychophysical studies on humans to determine how well they can judge heading and detect objects under these nonideal conditions and compare that to model responses. This is beyond the scope of the current paper but would be an interesting topic to pursue in the future. For a more detailed discussion of the limitations of the motion processing in this model, see Royden and Picone (2007).

Another requirement of the model as it is currently implemented is that the scene must contain local variations in the 2-D image velocities, which are necessary for the motion-subtraction stage. There is evidence to suggest that global mechanisms also play a role both in heading computation and in moving-object detection. Duijnhouwer, Beintema, van den Berg, and van Wezel (2006) examined an optic flow illusion that occurs when a radial flow field transparently overlaps a laminar flow field. In this case there is a shift in the perceived center of the radial flow field. Royden and Conti (2003) showed that this shift can be explained by a model using local motion subtraction as described here. Duijnhouwer et al. (2006) showed that when the radial and laminar fields do not overlap and are spatially separated by a 15° gap, there is still a perceptual shift in the perceived center, although the size of the effect is greatly reduced (17% of the effect seen when the fields overlap). This indicates there is a role played by global mechanisms in the computation of heading. Royden and Connors (2010) also showed that global mechanisms play a role in the detection of moving objects based on the angular deviations of the object motion from the radial flow field. The second layer of operators in the model uses large template cells that potentially could account for some of these global effects, or there could be some other mechanism that accounts for these effects. It seems likely that the visual system uses both local and global mechanisms when computing heading and detecting moving objects.

One test of a computational model is to determine whether it fails in the same way that humans fail under a variety of conditions. The motion processing of the current model has been tested under many nonideal conditions, e.g., noisy flow fields, degraded or sparse flow fields, nonuniform flow fields, and flow fields that generate visual illusions (Royden, 1997; Royden & Conti, 2003; Royden & Holloway, 2014). In all cases, the model performed at a level consistent with human performance, lending support to the idea that the visual system uses subtraction of tuned responses to compute heading and detect objects. In the current study, the model performed less well when the object was farthest from the background, at 250 cm from the observer, which would lead to the largest speed and disparity differences. One possible explanation for this decline in performance may be that the two measurements no longer fall along the approximately linear portion of the tuning curve, in either the motion or the stereo domain. Relying on these tuning curves suggests that the model may only provide accurate results within a limited domain of motion or stereo differences. If the human visual system uses the mechanism presented here, we would predict that humans would also show limitations in their ability to distinguish stationary and moving objects as the speed and/or disparity differences increase. This would be an interesting prediction to test psychophysically. Further tests of both humans and the model under the conditions outlined could add further support for this mechanism.

Relationship to psychophysics

It is clear from psychophysical experiments that, under many conditions, human observers can identify moving objects in an optic flow field using motion cues alone. Royden and Connors (2010) showed that people can use angular deviation of an object's motion from the optic flow field, and Royden and Moore (2012) showed that if an object is moving significantly faster or slower than the other objects in the scene, people can detect this moving object. A previous version of the current model used motion-tuned operators to identify moving objects in the scene (Royden & Holloway, 2014). This model performed the moving-object detection at a level consistent with the performance of human observers, with accuracy remaining fairly high in the presence of angular noise and sparse flow fields and for a variety of observer headings, observer rotations, object positions, and object speeds (Royden & Holloway, 2014). Thus, the model using motion-tuned operators alone behaves similarly to humans under many conditions.

One condition in which the motion-tuned operators cannot accurately distinguish between a moving object and a stationary object is when the object's image motion is along the radial optic flow lines, so that it is identified only by a difference in image speed. In this condition, it seems likely that the addition of independent depth cues, e.g., stereo, should aid moving-object detection for both the model and people. Royden and Moore (2012) showed that people can detect these as moving objects under certain conditions, but they did not test whether this detection improved with added depth cues. Rushton, Bradshaw, and Warren (2007) showed that a moving object within an optic flow field "pops out" when presented with both motion and stereo cues available, in the sense that the time it takes observers to identify its trajectory direction does not increase with increased numbers of distractor (stationary) objects in the scene. They further showed that the moving object does not pop out when the stereo cue is removed, as the time to determine the trajectory of the object increased as the number of distractors increased in this condition. In addition, they showed that people were better able to judge the trajectory of a moving object within an optic flow field when both motion parallax and stereo cues were present than they were when motion parallax alone was present in the stimulus (P. A. Warren & Rushton, 2009). These results are consistent with the idea that the addition of stereo cues aids in the detection of moving objects. The model presented here shows that a population of cells tuned to stereo disparity can enhance the ability to detect moving objects within an optic flow field generated by a moving observer.

Comparison with other models

To our knowledge, no other models use motion- and stereo-tuned operators to detect moving objects in the visual field of a moving observer. Several other models compute heading based on neural responses, which is the first stage of this model. As noted by Royden and Holloway (2014), it is likely that models that do not perform a motion subtraction, such as those put forward by Perrone and Stone (1994) or Hatsopoulos and Warren (1991), could probably use speed and direction of motion to detect moving objects in the scene. These models would have more difficulty including a comparison of stereo disparity with velocity, because the disparity of a surface will change with the changing fixation distance of the observer. Thus there is no fixed relationship between the disparity and image speed for a given point in the scene. These models would have to include a mechanism to take into account fixation distance when comparing disparity with image speed. In the model presented here, the use of motion differences and disparity differences make the measurement of fixation distance unnecessary.

Conclusions

We have shown that a model using motion- and stereo-tuned operators can determine the edges of moving objects in a scene through which the observer is moving. The model performs better than an earlier version that relies on motion alone to detect moving objects, because it can distinguish between moving and stationary objects within the scene under more conditions than the previous model could. The results

presented here show that for a moving observer, one need not compute the exact image velocity or stereo disparity to detect moving objects in a scene.

Keywords: optic flow, motion, stereo, moving-object detection, modeling

Acknowledgments

This work was supported by NSF grant #IOS-0818286 and by the College of the Holy Cross Alumni/Parent Summer Research Fellowship, Dan Kennedy '68. Please do not cite without permission.

Commercial relationships: none.
Corresponding author: Constance S. Royden.
Email: croyden@cs.holycross.edu.
Address: Department of Mathematics and Computer Science, College of the Holy Cross, Worcester, MA, USA.

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., & Andersen, R. A. (1998). Center-surround antagonism based on disparity in primate area MT. *The Journal of Neuroscience, 18,* 7552–7565.

Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Review of Neuroscience, 24,* 203–238.

DeAngelis, G. C., & Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *Journal of Neurophysiology, 89,* 1094–1111.

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. *The Journal of Neuroscience, 15,* 5192–5208.

Duijnhouwer, J., Beintema, J. A., van den Berg, A. V., & van Wezel, R. J. A. (2006). An illusory transformation of optic flow fields without local motion interactions. *Vision Research, 46,* 439–443.

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

Eifuku, S., & Wurtz, R. H. (1999). Response to motion in extrastriate area MSTl: Disparity sensitivity. *Journal of Neurophysiology, 82,* 2462–2475.

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.

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

Grigo, A., & Lappe, M. (1998). Interaction of stereo vision and optic flow processing revealed by an illusory stimulus. *Vision Research, 38,* 281–290.

Hatsopoulos, N. G., & Warren, W. H. (1991). Visual navigation with a neural network. *Neural Networks, 4,* 303–317.

Krekelberg, B., van Wezel, R. J. A., & Albright, T. D. (2006). Interactions between speed and contrast tuning in the middle temporal area: Implications for the neural code for speed. *The Journal of Neuroscience, 26,* 8988–8998.

Li, L., & Warren, W. H. (2004). Path perception during rotation: Influence of instructions, depth range, and dot density. *Vision Research, 44,* 1879–1889.

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. (1983a). 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.

Maunsell, J. H. R., & van Essen, D. C. (1983b). Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology, 49,* 1148–1167.

Pack, C. C., Hunter, J. N., & Born, R. T. (2005). Contrast dependence of suppressive influences in cortical area MT of alert macaque. *Journal of Neurophysiology, 93,* 1809–1815.

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

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

Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Computation, 6,* 390–404.

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. (1994). Analysis of misperceived observer motion during simulated eye rotations. *Vision Research, 34,* 3215–3222.

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. (2004). Modeling observer and object motion perception. In L. M. Vaina, S. A. Beardsley, & S. K. Rushton (Eds.), *Optic flow and beyond* (pp. 131–153). Dordrecht, the Netherlands: Kluwer.

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

Royden, C. S., Cahill, J. M., & Conti, D. M. (2006). Factors affecting curved vs. straight path heading perception. *Perception & Psychophysics, 68,* 184–193.

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, M. S. (1994). Estimating heading during eye movements. *Vision Research, 34,* 3197–3214.

Royden, C. S., & Hildreth, E. C. (1996). Human heading judgments in the presence of moving objects. *Perception & Psychophysics, 58,* 836–856.

Royden, C. S., & Holloway, M. A. (2014). Detecting moving objects in an optic flow field using direction- and speed-tuned operators. *Vision Research, 98,* 14–25.

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.

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. *The 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.

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, W. H., & Hannon, D. J. (1988). Direction of self-motion is perceived from optical flow. *Nature, 336,* 162–163.

Warren, W. H., & Hannon, D. J. (1990). Eye movements and optical flow. *Journal of the Optical Society of America A, 7,* 160–169.

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 Science, USA, 92,* 11303–11306.