A Bio-Inspired, Computational Model Suggests Velocity Gradients of Optic Flow Locally Encode Ordinal Depth at Surface Borders and Globally They Encode Self-Motion

Florian Raudies
fraudies@bu.edu
Center for Computational Neuroscience and Neural Technology and Center of Excellence for Learning in Education, Science, and Technology, Boston University, Boston, MA 02215 U.S.A.

Stefan Ringbauer
stefan.ringbauer@uni-ulm.de
Institute for Neural Information Processing, University of Ulm, 89069 Ulm, Germany

Heiko Neumann
heiko.neumann@uni-ulm.de
Center of Excellence for Learning in Education, Science, and Technology, Boston University, Boston, MA 02215, U.S.A., and Institute for Neural Information Processing, University of Ulm, 89069 Ulm, Germany

Visual navigation requires the estimation of self-motion as well as the segmentation of objects from the background. We suggest a definition of local velocity gradients to compute types of self-motion, segment objects, and compute local properties of optical flow fields, such as divergence, curl, and shear. Such velocity gradients are computed as velocity differences measured locally tangent and normal to the direction of flow. Then these differences are rotated according to the local direction of flow to achieve independence of that direction. We propose a bio-inspired model for the computation of these velocity gradients for video sequences. Simulation results show that local gradients encode ordinal surface depth, assuming self-motion in a rigid scene or object motions in a nonrigid scene. For translational self-motion velocity, gradients can be used to distinguish between static and moving objects. The information about ordinal surface depth and self-motion can help steering control for visual navigation.

1 Local Derivatives of Optic Flow Are Motivated by Psychophysics and Neurophysiology

Optic flow is defined as the change in the patterns of structured light on the retina (Gibson, 1950). Such changes are generated when an observer
moves, parts of the scene move, or both. Although optic flow is defined as vector field in the 2D image, it preserves some 3D attributes of the scene. For instance, the relative distance of local surface patches is preserved together with most parameters of 3D self-motion. We are interested in the estimation of types of self-motion and relative distances between surfaces.

One possibility to infer distance cues is by using motion parallax, the change of motion direction and speed between a near and a far object. Helmholtz (1925) described the importance of motion parallax for depth perception from optic flow: “But the moment he [the observer] begins to move forward, everything disentangles itself, and immediately he gets an apperception of the material contents of the woods and their relations to each other in space, just as if he were looking at a good stereoscopic view of it” (p. 295). Such discontinuities lead to strong changes in local flow (i.e., flow derivatives), which encode ordinal depth assuming a rigid scene.

Properties of optic flow derivatives and their interaction with other cues have been well studied. Koenderink (1986) and Braddick (1993) suggest humans are sensitive to divergence, curl, and shear in optic flow and that these form a perceptual basis system for the interpretation of optic flow and the estimation of self-motion. Morrone, Burr, and Vaina (1995) provide evidence that humans integrate motion from sectors in the image plane for divergence and curl patterns. Warren and coworkers compared the influence of optic flow against object position cues in steering control for curvilinear paths. For example, Li and Warren (2000) suggest that a curvilinear path is perceived due to the availability of dense motion parallax cues and at least one reference object present in the stimulus. Warren, Bruce, Wendy, Duchon, and Sahuc (2001) refined this view, suggesting that the control law of steering is governed by the linear combination of object position information and optic flow information weighted by the magnitude of the flow.

Aside from evidence from behavioral data, neurophysiology provides data on the cell’s selectivity for velocity gradients. Cells in the middle temporal (MT) area and the dorsal part of medial superior temporal area (MSTd) in macaque monkeys fire when stimulated by patterns of optic flow or flows with gradients. Duffy and Wurtz (1995) link the firing behavior of such cells to steering control based on the cell’s sensitivity to the translational direction of self-motion. Graziano, Andersen, and Snowden (1994) show cells whose firing is characterized in a 2D space with one axis ranging from contraction (CONT) to expansion (EXP) flow patterns and the orthogonal axis ranging from counterclockwise (CCW) to clockwise (CW) rotational flow patterns. Treue and Andersen (1996) show that MT cells are selective to velocity gradients that occur in stretching flows, compression flows, CW/CCW shear flows, or transitions between these flows. Orban and coworkers study the center and surround subfields of a cell’s receptive field. They found that the surround is antagonistic to the center and most often shows an elongation rather than a perfect radially symmetric shape (Raiguel, Van Hulle, Xiao, Marcar, & Orban, 1995) and some cells have the
antagonistic subfield spatially offset from the center subfield (Xiao, Raiguel, Marcar, Koenderink, & Orban, 1995; see Figure 2D). Cells with an elongated surround subfield might detect small, independently moving objects. Cells with only one antagonistic subfield component might compute velocity gradients. Area MST contains cell populations sensitive to speed gradients (Duffy & Wurtz, 1997). The latroventral region of MST (MSTl) shows sensitivity to motion only in the center and a stationary background or motion in the center and motion in the antipREFERRED direction in the surround (Eifuku & Wurtz, 1998). These cells in area MSTl can detect object motions while the background is stationary, or they can detect independently moving objects (IMOs) when the background is moving.

Figure 1A illustrates cell sensitivities in relation to properties of optic flow generated by self-motion. Assume a forward/rightward self-motion. Furthermore, the scene is composed of a stationary background, one close stationary object, and another close IMO. Four different transitions between local flows at the occlusion edges occur in this scenario. First, at the stationary object, image speeds change from fast to slow (see Figure 1A, i). Along the tangential flow direction, this is a local deceleration. A second transition in flow occurs from slow to fast (see Figure 1A, ii), which is a local acceleration in tangent flow direction. A third transition is the formation of a local source pattern of flow (see Figure 1D, iii). The singularity point of this pattern is also called the focus of expansion (FOE) (Gibson, 1950). We call such a pattern an EXP flow pattern. This EXP pattern is not only locally defined but can extend over a large region of the visual field and thus allows a globally consistent interpretation as well. A fourth transition occurs at the edge of the IMO (see Figure 1A, iv). In addition to a speed change, as in the case of a stationary object, a local change in direction occurs. Such a combination of speed and direction change generates spiral flow patterns. Figure 1B provides a scheme for the encoding of velocity gradients with its local and global interpretation. If the velocity gradient points to the right, this globally encodes an EXP flow pattern and locally an acceleration (see Figure 1B, i). If the velocity gradient points upward, this globally encodes a CCW flow pattern and locally a CW shear edge (see Figure 1B, ii). Patterns for CONT and CW flows are analogous (see Figure 1B, iii and iv). A linear superposition of EXP/CONT with CW/CCW flow patterns indicates a spiral inward or outward flow pattern.

With this work on velocity gradients, we follow three objectives. First, we want to show that velocity gradients globally encode the type of self-motion. Second, we want to establish a link between velocity gradients and their local encoding of ordinal depth at surface borders assuming self-motion in a rigid scene. Our third objective is to show that spiral flow patterns occur for IMOs, at boundaries of object motions, or for local regions of flow with high curvature as generated, for example, by curvilinear self-motion. Thus, we provide behavioral arguments for the processing of velocity gradients—that of deciding about self-motion or relative depth order.
Figure 1: A motivation to study derivatives of flow. (A) A scenario that includes a stationary and a moving object. The transition between objects and background combined with the self-motion produces characteristic flow patterns: (i) local deceleration, (ii) local acceleration, (iii) a pattern of expansion flow, and (iv) a pattern of spiral inward motion. (B) The velocity gradient space as a representation of derivatives of flow. The direction of the gradient vector indicates the pattern, globally, and the characteristics of the local flow discontinuity, locally.
We follow a bio-inspired, computational modeling approach. The detection of optic flow from image sequences builds on prior work (Bayerl & Neumann, 2004; Raudies, Mingolla, & Neumann, 2011). In this letter, we develop a mechanism for the detection of velocity gradients from detected optic flow, which builds on the asymmetric cells reported for area MT (Xiao et al., 1995).

Our letter is organized as follows. In section 2, we describe our computational model for velocity gradients. Section 3 shows the simulation results of the model, with a discussion of the main findings. Our conclusion is in section 4. A preliminary version of the computational model developed here has been published as a conference article (Ringbauer, Bayerl, & Neumann, 2007).

2 A Bio-Inspired, Computational Model of Velocity Gradients

Our model captures and describes the local changes in the direction and speed of vectors in optic flow. First, we briefly review the model we use for the computation of optic flow from video sequences (Raudies et al., 2011). Then we describe the mechanism for the computation of velocity gradients.


Velocity vectors are defined by speed and direction and translate into shifts between frame pairs when multiplied by the reciprocal value of the temporal sampling rate. Initial velocities are computed using a correlation mechanism, which uses two successive frames (Hassenstein & Reichardt, 1956). Our model represents velocities as likelihood values in a velocity space with log-polar coordinates. This log-polar encoding is inspired by the firing characteristics of cells in area MT, which can be described as a gaussian that encodes the presented visual motion in log-polar coordinates (Rodman & Albright, 1987).

Initial velocity estimates can be ambiguous, for example, due to the aperture problem (Wallach, 1935). In our model, local estimates are integrated over the spatial and temporal domain to resolve such ambiguity. We propose a three-level cascade and two model areas to integrate initial motions across space and time. Our three-level processing cascade is inspired by the organization of cortex into mainly six layers but is abstract with respect to termination or origin of forward and feedback signals in layers of cortex, as described, for example, in Felleman and Van Essen (1984). But it contains the same patterns of forward, feedback (FB), and lateral (columnar) connectivity and an organization of function into different areas. Each model area is composed of a three-level processing cascade (Bayerl & Neumann, 2004; Bouecke, Tlapale, Kornprobst, & Neumann, 2011; Raudies et al., 2011). The first stage of our cascade spatially integrates signals (Born & Tootell, 1992) and employs a nonlinear transform attributed to synapses and dendritic
tree (Angelucci et al., 2002). Signal integration and nonlinear transform are formalized in the ordinary differential equation (ODE):

\[
\dot{x}^{(1)} = -x^{(1)} + f_{\text{sample}} \left( [x^{FF}]^\alpha \ast \Lambda_{\text{space}} \ast \Lambda_{\text{vel}} \right).
\] (2.1)

The signals \(x^{(1)}\) and \(x^{FF}\) model the membrane potential, which has an almost linear relationship to the firing rate of a neuron (Stafstrom, Schwindt, & Crill, 1984). In our model, this membrane potential represents the likelihood for a visual image motion. In our model, \(x^{(1)}\) exists for each spatial location and represented velocity. For an image resolution of 640 × 480 pixels, 16 motion directions, and 6 motion speeds, this amounts to 640 × 480 × 16 × 6 = 29,491,200 model cells for the encoding of motion likelihood values. For model area V1, the incoming signal is defined by the likelihood values of the correlation detector. This incoming signal is transformed by a quadratic function, \(\alpha = 2\), followed by the convolution (symbol \(*\)) with the gaussian \(\Lambda_{\text{space}}\). The convolution with the gaussian appropriately reduces high frequencies (Nyquist, 1924) before sampling with \(f_{\text{sample}}\). We implement arbitrary sampling rates using a linear interpolation. After sampling, the resulting space is convolved with the gaussian \(\Lambda_{\text{vel}}\). This convolution blurs likelihoods in the velocity space. All parameter values and variables in equations are kept dimensionless, since we do not attempt to model the detailed biophysical processes of a cell or its channels.

The second stage of the three-level cascade integrates the FB signal that originates from areas higher up in the hierarchy than the one considered. In our model, FB provides context information along the top-down stream of processing. Here, disambiguated motion signals are delivered from model area MT to model area V1. This FB helps to enhance likelihood values in the motion encoding of area V1 that correspond to disambiguated likelihoods from area MT. Due to the closed loop processing between model area V1 and MT, the disambiguation of motion signals can propagate in visual space when processing coherent motion over several image frames. In our model, FB modulates the driving signal (Bullier, Hupé, James, & Girard, 2001). A linking principle (Eckhorn, Reitboeck, Arndt, & Dicke, 1990) suggests that FB only enhances the driving signal, but FB alone cannot generate activity. It always requires driving input. Formally, this combination of the FB signal \(x^{FB}\) and the driving signal \(x^{(1)}\) is defined by

\[
\dot{x}^{(2)} = -x^{(2)} + x^{(1)} \cdot (1 + \beta \cdot x^{FB}).
\] (2.2)

\(^{1}\)In our implementation, we solve all ODEs by using their steady-state solution. Iterations between different frames of the input sequence are iterations of these steady-state solutions that are assumed for each frame.
The parameter $\beta$ regulates the strength of the FB signal in relation to the driving forward signal. The modulation property of the FB signal is explained as follows. If $x^{FB}$ is nonzero, it enhances compatible likelihood values in the driving signal $x^{(1)}$. But this happens only if such compatible likelihood exists. In case a driving signal is present that does not receive any FB, then the driver $x^{(1)}$ is left unchanged. Otherwise, if the driving signal $x^{(1)} \approx 0$, the FB has no influence on the integration of signal activity for the second-level stage likelihood $x^{(2)}$.

The third stage of our three-level cascade normalizes likelihood values across features at each location in the visual field. Lateral or columnar connectivity can explain the signal integration within the surround of a receptive field (Stettler, Das, Bennett, & Gilbert, 2002). Our model assumes these lateral interactions to have an inhibitory net effect. This inhibition is implemented as divisive inhibition. The division keeps likelihoods across different encoded motions bounded. The combination of FB and divisive normalization leads to a biased competition, deemphasizing the likelihood for features that did not receive supportive FB. Formally, this third-stage processing mechanism is expressed by

$$
\dot{x}^{(3)} = -\gamma \cdot x^{(3)} + x^{(2)} * \Lambda^+ - x^{(3)} \cdot (x^{(2)} * \Lambda^-).
$$

Equation 2.3 integrates the second-stage signal $x^{(2)}$ in combination with a center-surround mechanism. The excitatory or supportive center likelihood is computed by the convolution of the incoming signal $x^{(2)}$ with the kernel $\Lambda^+$. The inhibitory or suppressive surround likelihood is computed by convolution of $x^{(2)}$ with the kernel $\Lambda^-$. Typical parameterizations assume gaussians for both kernels, whereas $\Lambda^-$ is several standard deviations larger than $\Lambda^+$. All convolutions range over dimensions of motion speed and direction. A listing of parameters for model areas V1 and MT is given in Raudies et al. (2011).

Our model of velocity gradient computation builds on the representation of optic flow in model area MT. Velocity gradients are detected based on MT-encoded likelihoods and are further processed using the same three-level processing cascade as described. Next, we describe the detection scheme and model areas, which processes velocity gradient information.

### 2.2 Bio-Inspired Model Mechanisms for the Detection and Integration of Velocity Gradients.

Our mechanism for the detection of velocity gradients is motivated by the 50% of MT neurons with asymmetric receptive field shapes that have the suppressive subfield only to one side of the excitatory subfield (Xiao, Raiguel, Marcar, & Orban, 1997; Born & Bradley, 2005). Schematically the effect of the excitatory center and inhibitory surround of these asymmetric MT neurons can be formalized by a filter kernel with one
positive and one negative subfield that is offset to one side of the positive subfield. Such a combined filter with a positive and a negative subfield can compute directed spatial derivatives if applied to the spatial domain. If such a filter is applied to the motion direction domain, it computes sensitivity for motion direction differences. Similarly, if the filter is applied to the motion speed domain, it computes sensitivity for motion speed differences. We use all three types of filtering for the detection of velocity gradients: filtering in the spatial, the motion direction, and the motion speed domain.

In our detector model, the asymmetric filter kernels operate on the likelihood values that encode visual motions. (Figure 3 provides a figural description of all mechanisms of the detector.) We define the asymmetric filter kernel as

\[
b^{+/−}_{dir}(φ, \Delta φ, s) = \frac{1}{2 \cdot \pi \cdot s^2} \exp \left( -\frac{1}{2} \cdot \left( \frac{u(φ, \Delta φ)}{s} \right)^2 - \frac{1}{2} \cdot \left( \frac{v(φ, \Delta φ) \mp s}{s} \right)^2 \right),
\]

with its direction composed of \( φ \) and \( \Delta φ \) by using the rotated coordinate system,

\[
\begin{pmatrix}
u(φ, \Delta φ) \\
u(φ, \Delta φ)
\end{pmatrix} = \begin{pmatrix}
\cos(ψ(φ, \Delta φ)) & -\sin(ψ(φ, \Delta φ)) \\
\sin(ψ(φ, \Delta φ)) & \cos(ψ(φ, \Delta φ))
\end{pmatrix} \begin{pmatrix}
x \\
y
\end{pmatrix}.
\]

For velocity gradients tangent to the motion direction, we set \( ψ(φ, \Delta φ) = φ + \pi/2 - \Delta φ \), and for velocity gradients normal to the motion direction, we set \( ψ(φ, \Delta φ) = φ + \pi - \Delta φ \). The direction of these filters is orthogonal to the local motion direction due to the employed encoding of motion directions by a distribution of likelihood values in spatial and motion domains. For instance, for an EXP flow pattern, the local motion direction is radially directed outward from the center. The encoding of motion direction varies orthogonal to the local motion direction. To detect a change in the likelihood values for motion directions in this expansion optic flow, the filter kernel has to be orthogonal to the radial direction or motion direction. Receptive fields of cells in area MT are elongated orthogonal to their preferred motion direction as well (Raiguel et al., 1995). The filters from equation 2.4 depend, furthermore, on the motion direction difference \( \Delta φ \) and motion speed \( s \). A faster speed uses a larger filter kernel. The standard deviation of a gaussian is directly proportional to the speed \( s \), as is the spatial shift of features between two frames. A motion direction difference \( \Delta φ \) changes the angular difference between the encoded motion direction \( φ \) and the probed motion direction \( ψ \). In our example of an EXP flow pattern, this difference is \( \pi/2 \). Our detector applies these asymmetric filters in the spatial domain to
compute likelihood values for a direction difference $\Delta \phi$ and speed $s$ and sums the filtering result over all possible reference directions $\phi$:

$$g^{+/-}_{dir}(x, y, \Delta \phi, s) = \sum_{\phi} m(x, y, \phi, s) * b^{+/-}_{dir}(\phi, \Delta \phi, s). \quad (2.6)$$

This gives the gradient likelihood $g^{+/-}_{dir}$ for directions and two channels, a $+$ channel and a $-$ channel. The summation over filters with different directions can be interpreted as a filtering mechanism in the motion direction domain. The symbol * denotes the filtering in the spatial domain. Thus, equation 2.6 applies filtering in the spatial domain in combination with a filtering in the motion direction domain. These are two of the domains described above where filtering is applied. In equation 2.6, we indicate the likelihood for the excitatory channel (center) by the $+$ symbol and the likelihood for the inhibitory channel (surround) by the $-$ symbol. Next, we formulate a similar mechanism that operates on the motion speed domain. We define the gaussians in the speed domain,

$$b_{spd}(s_i; \sigma_s) = \frac{1}{\sqrt{2\pi} \cdot \sigma_s} \exp \left( -\frac{1}{2} \left( \frac{s - s_i}{\sigma_s} \right)^2 \right) \quad \text{for } i = 1, 2. \quad (2.7)$$

Speeds $s_1$ and $s_2$ define the speed difference $\Delta s = s_2 - s_1$ to be detected. To simplify, we keep the standard deviation $\sigma_s$ of the gaussian constant. These gaussians are used to filter the likelihood values from equation 2.6 that are sensitive to motion direction differences. The filtering is explicitly written using a summation over speeds combined with the function of the speed differences $s - s_i$ from equation 2.7. All of these operations are summarized as

$$g^{+/-}_{dir,s}(x, y, \Delta \phi, s_i) = \sum_{s} g^{+/-}_{dir}(x, y, \Delta \phi, s) \cdot b_{spd}(s_i; \sigma_s) \quad \text{for } i = 1, 2. \quad (2.8)$$

This computes the auxiliary term $g^{+/-}_{dir,s}$, which can be interpreted as the likelihood for directional velocity gradients for a specific speed $s_i$. The index $i = 2$ denotes the faster speed, and the index $i = 1$ denotes the slower speed. Likelihood values for motion direction differences and motion speeds are used to compute the difference of likelihoods that encode a faster and a slower speed:

$$g(x, y, \Delta \phi, \Delta s) = [g^{+}_{dir,s}(x, y, \Delta \phi, s_2) - g^{-}_{dir,s}(x, y, \Delta \phi, s_1)]^+ \quad \text{for } \Delta s = s_2 - s_1. \quad (2.9)$$
The difference is half-wave rectified, which is indicated by the symbol $[\cdot]^+$. Equation 2.9 defines the likelihood of a detected velocity gradient by its motion direction difference $\Delta \phi$ and motion speed difference $\Delta s$. Our detector applies filtering in the spatial domain, motion direction domain, and motion speed domain.

Figure 2 summarizes all computational steps of our detector. Computations start with filters that contain a positive and a negative subfield. These are depicted by circles marked by symbols $-$ and $+$. Our example shows the likelihood encoding from model area MT for an EXP flow pattern. High likelihoods occur, for example, to the right of the image center in the maps encoding a motion direction of zero degree. In addition, high-likelihood values shift from the inside radially outward. This encodes the increase of motion speed that occurs from the center to the periphery. For encodings of increased motion direction, high-likelihood regions shift counterclockwise as going from top to bottom. In Figure 2, filter kernels are oriented perpendicular to the expected transition of high likelihood between neighboring maps. In the example, such transitions occur between $22.5^\circ$, $67.5^\circ$, ..., and $337.5^\circ$. In practice these transitions are gradual; however, for simplicity, we drew sharp boundaries in Figure 2. The filtering results for the excitatory channel ($+$) and inhibitory channel ($-$) are summed over directions. This summation is indicated by the plus signs in each column showing motion likelihood maps. The summation results in likelihood values that we show as two channels depicted by the bottom-most two rows. Likelihood values from different motion speeds are multiplied with the indicated gaussians—the upper one shifted to higher speeds to represent the gaussian for $s_2$ and the lower one shifted to lower speeds to represent the gaussian for $s_1$. After the multiplication with the gaussians, the resulting product values are summed over speeds. This summation is denoted by the plus symbols in rows for the likelihood spaces of the last two rows in Figure 2. The half-wave rectified difference between the summed product values for higher and lower speed defines the likelihood for a velocity gradient. The sampling of combinations of motion direction differences $\Delta \phi$ and speed differences $\Delta s$ computes the likelihood map for velocity gradients.

Model areas MT and MST are defined by the same three-level processing cascade that we described in equations 2.1 to 2.3. For model area MT, the cascade is defined by

$$\dot{x}^{(1)}_{MT} = -x^{(1)}_{MT} + \left[ g \right]^a \ast \Lambda^b_{\text{grad}},$$

$$\dot{x}^{(2)}_{MT} = -x^{(2)}_{MT} + x^{(1)}_{MT} \cdot (1 + \beta \cdot x^{FB}_{MST}), \text{ and}$$

$$\dot{x}^{(3)}_{MT} = -\gamma \cdot x^{(3)}_{MT} + x^{(2)}_{MT} - x^{(3)}_{MT} \cdot \sum_{\phi, s} x^{(2)}_{MT}.$$  

Equation 2.10 integrates the detected gradient likelihoods from equation 2.9. The likelihood values are transformed applying the nonlinearity setting
Figure 2: The filtering schema for the computation of velocity gradients assuming a population encoding of motion velocities. (A) The upper half shows the encoding of velocities in likelihood maps with coordinates $x$ and $y$ for motion directions $\phi$ and motion speeds $s$, as indicated by the coordinate system that is plotted for the case $s = 1$ pixel/frame and $\phi = 0^\circ$. Antagonistic filters are applied to each of the maps with the orientation difference $\Delta \phi + \pi/2$ between the encoded motion directions and filter orientation. This computes the velocity gradient tangent to the motion direction. Filtering results for the positive and negative filter kernel are summed over directions. (B) Sums for each motion speed are multiplied with the gaussians, $b_{s\phi}(s, \sigma)$, and the products are summed over speeds to compute $g_{dir,s}^{+/--}$. (C) We define the rectified difference between the sums as the likelihood for a velocity gradient.
\(\alpha = 2\). The squared likelihood values are blurred by convolving with the
gaussian kernel \(\Lambda_{\text{grad}}^b\) which is parameterized by the standard deviation
\(\sigma_{\Delta \varphi} = 16.875^\circ\) for motion direction differences and the standard deviation
\(\sigma_{\Delta s} = 0.0567\) pixel per frame for motion speed differences. We assume a
maximum speed difference of 1.7 pixels per frame that is equally distributed
to six cells for the computation of the standard deviation \(\sigma_{\Delta s}\). In addition,
we assume cyclic boundary conditions for the domain of motion direction
differences and Neumann boundary conditions for the domain of motion
speed differences. Equation 2.11 integrates the FB signals \(x_{\text{MST}}^{FB}\) from model
area MST. Similar to the FB in motion processing, this FB can only enhance
existing nonzero likelihoods \(x_{\text{MST}}^{(1)}\) and cannot generate likelihood without
driving support. Equation 2.12 normalizes the likelihood values by division
with the sum of all likelihoods for different velocity gradients at one spatial
location. The signal \(x_{\text{MST}}^{(2)}\) from equation 2.12 serves as input to the processing
cascade of model area MST. This cascade consists of
\[
\dot{x}_{\text{MST}}^{(1)} = -x_{\text{MST}}^{(1)} + f^{1:2}\left(\left(x_{\text{MST}}^{(3)} * \Lambda_{\text{space}}^b\right)^\alpha * \Lambda_{\text{grad}}^b\right)
\]
\[
\dot{x}_{\text{MST}}^{(3)} = -\gamma * x_{\text{MST}}^{(3)} + x_{\text{MST}}^{(1)} - x_{\text{MST}}^{(3)} \cdot \sum_{\phi, \beta} x_{\text{MST}}^{(1)}.
\]
The FB equation at the second stage of the cascade, \(x_{\text{MST}}^{(2)}\), is left out because
model area MST does not receive top-down input in the current implemen-
tation. Equations 2.13 and 2.14 have essentially the same function as those
of model area MT (see equations 2.10 and 2.12). Additional functions are the
spatial filtering with the gaussian \(\Lambda_{\text{space}}^b (\sigma_{\text{space}} = 1\) pixel) and the sampling
by 1:2 of model area MT likelihood values before feeding them into the
cascade. Table 1 reports all parameters of the model for the processing of
velocity gradients.

A summary of our model for motion and velocity gradient detection and
integration with its likelihood representation in shown in Figure 3. Process-
ing starts with feeding the images to our model area V1 (see Figure 3A).
This model area V1 detects initial likelihood values for motion velocities
and uses the velocity representation shown in Figure 3B. Each square tile in
the velocity representation contains the likelihood for a velocity. These like-
lihood values are processed by the three-level cascade in V1 and fed into our
model area MT. The model area MT in turn feeds a disambiguated motion
signal, gained by integration over a larger region in the visual field, back to
model area V1. These forward/FB signals are denoted by the two arrows
between boxes V1 and MT in Figure 3A. Likelihood values for velocity gra-
dients are detected based on the likelihood representation of model area MT
by asymmetric filters depicted as circles with a + and -. These likelihoods
for velocity gradients are encoded in the gradient representation shown
in Figure 3C. Each square tile contains a likelihood that encodes a motion
direction difference combined with a motion speed difference. Model area
Table 1: Parameter Values of the Velocity Gradient Model.

<table>
<thead>
<tr>
<th>Description</th>
<th>Identifier and Value</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradient detector</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard deviation for gaussian in speed domain</td>
<td>$\sigma_s = 0.5$ pixel/frame</td>
<td>2.7</td>
</tr>
<tr>
<td>Model area MT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonlinearity</td>
<td>$\alpha_{MT} = 2$</td>
<td>2.10</td>
</tr>
<tr>
<td>Feedback parameter</td>
<td>$\beta_{MT} = 10^1$</td>
<td>2.11</td>
</tr>
<tr>
<td>Normalization parameter</td>
<td>$\gamma_{MT} = 10^{-5}$</td>
<td>2.12</td>
</tr>
<tr>
<td>Model area MST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaussian</td>
<td>$\sigma_{space} = 1$ pixel</td>
<td>2.13</td>
</tr>
<tr>
<td>Sampling rate MT:MST</td>
<td>$r_{MT:MST} = 0.25$</td>
<td>2.13</td>
</tr>
<tr>
<td>Nonlinearity</td>
<td>$\alpha_{MST} = 2$</td>
<td>2.13</td>
</tr>
<tr>
<td>Normalization parameter</td>
<td>$\gamma_{MST} = 10^{-2}$</td>
<td>2.14</td>
</tr>
</tbody>
</table>

Note: Parameter values for the V1/MT motion processing model are reported in Raudies et al. (2011).

MT employs the three-level cascade to likelihood values for velocity gradients, and so does model area MST. The gradient processing between model areas MT and MST is analogous to the motion processing between areas V1 and MT. Our model uses the same forward/FB scheme for gradients as it uses it for motions, depicted by the arrows between the MT and MST boxes in Figure 3A. To interpret and visualize motions and gradients, we read out likelihood values from area MT that encode motions or gradients as well as likelihood values from area MST that encode gradients. These latter likelihood values provide information about the type of self-motion. This model does not employ motion processing within area MST, which is a simplification compared to prior model versions (Raudies et al., 2011). Our model employs no gradient processing at the level of area V1 due to the evidence that cells showing functionality for gradient detection have been reported for MT/MST but not for V1.

2.3 Read-Out of Population-Encoded Motion Velocities and Gradients. We use a weighted-vector average to read out the population codes depicted in Figures 3B and 3C. In our implementation of the read-out method, we use the likelihood values $x^{(3)}_{MT,m}$ for motion velocities and velocity gradients, which we denote by $x^{(3)}_{MT,m}$ and $x^{(3)}_{MT,g}$, respectively. Motion velocities are computed by

$$
\begin{pmatrix}
\hat{x}(x, y) \\
\hat{y}(x, y)
\end{pmatrix} = \frac{N_{\phi,s}}{\sum_{i=1}^{N_{\phi,s}} x^{(3)}_{MT,m}(x, y, \phi_i, s_i) \cdot s_i \cdot \begin{pmatrix}
\cos \phi_i \\
\sin \phi_i
\end{pmatrix}} \sum_{i=1}^{N_{\phi,s}} x^{(3)}_{MT,m}(x, y, \phi_i, s_i)
\tag{2.15}
$$
Figure 3: Our biologically inspired model and the representation of image velocity and velocity gradients. (A) Box arrow diagram with forward and FB connections between model areas V1, MT, and MST. Motion is detected using Gabor filters, and velocity gradients are detected using antagonistic filters. Information about image velocity is spatially integrated in area MT, and information about velocity gradients is spatially integrated in area MST. (B) Velocity information is encoded as a map of likelihood values sampling motion direction and speeds. (C) Gradient information is encoded as a map of likelihoods sampling differences of motion directions and differences of motion speeds. Note that this depiction does not reflect the actual number of motion directions and speeds used in the simulations.
for each position \((x, y)\) sampled in the visual field. Equation 2.15 computes the weighted vector average using the likelihood values \(x_{MT,m}^{(3)}\). The weights are multiplied by the encoded velocity vector, which is defined by \(s_i \cdot (\cos \phi_i, \sin \phi_i)\). Velocity gradients are analogously computed replacing the activity \(x_{MT,m}^{(3)}\) by \(x_{MT,g}^{(3)}\) and the vectors \(s_i \cdot (\cos \phi_i, \sin \phi_i)\) by \(\Delta s_i \cdot (\cos \Delta \phi_i, \sin \Delta \phi_i)\).

3 Results

We show simulation results for the global and local interpretation of velocity gradients using rigid and nonrigid scenes and real-world videos, which contain IMOs.

3.1 A Superposition of Translational and Rotational Self-Motion Along or Around the Optical Axis Is Represented as a Spiral Motion Pattern. This simulation probes the capability of our model to process and represent spiral motion patterns. We render a video for a flight with a pinhole camera through a 30 m long cylindrical tunnel that is closed at both ends (Persistence of Vision Pty. Ltd., http://www.povray.org/). At all times, this pinhole camera is in the center of each fronto-parallel section, which results in a circular disk of 5 m diameter corresponding to a \(50^\circ \times 50^\circ\) visual field.

Initially the camera is at 5 m distance from one end of the tunnel, as depicted in Figure 4A. We show a snapshot of the initial frame in Figure 4B to give an impression of the video. During the video, the camera translates (denoted by the variable \(v_z\)) and rotates along the optical axis (denoted by \(\omega_z\)). Figure 4C shows the values of translation and rotation during the flight. A spiral motion occurs when translation and rotation are present. Figure 4D shows flow fields computed by our bio-inspired model, which are read out from model area MT (see equation 2.15).

The first plot of Figure 4D shows an expansion flow with radially outward directed arrows. Figure 4E shows the activity of the detected velocity gradients as encoded in model area MST. The cyclic encoding ranges from CONT to CCW, EXP, and CW. Likelihoods of speed differences and spatial position were summed for this plot. Figure 4E shows high activity for EXP during the first 15 frames of the sequence and transitions into high activity for CONT for frames 15 to 35, and so on. Figures 4F and 4G show the tangent and normal velocity gradient for sampled frames, respectively. In contrast to Figure 4E, these plots show the direction and strength of pattern responses, whereas the strength is encoded by the length of the vectors. At the FOE, velocity gradients are short, that is, slow in speed (see the first two plots of Figures 4F and 4G). Similarly, velocity gradients at the center of rotation (COR) are short in length (see the last two plots of Figures 4F and 4G). In our example, translations generate lower image speeds than
rotations. Subsequently, velocity gradients for rotational self-motion have longer vectors than those for translational self-motion. Compare the first two with the last two plots in Figures 4F and 4G with each other, respectively. Due to boundary effects, velocity gradients for expansion (frame 9) and contraction (frame 19) at the boundary differ in their length as well (see the first two plots in Figure 4F and 4G). This can also be observed in the flows that are depicted in Figure 4D. Next, we show the tuning of our model MST cells.

The tuning width of a model MSTd cell is measured using optic flow generated by a simulated translation along the optical axis, \( v_z = 0.1 \text{ m/frame} \), or rotation around the optical axis in front of a back plane, initially 5 m away. To match the produced speed distribution in the image plane for translation and rotation, we equal the flow magnitude of the generated flows, yielding \( v_z/Z = \omega_z \) (rotational velocity equals the ratio between the translational velocity along the optical axis and the distance \( Z \)). For our configuration, we compute the rotational velocity by \( \omega_z = 0.1 \text{ m/frame}/(5 \text{ m}) = 0.02 \text{ rad/frame} \) or \( \omega_z = 1.15^\circ/\text{frame} \). For spiral motion, we use only half of the translational and rotational velocity to match the speed distribution because translational and rotational flow superimpose linearly. In total, we use eight flow patterns—EXP, CONT, CW, CCW, and four combinations thereof—produce spiral flow patterns. Table 2 shows parameters of all flow patterns.

Figure 4I shows the tuning curve for a model MST cells, and Figure 4J shows the corresponding tuning curve from a recorded cell (Graziano et al., 1994). To fit these curves, we use the gaussian model function:

\[
f_{\text{Gauss}}(x; a, b, \sigma, \mu) = a \cdot \exp\left(-\frac{(x - \mu)^2}{2 \cdot \sigma^2}\right) + b,
\]

Figure 4: The encoding of global velocity gradients for a flight through a tunnel. (A) Top view of the tunnel and initial camera position. (B) Image frame for the initial starting position or frame number 0. (C) Linear and rotational velocity along or around the optical axis of the camera for 96 frames. Numbers in circles mark frames of the video. (D) Flows for these marked frames. (E) Likelihood values for flow patterns. A high likelihood is encoded in light and a low likelihood in dark. White dashed lines indicate the position of the marked frames. (F) Velocity gradients in tangential motion direction. (G) Velocity gradients normal to the motion direction. (H) Encoding schema of velocity gradients. The marked point indicates a clockwise, spiral, inward flow pattern. (I) The curve of a simulated model cell tuned to clockwise, spiral inward motion shows its response likelihood to eight flow patterns (circles). These data points were fitted by a gaussian curve. (J) Data and fitted gaussian curve from a recorded cell that is tuned to the same motion. (This panel has been redrawn from Graziano et al., 1994, Figure 7B, p. 59.) Flows were read out from model area MT, and to avoid clutter, we sampled them four times. Vector fields that show velocity gradients are sampled four times and scaled by a factor of eight.
Table 2: Motion Parameters Used to Generate Motion Types to Measure the Tuning of a Simulated MSTd Cell.

<table>
<thead>
<tr>
<th>Motion Type</th>
<th>$v_z$ (meter/frame)</th>
<th>$\omega_z$ (°/frame)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counterclockwise</td>
<td>0.00</td>
<td>1.150</td>
</tr>
<tr>
<td>Spiral counterclockwise and contraction</td>
<td>−0.05</td>
<td>0.625</td>
</tr>
<tr>
<td>Contraction</td>
<td>−0.10</td>
<td>0.000</td>
</tr>
<tr>
<td>Spiral clockwise and contraction</td>
<td>−0.05</td>
<td>−0.625</td>
</tr>
<tr>
<td>Clockwise</td>
<td>0.00</td>
<td>−1.150</td>
</tr>
<tr>
<td>Spiral clockwise and expansion</td>
<td>0.05</td>
<td>−0.625</td>
</tr>
<tr>
<td>Expansion</td>
<td>0.10</td>
<td>0.000</td>
</tr>
<tr>
<td>Spiral counterclockwise and expansion</td>
<td>0.05</td>
<td>0.625</td>
</tr>
</tbody>
</table>

where $a$ denotes the amplitude, $b$ the baseline firing rate in the recordings or baseline likelihood for our model, $\mu$ the mean value of the gaussian, and $\sigma$ the standard deviation of the gaussian. We fitted this gaussian model to the mean values of likelihood responses of the model and the mean spike rates of the recorded data. For our model cells, these values are $a = 0.74$ likelihood, $b = 0.28$ likelihood, $\mu = 182.5^\circ$, and $\sigma = 58.9^\circ$. Fitting recorded mean firing rates give $a = 33.84$ spikes/sec, $b = 0.73$ spikes/sec, $\mu = 180.0^\circ$, and $\sigma = 45.0^\circ$. The amplitude and baseline cannot be directly compared due to the different scale in use. The tuning bandwidth ($\sigma$) of our model MST cell is wider than that of its biological counterpart, and our model MST cell baseline likelihood is higher with respect to the amplitude (ratio $b/a$) than the recorded baseline firing rate. In all, our model cell is in the parameter range of recorded MSTd cells. The slight differences between model and data could be further reduced by, for example, strengthening the competitive interactions between flow pattern cells. In the next simulation, we show local properties of velocity gradients instead of their global properties that we interpreted as selectivity to flow patterns, which cover a large region of the visual field.

3.2 For Self-Motion, Local Acceleration and Deceleration Occur at Depth Discontinuities and Encode Ordinal Surface Depth at Kinetic Contours. To illustrate the local encoding properties of velocity gradients, we use the Flower Garden sequence (available at http://www.cs.brown.edu/~black/) which is well known in the computer vision community and contains strong depth discontinuities. Figure 5A shows a top view of the scenario with the camera in sideward translation. Figure 5B shows an image frame. Due to the depth dependency of the translational flow, the tree has larger image velocities than the background because it is closer. This is also visible in the flow in Figure 5C computed with our model. Figures 5D and 5E show the corresponding velocity gradient. For a local interpretation, we use the tangent velocity gradients. At the transition
Figure 5: Local velocity gradients of acceleration and deceleration occur at depth discontinuities for self-motion. (A) Top view of the scenario that shows position and movement of the camera. (B) Image frame number 23 of the sequence when counting from zero. (C) Detected flow. (D) Velocity gradient in tangential motion direction encoded as a vector field. (E) Velocity gradient normal to the motion direction. (F) Encoding schema for velocity gradients. Flow and velocity gradient are read out from model area MT with reference to the frame pair 23–24. To avoid clutter in depicting flows, we sampled them four times. Vector fields that show velocity gradients are scaled by a factor of 4.

between the background and tree, on the right edge of the tree trunk, a local acceleration appears when traversing from right to left. This transition from slow to fast along the local motion direction depicts a disocclusion edge (Beck, Ogniben, & Neumann, 2008). Locally, this transition is encoded as an EXP flow pattern. The velocity gradient points to the right; compare also with the encoding of velocity gradients in Figure 5F.
This local acceleration is similar to the EXP flow pattern because such a pattern also shows a local acceleration at each point in the optic flow field, excluding the singularity of the FOE. On the left side of the tree trunk, a transition from tree to background appears along the local motion direction depicting an occlusion edge. This is a transition from fast to slow speeds and can be seen as a local deceleration similar to what appears in an CONT flow pattern. Thus, velocity gradients in Figure 5D point to the left, which encodes a CONT pattern. Normal velocity gradients in Figure 5E show mainly the same; however, to achieve the same interpretation as in Figure 5D, the encoding of the velocity gradient space in Figure 5E would have to be rotated by 90° in the mathematical positive direction. These normal velocity gradients are a representation of the local flow perpendicular to the gauge coordinate system that is oriented along the tangential velocity gradients.

In the next simulation, we study the encoding of local velocity gradients for nonrigid scenes composed of object motions in front of a stationary camera.

3.3 Local Velocity Gradients Appear at Motion Boundaries in Scenes with Nonrigid Object Motions. For the computation of velocity gradients in nonrigid scenes, we chose the three image sequences Army, Wooden, and Mequon from the Middlebury benchmark (Baker et al., 2011). Each sequence is composed of eight image frames, and we show the previous to last image frame in Figures 6A, 6D, and 6G, respectively. In our model, we use all eight image frames and display the computed flow and velocity gradients of the last image pair. Figures 6B, 6E, and 6H show the flows and 6C, 6F, and 6I show the velocity gradients. To facilitate the interpretation of the velocity gradients, we labeled regions according to their local pattern response. For the Army sequence, the background on the left-upper quadrant is rotating counterclockwise (see Figure 6B), which is detected by our model (CCW response; see Figure 6C). At the transition between the fence and curtain in the upper half of the image, a local deceleration (CON response) is detected. The letter “O” and the square block at the bottom of the image sequence rotate counterclockwise, again detected by our model (CCW response). For the Wooden sequence, the upper part in the image rotates counterclockwise (see Figure 6E), which is detected by our model (CCW response; see Figure 6F). At the transition between the rotating block and the lower part, a clockwise pattern (CW response) is detected due to the upper part moving rightward and the lower part moving leftward (see Figure 1B, iv). On the left side of the same block, a spiral pattern (SP response) is detected. This pattern is generated due to the transition of a right-downward motion of the block into a leftward motion of the background. For the Mequon sequence, a greater variety of patterns occurs and is detected. Parts of the T-shirt above the diagonal rotate clockwise (see Figure 6H). Accordingly, velocity gradients indicate a local CW pattern (see Figure 6I). Local accelerations and decelerations appear at the depth discontinuities between the two figures and the T-shirt (patterns EXP and CON). A spiral (SP) pattern is detected
Figure 6: Velocity gradients detected for a nonrigid scene, are consistent with a local interpretation. (A) Frame 13 of the Army sequence, (B) the detected image motion, and (C) velocity gradients. (D) Frame 13 of the Wooden sequence, (E) the detected image motion, and (F) velocity gradients. (G) Frame 13 of the Mequon sequence, (H) image motion, and (I) velocity gradients. For all three sequences, motions and velocity gradients are computed for the frame pair 13–14 and are read out from model area MT. To avoid clutter in depicting flows, we sampled them four times. Vector fields showing velocity gradients are scaled by a factor of two (B), one half (E), one (H), eight (C), and four (F and I).

for the torque that appears on the right side in the image. Visual inspection and a case-by-case comparison show that the strong velocity gradients are consistent with the underlying local velocity changes.

3.4 Spiral Motions Indicate the Presence of Independently Moving Objects (IMOs) But Are Not Restricted to IMOs. We use three image sequences to study the behavior of the velocity gradients under general self-motion in combination with IMOs. The first sequence is the Yosemite sequence with clouds. This sequence shows a simulated flight with the linear velocity \( \vec{v} = (0, 0.17, 0.98) \cdot 34.8 \text{ pixels per frame} \) and rotational velocity \( \vec{\omega} = (0.14, 0.98, 0.17) \cdot 0.095 \text{ degrees per frame} \) (Heeger & Jespon, 1992). The clouds move two pixels per frame to the right. Figure 7A shows the fourteenth image frame and 7B and 7C show computed flow and velocity gradient for the frame pair 14–15, respectively. In the lower part of the
Figure 7: Spiral velocity gradients as an indicator for independently moving objects. (A) The thirteenth frame of the Yosemite sequence (available at http://www.cs.brown.edu/~black/). (B) Detected image motion for frame pair 13–14 and (C) velocity gradients. (D) Sequence with a moving object (rectangle). (E) Detected flow for frame pair 9–10 and (F) velocity gradients. (G) Same sequence but with a stationary object. (H) Detected flow and (I) velocity gradients. To avoid clutter in depicting flows, we sampled them four times. Vector fields showing velocity gradients are scaled by a factor of four.

image, an EXP pattern is detected, and at the horizon a CW pattern. The CW pattern is generated due to the rightward motion in the sky above the leftward motion close to the horizon (see Figure 1B, iv). In the left part of the image below the horizon, a spiral motion (SP) is detected, which is indicative of the combination of rotational and translational self-motion. The next two image sequences are identical except that the rectangular object moves for Figures 7D to 7F and is stationary for Figures 7G to 7I. The camera moves with $\vec{v} = (0.05, 0, 0.05)$ m per frame and $\vec{\omega} = (0, 0.6, 0.0)$ degrees per frame and points 10 degrees to the ground. The object has a leftward motion of 0.05 meters per frame. The ground is 1.5 m below the camera, and initially the back plane is 10 m away, while the object has an initial distance of 4 m.
Figure 7F shows SP velocity gradients at the four corners of the IMO. In addition, an SP gradient is detected in the lower-right corner of the image. In Figure 7I, the static object shows SP, EXP, and CON patterns at its boundary. These responses are weaker compared to those of the IMO. These three examples suggest that the interpretation of velocity gradients for IMOs is consistent under a variety of general motions, including rotations.

4 Discussion

We developed a bio-inspired, computational model for the detection of velocity gradients from optic flow, which in turn was estimated from rendered or real-world videos. Our model uses an encoding of motion speed differences and motion direction differences. Likelihood values represent the presence or absence of a velocity gradient in each spatial location of the visual field. Model MSTd cells represent a continuum of spiral motions with a tuning similar to their biological counterpart (see Figure 4). Optic flow for self-motion in a rigid scene contains acceleration or deceleration at depth discontinuities. Due to the characteristics of translational optic flow, close surfaces generate faster image velocities than surfaces that are farther away. A local acceleration for the tangent velocity gradient encodes a transition from far to close along the motion direction. A local deceleration encodes a transition from close to far, again along the motion direction. Cells selective to velocity gradients can encode the ordinal surface depth at surface borders as transition from close to far or far to close.

Our simulations suggest an interpretation of scene structure from optic flow fields based on velocity gradients. A global interpretation of velocity gradients allows for a discrimination of types of self-motion, while a local interpretation of velocity gradients yields ordinal depth at surface borders assuming self-motion in a rigid scene. Spiral velocity gradients appear for IMOs and static objects assuming general motion. For only translational self-motion, these spiral velocity gradients indicate IMOs. In a scenario of only 2D shift motions in the image plane, IMOs are identified by CW or CCW velocity gradients. Table 3 summarizes these results on IMOs.

Velocity gradients are more powerful in their encoding capabilities than speed gradients. Tsotsos et al. (2005) propose the computation of speed gradients using the same likelihood encoding of motion velocities that we use for our model. For translational self-motion across a tilted plane that contains a depth gradient, the encoding of speed gradients can drift from global EXP/CONT flow patterns toward spiral motion. For this tilted plane, the fastest neighboring motion is not necessarily located along the reference motion direction but, for example, to the left or right of it. Thus, speed gradients indicate a spiral motion pattern for planes with a depth gradient. Our velocity gradient has no drift to spiral motion patterns, in this case, of a plane with a depth gradient because its gradient is computed tangential
Table 3: Local Velocity Gradients at Object and Motion Discontinuities Assuming Translation, Rotation, or Shift for the Self-Motion or Object Motion.

<table>
<thead>
<tr>
<th>Type of Motion</th>
<th>Object That Introduces a Discontinuity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-motion</td>
<td>Object motion</td>
</tr>
<tr>
<td>Translation</td>
<td>Translation and rotation</td>
</tr>
<tr>
<td></td>
<td>EXP, CONT, CW, CCW, SP</td>
</tr>
<tr>
<td>Translation and</td>
<td>Translation and rotation</td>
</tr>
<tr>
<td>rotation</td>
<td>EXP, CONT, CW, CCW, SP</td>
</tr>
<tr>
<td>Shift</td>
<td>Shift</td>
</tr>
<tr>
<td></td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>EXP, CONT, CW, CCW, SP, SP</td>
</tr>
</tbody>
</table>

Notes: Local velocity gradients are EXP for expansion or acceleration, CONT for contraction or deceleration, CW for clockwise rotation, CCW for counterclockwise rotation, and SP for spiral patterns. Translations are $\vec{v} = (v_x, v_y, v_z)$, rotations are $\vec{\omega} = (\omega_x, \omega_y, \omega_z)$, and shifts are $\vec{v} = (v_x, v_y, 0)$.

or normal to the local motion direction. This computes the gradient of speed and direction separately instead of computing the direction of the highest speed increase for the speed gradients. Thus, velocity gradients discriminate more motion patterns than speed gradients.

With our model, we propose a scheme for the detection of velocity gradients from likelihood-encoded image motions in which the mechanisms of gradient computation are predefined and hard-wired mechanistically. Globally these velocity gradients represent optic flow patterns. An alternative approach is to learn selectivity for optic flow patterns. Beardsley and Vaina (1998) suggest a two-layer backpropagation network to study the mechanisms that lead to position invariance, gaussian tuning profiles, and multicomponent selectivity (Duffy & Wurtz, 1991a, 1991b) and a continuum of selectivity to spiral motions (Graziano et al., 1994). Zemel and Sejnowski (1998) suggest a network consisting of an input layer that models area MT neurons using a population code with speed and direction selectivity, a middle layer to model area MST. The output layer copies the input layer. After training, MST cells show selectivity to the continuum of spiral motions. Furthermore, the hidden layer cell responses are used to recover heading employing an additional network. Except for the singularity points FOE or COR, our model MST velocity gradients show a position-invariant sensitivity in the motion pattern encoding of model MSTd cells. The tuning profile of our model MSTd cells is a gaussian as well (see Figure 4I). However, our model does not incorporate cells with multicomponent selectivity (Duffy & Wurtz, 1991a, 1991b). These models focus on the ability of self-organization of flow pattern–selective cells at the midlevel processing of optic flow in cortex. Complementary to this, we propose biologically plausible mechanisms based on the current knowledge of flow pattern–sensitive cells as recorded in cortical areas MT and MST, respectively. We also emphasize that
the proposed neural mechanism achieves robust computational behavior when applied to real-world video.

In another model of cells in the dorsal part of MST, optic flow patterns are mapped into the log-monopole domain (Grossberg, Mingolla, & Pack, 1999). This mapping contributes to the nearly position invariance of MST cells; a selectivity for the FOE of these cells is absent. However, the model is still able to recover heading with respect to gaze from a population of MST model cells that are selective to EXP, CONT, CW, and CCW flow patterns. When we use the log-monopole mapping, these complex flow patterns are transformed into simple laminar patterns of flow, much like our velocity gradient representation. Selectivity is encoded in the same gradient space where one axis denotes CONT-EXP selectivity and the other, orthogonal axis CCW-CW selectivity. In their model, the angle defined by responses encoded along the two axes indicates the heading direction. Our model focuses on the local (model area MT) and global (model area MST) interpretation of velocity gradients.

Models with the same dynamical equations and three-level processing cascade have been proposed in earlier work. Bayerl and Neumann (2004) suggested the processing cascade to model optic flow processing in cortical areas V1 and MT (compare also Bouecke et al., 2011). In particular, their model shows the disambiguation of initial motions in model area V1 by MT FB over time. Time was simulated by the integration of motion from successive image frames. Raudies et al. (2011) extended the model by refining its interactions in velocity space of model area MT. A soft competition between velocities allows the stable representation of multiple likelihoods and, thus, visual motions in configurations of transparent moving surfaces. Escobar and Kornprobst (2012) extended the V1/MT model of visual motion processing by the filtering of MT motion likelihoods with symmetric and asymmetric weight kernels. Using these filtering results, they could show an improvement in the classification of biological motion stimuli compared to using only motion signals. Our current model builds on these earlier models. It uses the likelihood representation of model area MT to detect velocity gradients. These velocity gradients are integrated spatially and temporally by replicating the three-level processing cascade for model areas MT and MST concerned with the processing of velocity gradients (see Figure 3).

Methods of texture segmentation use the same idea of a local analysis in tangent and normal direction. Ben-Shahar and Zucker (2003) analyze texture flow, which is defined by the main orientation of a texture item at each location using two curvatures—one in tangential and another in normal direction. Our model does not aim for a flow-based segmentation of image regions. We used a local gauge coordinate system defined by tangent and normal direction to simplify the notation of complex EXP, CONT, CW, and CCW flow patterns.

Template-based neural networks for self-motion estimation use similar flow patterns as ours. Perrone and Stone (1994) propose a flow template
matching method for the estimation of self-motion. Flow templates for expansion with different FOEs, like MST cell’s selectivity described by Duffy and Wurtz (1995), are modeled in combination with flow templates for fixating rotational flow. In addition to these different self-motions, possible depths are sampled in the range of 2 m to 32 m at a walking speed of 1 m per second. All flow templates are matched against the sensed flow, and the template with the best match wins. The parameters of this winning template define the self-motion estimate. Lappe and Rauschecker (1993) propose a neural network consisting of three layers: an input, middle, and output layer. The input layer represents optic flow in a population code of direction and speed selectivity similar to our model, not using a log-encoding for motion speeds. By projection to the middle layer, the dependency on rotational self-motion parameters is cancelled using a subspace of altered flow templates that is independent of rotation and depth (see Heeger & Jepson, 1990), for details of the subspace construction. The output layer consists of templates sensitive for different heading directions while assuming a constant speed for locomotion. Our model extracts optic flow from image sequences and emphasizes the detection of types of self-motion through velocity gradients rather than estimating parameters of 3D translation and 3D rotation.

Acknowledgments

F.R. is supported in part by CELEST, a National Science Foundation Science of Learning Center (NSF SMA-0835976), and the Office of Naval Research (ONR N00014-11-1-0535 and ONR MURI N00014-10-1-0936). H.N. and S.R. acknowledge support by a grant from the German Federal Ministry of Education and Research, project 01GW0763, Brain Plasticity and Perceptual Learning. H.N. is further supported by the Transregional Collaborative Research Center, “A Companion Technology for Cognitive Technical Systems” funded by the German Research Foundation (DFG).

References


Received July 29, 2012; accepted March 7, 2013.