Emergence of Position-Independent Detectors of Sense of Rotation and Dilation with Hebbian Learning: An Analysis

Kechen Zhang
Martin I. Sereno
Margaret E. Sereno*
Department of *Cognitive Science, University* of *California, San Diego,
La Jolla, CA* 92093-0515 *USA*

We previously demonstrated that it is possible to learn position-independent responses to rotation and dilation by filtering rotations and dilations with different centers through an input layer with MT-like speed and direction tuning curves and connecting them to an MST-like layer with simple Hebbian synapses (Sereno and Sereno 1991). By analyzing an idealized version of the network with broader, sinusoidal direction-tuning and linear speed-tuning, we show analytically that a Hebb rule trained with arbitrary rotation, dilation/contraction, and translation velocity fields yields units with weight fields that are a rotation plus a dilation or contraction field, and whose responses to a rotating or dilating/contracting disk are exactly position independent. Differences between the performance of this idealized model and our original model (and real MST neurons) are discussed.

1 Introduction

A major stream of motion information processing in the primate visual system goes from layer 4B in primary visual cortex (Vl) to the middle temporal area (MT) and then to the medial superior temporal area (MST) (for reviews see Sereno and Allman 1991; Felleman and Van Essen 1991). Most neurons in area MT have moderate-sized receptive fields, and a subset is tuned to the local pattern velocity (Movshon *et al.* 1985). Neurons in the dorsal part of MST, by contrast, have much larger receptive fields and some are selective to higher order motion features—for example, rotation (either clockwise or counterclockwise, but not both), and dilation or contraction (but not both) on the frontoparallel plane (Saito *et al.* 1986; Sakata *et al.* 1986; Tanaka and Saito 1989; Tanaka *et al.* 1989; Duffy and Wurtz 1991a,b). Detecting rotation, dilation, and contraction provides useful information about an animal's motion relative to the environment or about the intrinsic motion of an object (Koenderink and

---

*Present address: Department of Psychology, University of Oregon, Eugene, OR 97403.

van Doorn 1975, 1976; Longuet-Higgins and Prazdny 1980; Koenderink 1986).

An interesting property is that some dorsal MST neurons give nearly identical responses to a rotation, or dilation or contraction, no matter where the center of the velocity flow is located. We sought to find a neural mechanism for this position invariance. To be selective to a rotation or dilation/contraction with a fixed center, the receptive field of an MST neuron need just be composed of the MT neurons whose preferred directions are arranged circularly or radially around that center (Saito et al. 1986). At first glance, this simple mechanism would not seem to be able to support an invariant response when the position of the center changes (Saito et al. 1986; Tanaka et al. 1989; Duffy and Wurtz 1991b).

Two previous proposals for a position-independent mechanism assume a homogeneous organization for an MST receptive field to ensure that all its subfields have identical structure and function. In one model, the local rotation and dilation of the velocity field is first derived and then summed up across space to get invariant responses (Duffy and Wurtz 1991b). This algorithm requires that MT neurons be selective to local rotation and dilation/contraction, which is generally not the case (Tanaka et al. 1986). Another model makes use of partially overlapping compartments in an MST receptive field (Saito et al. 1986). But this model needs a special surround effect in MT neurons to prevent many compartments from being activated simultaneously, the exact mechanism of which awaits further experimental proof.

A simpler yet counterintuitive solution was discovered in a computer simulation experiment using a feedforward network and unsupervised learning (Sereno and Sereno 1990, 1991). That work was based on a previous study in which Hebbian learning was used to find a solution to the aperture problem in a two-layer feedforward network corresponding to the connections from V1 $\rightarrow$ MT (Sereno 1989). When a similar network (with a larger interlayer divergence) representing MT $\rightarrow$ MST connections is trained with rotation, dilation, and contraction using a Hebb rule and input-layer units with MT-like tuning curves, MST-like units with position-independent responses emerge. Surprisingly, such rotation or dilation/contraction detectors turned out to have inhomogeneous receptive fields with a circular, spiral, or radial arrangement of local direction selectivity, just as in the simple mechanism mentioned before.

In this letter we analyze a modified version of the original model in Sereno and Sereno (1991). The input layer of the modified model has broader (cosine) tuning curves than in the original model (and broader than those of real MT neurons), but it allows us to derive explicit expressions for the course of learning empirically observed in the original model. The modified model gives rise to MST-like units that linearly decompose the flow field into flow field components — for example, a clockwise rotation-preferring unit will respond as well to the rotation in a clockwise spiral as to a pure clockwise rotation, ignoring any added di-

lation/contraction. By contrast, in our original model, the sharper tuning curves for the MT-like units result in MST-like units whose response falls off as other optic flow components are added. This smooth fall-off has also been observed with real MST neurons (Graziano *et al.* 1990; Orban *et al.* 1992). It is important to note, however, that the basic mechanism of position-invariant response to flow field stimuli (a position-variant direction-tuning template) is identical in both the idealized model with cosine tuning curves as well as the original model with narrower tuning curves. Linear decomposition might yet be found in areas beyond MST. It would be useful for filtering out certain movement components (e.g., translation) while exactly signaling the magnitude of others of interest (e.g., dilation). On the other hand, tighter input-layer tuning curves allow individual output layer units to code more information about a flow field (see Discussion).

2 A Mechanism for Position Independence

First, we show the basic principles for the position-independent responses, as initially revealed by computer simulation (and recently independently derived in similar form by Poggio *et al.* 1990, 1991). Let $\mathbf{v} = \mathbf{v}(\mathbf{r})$ be the velocity field on the image plane (frontoparallel plane), where the vector $\mathbf{r} = x\mathbf{i} + y\mathbf{j}$ denotes the position with $\mathbf{i}$ and $\mathbf{j}$ being the unit vectors of x and y axes. Consider MT-like units that are sensitive to the local stimulus velocities $\mathbf{v}$. Each MT-like unit has a preferred direction. Given $\mathbf{v}$ as the stimulus velocity at a fixed position, the response or activation a of the MT-like unit at that position is assumed to be proportional to the velocity component in the preferred direction, or

$$a = cv\cos(\theta - \phi) = c\mathbf{d}_\phi \cdot \mathbf{v} \tag{2.1}$$

where $v = |\mathbf{v}|$ is the stimulus speed, c is a constant coefficient representing the slope of the (linear) speed tuning, θ is the direction angle of $\mathbf{v}$, and $\mathbf{d}_\phi$ is the unit vector for the preferred direction angle ϕ. In other words, the unit has linear response to the speed v and a sinusoidal direction tuning curve with the maximum at the preferred direction ϕ (Fig. 1).

The artificial MT-like units resemble the real neurons in area MT of monkey in certain respects (Rodman and Albright 1987). For most MT neurons, speed does not alter the shape of direction tuning curves, which implies a multiplicative interaction of the speed tuning and the direction tuning as used in expression 2.1. The linear speed tuning is a reasonable approximation for small speeds, although the response is sometimes reduced when the speed exceeds an optimum value. The sinusoidal direction tuning is broader than a typical real MT neuron (Maunsell and Van Essen 1983). Also, the responses of real MT neurons to the antipreferred direction are usually smaller. We retain expression 2.1 for its simplicity and ease of analysis.
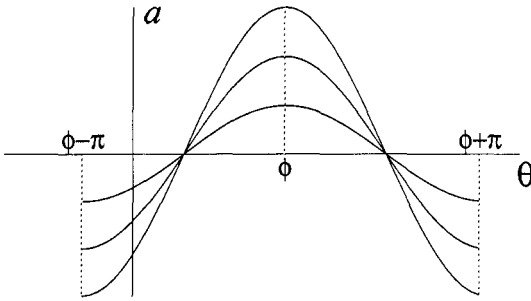
Figure 1: A family of direction tuning curves for different speeds.

Now consider an MST-like unit that receives inputs from many MT-like units. It is convenient to define the weight *vector field* of an MST-like unit. For any position on the image plane, we define the weight vector $\mathbf{w}$ at that point as

$$\mathbf{w} = cw\mathbf{d}_\phi \tag{2.2}$$

where $\mathbf{d}_\phi$ is the preferred direction of the MT-like unit at that position, w is the scalar weight for its connection to the MST-like unit, and c is the same constant coefficient as in 2.1. The total input $\mathbf{I}$ to the MST-like unit is assumed to be the weighted sum of the inputs from all the MT-like units within the receptive field of the MST-like unit:

$$I = \sum_\mathbf{r} wa = \sum_\mathbf{r} \mathbf{w} \cdot \mathbf{v} \tag{2.3}$$

where the simple relation (see equations 2.1 and 2.2) $wa = wc\mathbf{d}_\phi \cdot \mathbf{v} = \mathbf{w} \cdot \mathbf{v}$ has been used. The output of the MST-like unit is simply

$$O = \sigma(I)$$

where $\sigma(\ )$ is a sigmoid function.

If the weight vector field of an MST-like unit is itself a rotational field, namely,

$$\mathbf{w} = \mathbf{\Omega} \times \mathbf{r} = -\Omega y\mathbf{i} + \Omega x\mathbf{j} \tag{2.4}$$

where vector $\mathbf{\Omega} = \Omega\mathbf{k}$ can be regarded as the "angular velocity" for the weight vector field $\mathbf{w}$, with $\mathbf{k}$ being the unit vector of the z axis (perpendicular to the image plane), then we can prove that the MST-like unit's response O to a rotating disk of angular velocity $\boldsymbol{\omega} = \omega\mathbf{k}$ depends only on the angular speed ω of the stimulus but not on the location of the stimulus disk.
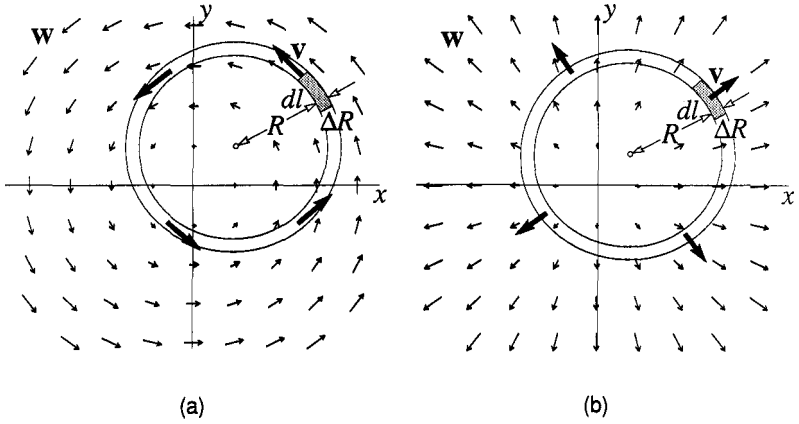
Figure 2: Response elicited by a rotating (a) or dilating (b) ring in a receptive field with circular or radial distribution of direction selectivity is independent of the position of the ring (see text).

To show this, decompose the stimulus disk into many concentric rings and calculate the response elicited by a single rotating ring of radius R and width ΔR (Fig. 2a). Let $\mathbf{v}$ be the velocity field of the stimulus ring, and $\Delta I = \sum \mathbf{w} \cdot \mathbf{v}$ be the increment to the total input to the MST-like unit contributed by all the MT-like units within the area covered by the ring. We treat the weight vector field $\mathbf{w} = \mathbf{w}(\mathbf{r})$ as depending continuously on the position $\mathbf{r}$. The number of units for a unit area on the image plane is assumed to be a constant, and is taken as unity for simplicity. Replacing the sum by the integration along the ring, we get

$$\Delta I = \Delta R \oint \mathbf{w} \cdot \mathbf{v} \, dl = v \Delta R \oint \mathbf{w} \cdot d\mathbf{l} = v \Delta R \int_{S} (\nabla \times \mathbf{w}) \cdot \mathbf{k} \, dS \qquad (2.5)$$

where $d\mathbf{l} = (\mathbf{v}/v) dl$ with $\mathbf{v}/v$ being the unit vector in the circular direction, and S is the area enclosed by the ring. The last equality is a direct application of Stokes' theorem, where dS is the area element. Since the weight vector field 2.4 has a constant curl $\nabla \times \mathbf{w} = (\partial w_y/\partial x - \partial w_x/\partial y)\mathbf{k} = 2\Omega\mathbf{k}$, the last integral in equation 2.5 is equal to $\int_{S} 2\Omega \, dS = 2\pi\Omega R^2$. Since the rotational speed v of the stimulus ring is proportional to its radius ($v = \omega R$), we finally obtain

$$AI = 2\pi\omega\Omega R^3 \Delta R$$

which is independent of the position of the stimulus ring. The total response to the disk of radius ρ is therefore

$$O = \sigma(I) = \sigma\left(\pi\omega\Omega\rho^4/2\right)$$

which is also position independent. If the angular speed ω changes sign, that is, the disk rotates in the opposite direction, the total input I also changes sign.

To get position-independent responses to dilation or contraction, just let the weight vector field be

$$\mathbf{w} = \Lambda\mathbf{r} = \Lambda x\mathbf{i} + \Lambda y\mathbf{j} \tag{2.6}$$

which is itself a "dilation" when constant $A > 0$ and a "contraction" when $A < 0$. This vector field has constant divergence $\nabla \cdot \mathbf{w} = \partial w_x/\partial x + \partial w_y/\partial y = 2\Lambda$ (cf. the expression for constant curl above, except that div is a scalar). The response elicited by a dilating ring (Fig. 2b) is independent of the position of its center. The proof is similar, but Gauss' theorem is used to evaluate the integral:

$$\oint \mathbf{w} \cdot \mathbf{v}\, dl = v \oint \mathbf{w} \cdot \mathbf{n}\, dl = v \int_S \nabla \cdot \mathbf{w}\, dS = 2\pi\lambda\Lambda R^3$$

where $\mathbf{n} = \mathbf{v}/v$ is the unit vector in the radial direction of the ring, and $v = \lambda R$, where λ specifies the rate of dilation (as ω specifies the rotation speed). It follows that the response to a dilating disk is also position independent. In the case that the stimulus is a contraction, the input just changes sign.

This result can be intuitively appreciated by considering Figure *3.* In (a), the stimulus is centered. Since the local stimulus direction $\mathbf{v}(\mathbf{r})$ always agrees with the weighted local preferred direction $\mathbf{w}_\phi(\mathbf{r})$ in the receptive field, the dot product between each pair is positive, though small. In (b), with the stimulus center situated to the right of the receptive field center, local direction selectivity and local stimulus direction clash near the center of the receptive field—the dot products there are actually negative; but the negative terms are compensated, exactly, as we have seen, by the larger positive dot products in the periphery of the receptive field.

3 Development of Weight Vector Field under a Hebb Rule _____

Now we consider the general manner in which weight vector field changes during Hebbian learning. At each position on the image plane we use a set of MT-like units with different preferred directions. Let $w_\phi(\mathbf{r})$ denote the weight of the MT-like unit at position $\mathbf{r}$ with preferred direction ϕ. As before, the response or activation of the MT-like unit to the velocity field $\mathbf{v}(\mathbf{r})$ is

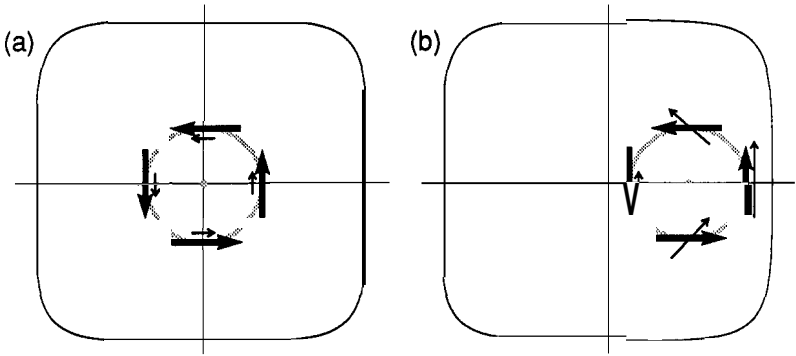$$a_\phi(\mathbf{r}) = c_\phi(\mathbf{r})\mathbf{d}_\phi \cdot \mathbf{v}(\mathbf{r}) \tag{3.1}$$

Figure 3: An intuitive interpretation of the mechanism for position independence. Negative dot products near the center of the receptive field in (b) are compensated by larger ones peripherally to give the same sum as in (a).
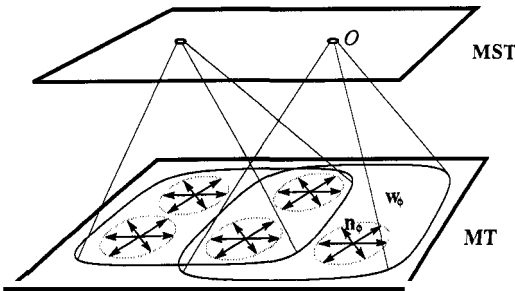

Figure 4: Each unit in the MST layer receives inputs from MT-like units at different positions and with different preferred directions (indicated by arrows).

and the weight vector is defined as $\mathbf{w}_\phi(\mathbf{r}) = c_\phi(\mathbf{r}) w_\phi(\mathbf{r}) \mathbf{d}_\phi$, with $\mathbf{d}_\phi$ again being the unit vector for the preferred direction of an MT-like unit. The total input I to the MST-like unit is the weighted sum of the responses from all MT-like units within the receptive field (Fig. 4), i.e., summing

over different positions in the receptive field as well as different preferred directions:

$$I = \sum_{\mathbf{r}} \sum_{\phi} w_\phi(\mathbf{r}) a_\phi(\mathbf{r}) = \sum_{\mathbf{r}} \sum_{\phi} \mathbf{w}_\phi(\mathbf{r}) \cdot \mathbf{v}(\mathbf{r}) \tag{3.2}$$

where the identity $w_\phi(\mathbf{r}) a_\phi(\mathbf{r}) = w_\phi(\mathbf{r}) c_\phi(\mathbf{r}) \mathbf{d}_\phi \cdot \mathbf{v}(\mathbf{r}) = \mathbf{w}_\phi(\mathbf{r}) \cdot \mathbf{v}(\mathbf{r})$ is used.

We can treat the system as if there were only one MT-like unit specified by the weight vector at each position (call this the equivalent weight vector)

$$\mathbf{w}(\mathbf{r}) := \sum_\phi \mathbf{w}_\phi(\mathbf{r}) = \sum_\phi c_\phi(\mathbf{r}) w_\phi(\mathbf{r}) \mathbf{d}_\phi$$

so we can write equation **3.2** as

$$I = \sum_{\mathbf{r}} \mathbf{w}(\mathbf{r}) \cdot \mathbf{v}(\mathbf{r}) \tag{3.3}$$

which is exactly the same as expression **2.3** in the previous section. **As** before, the output of the MST-like unit is

$$O = \sigma(I)$$

Suppose the increment of the weight in each training step follows a simple Hebb rule

$$\Delta w_\phi(\mathbf{r}) = \epsilon a_\phi(\mathbf{r}) O \tag{3.4}$$

In the present model no explicit distinction has been made between excitatory and inhibitory synaptic connections, and the weights are allowed to change sign. Since $\mathbf{w}(\mathbf{r}) = \sum_\phi c_\phi(\mathbf{r}) w_\phi(\mathbf{r}) \mathbf{d}_\phi$, the corresponding increment of the equivalent weight vector field is

$$\Delta \mathbf{w}(\mathbf{r}) = \sum_\phi c_\phi(\mathbf{r}) \Delta w_\phi(\mathbf{r}) \mathbf{d}_\phi = \sum_\phi \left[\epsilon c_\phi^2(\mathbf{r}) \mathbf{d}_\phi \cdot \mathbf{v}(\mathbf{r}) O \right] \mathbf{d}_\phi$$

where the last equality is obtained by substituting equation **3.1** into **3.4**. The coefficient $c_\phi(\mathbf{r})$ is assumed to be a random variable with uniform distribution across the angle ϕ. Let $\bar{c}^2$ be the average of $c_\phi^2(\mathbf{r})$. It is assumed to be a constant across the image plane. As an approximation for large number of units, we have

$$\Delta \mathbf{w}(\mathbf{r}) = \bar{c}^2 \sum_4 [\mathbf{d}_\phi \cdot \epsilon O \mathbf{v}(\mathbf{r})] \mathbf{d}_\phi \tag{3.5}$$

To simplify this expression, first note that if there are $n (\geq \mathbf{3})$ unit vectors $\{\mathbf{d}_\phi\}$ distributed evenly around the unit circle, then

$$\sum_\phi (\mathbf{d}_\phi \cdot \mathbf{V}) \mathbf{d}_\phi = \frac{n}{2} \mathbf{V} \tag{3.6}$$

holds for all vector $\mathbf{V}$. For a proof, write each vector as a complex number, namely, $\mathbf{V} = \rho e^{i\theta}$ and $\mathbf{d}_\phi = e^{i\phi}$, where θ is the direction angle of $\mathbf{V}$ and $\rho = |\mathbf{V}|$ is the radius. Because $\mathbf{d}_\phi \cdot \mathbf{V} = \rho \cos(\theta - \phi) = \frac{1}{2}\rho \left(e^{i(\theta - \phi)} + e^{-i(\theta - \phi)} \right)$ is a real number, we have

$$\sum_\phi (\mathbf{d}_\phi \cdot \mathbf{V}) \mathbf{d}_\phi = \sum_\phi \frac{1}{2}\rho(e^{i\theta}e^{-i\phi} + e^{-i\theta}e^{i\phi})e^{i\phi} = \frac{n}{2}\rho e^{i\theta} + \frac{1}{2}\rho e^{-i\theta} \sum_\phi e^{2i\phi}$$

Since ϕ is evenly distributed around the circle, $\sum_\phi e^{2i\phi} = 0$. This proves equation 3.6.

Assuming that the preferred directions of the MT-like units at each spatial position are evenly distributed, we can employ formula 3.6 by identifying $\mathbf{V}$ with $\bar{c}^2\epsilon O\mathbf{v}(\mathbf{r})$ so that equation 3.5 can be rewritten as

$$\Delta\mathbf{w}(\mathbf{r}) = \frac{1}{2}n\epsilon\bar{c}^2 O\mathbf{v}(\mathbf{r}) \tag{3.7}$$

where n is the number of the MT-like units at each position. This increment is caused by a single training step with the velocity field $\mathbf{v}(\mathbf{r})$. After training with a sequence of velocity fields, the equivalent weight vector field adds up to

$$\mathbf{w}(\mathbf{r}) = \mathbf{w}_0(\mathbf{r}) + \frac{1}{2}n\epsilon\bar{c}^2 \sum_t O\mathbf{v}(\mathbf{r}) \tag{3.8}$$

where $t(= 0, 1, 2, \ldots)$ stands for all time steps in the training and $\mathbf{w}_0(\mathbf{r})$ is the initial weight vector. In conclusion, the final equivalent weight vector field is just proportional to the sum of the training velocity fields weighted by the corresponding responses of the MST-like unit.

4 Training with Translation, Rotation, Dilation, and Contraction ____

We are now ready to consider the training with translation, rotation, dilation, and contraction velocity fields. To begin with, suppose for a single training step the velocity field is a rotation centered at $\mathbf{r}_c$

$$\mathbf{v}(\mathbf{r}) = \boldsymbol{\omega} \times (\mathbf{r} - \mathbf{r}_c) \tag{4.1}$$

and in different steps both the angular velocity $\boldsymbol{\omega}$ and the center $\mathbf{r}_c$ vary randomly. Substituting equation 4.1 into 3.8 and ignoring the initial weight vector for its smallness, we obtain the final weight vector field

$$\mathbf{w}(\mathbf{r}) = \eta \sum_t O\mathbf{v}(\mathbf{r}) = \eta \left(\sum_t O\boldsymbol{\omega} \right) \times \mathbf{r} - \eta \sum_t (O\boldsymbol{\omega} \times \mathbf{r}_c)$$

where $\eta = n\epsilon\bar{c}^2/2$ is a constant. This can be identified with the rotational field

$$\mathbf{w}(\mathbf{r}) = \boldsymbol{\Omega} \times (\mathbf{r} - \mathbf{r}_0) \tag{4.2}$$
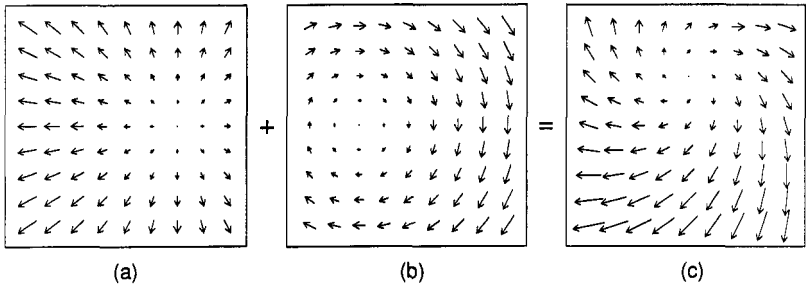
Figure 5: The final weight vector field is generally composed of a rotation field (b) and a dilation or contraction field (a). The result (c) is a spiral field.

where the weight field "angular velocity" Ω and the weight field center $\mathbf{r}_0$ are defined by $\Omega := \eta \sum_t O w$ and $\Omega \times \mathbf{r}_0 := \eta \sum_t (O \boldsymbol{\omega} \times \mathbf{r}_c)$. The latter equation has a unique solution of $\mathbf{r}_0$ as long as $\Omega \neq 0$. In the special case $\Omega = 0$, $\mathbf{w}(\mathbf{r})$ is a constant vector field (translation).

Similarly, training with dilation or contraction

$$\mathbf{v}(\mathbf{r}) = \lambda(\mathbf{r} - \mathbf{r}_c)$$

with rate λ and center $\mathbf{r}_c$ varying in time will lead to the final weight vector field

$$\mathbf{w}(\mathbf{r}) = \Lambda(\mathbf{r} - \mathbf{r}_0) \tag{4.3}$$

where $\Lambda := \eta \sum_t O X$ and $\Lambda \mathbf{r}_0 := \eta \sum_t O \lambda \mathbf{r}_c$. This is either a dilation ($\Lambda > 0$) or a contraction ($\Lambda < 0$). In the special case $\Lambda = 0$, $\mathbf{w}(\mathbf{r})$ is a constant (translation).

Note that expressions 4.2 and 4.3 are just what are required for position-independent responses (cf. equations 2.4 and 2.6). It should be realized that the center $\mathbf{r}_0$ does not affect the curl and divergence of a vector field, and thus does not affect our previous conclusions.

For training with a mixture of translations, rotations, dilations, and contractions, it is readily shown by similar argument that the final weight vector field takes the form

$$\mathbf{w}(\mathbf{r}) = \Omega \times (\mathbf{r} - \mathbf{a}) + \Lambda(\mathbf{r} - \mathbf{b}) + \mathbf{c}$$

It can always be written equivalently as

$$\mathbf{w}(\mathbf{r}) = \Omega \times (\mathbf{r} - \mathbf{r}_0) + \Lambda(\mathbf{r} - \mathbf{r}_0)$$

which is a spiral centered at $\mathbf{r}_0$ (Fig. 5). An MST-like unit with a spiral

weight vector field has position-independent responses to a particular sense of rotation as well as to either a dilation or contraction.

Even if the training velocity fields have a zero average, for example, clockwise and counterclockwise rotations have an equal chance of appearing, the weight vector field is still expected to grow with time. We consider the simple case where all rotations and dilations/contractions are centered at the same point so that the development of the two corresponding components is strictly independent. Consider the initial stage of development for training with, say, rotation fields alone. Now we need consider only the linear range of the sigmoid function σ, and for simplicity we assume $O = I$. According to equations 3.7 and 3.3, at time step $t + 1$

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta O_t \mathbf{v}_t = \mathbf{w}_t + \eta \left(\sum_{\mathbf{r}} \mathbf{w}_t \cdot \mathbf{v}_t \right) \mathbf{v}_t$$

where the subscripts refer to time. Thus

$$O_{t+1} = \sum \mathbf{w}_{t+1} \cdot \mathbf{v}_{t+1} = \sum \mathbf{w}_t \cdot \mathbf{v}_{t+1} + \eta \left[\sum \mathbf{w}_t \cdot \mathbf{v}_t \right] \left[\sum \mathbf{v}_t \cdot \mathbf{v}_{t+1} \right]$$

It can be expressed as

$$O_{t+1} = A\Omega_t \omega_{t+1} + \eta(A\Omega_t \omega_t)(A\omega_t \omega_{t+1}) \tag{4.4}$$

where Ω_t and ω_t are the angular speeds for the vector fields $\mathbf{w}_t$ and $\mathbf{v}_t$ at time t, respectively, and A is a constant depending on the size and shape of the receptive field as well as the position of the rotation center.

Imagine an ensemble of parallel training sessions starting from different initial weights and using different rotation sequences of random angular speeds, which are independent of each other while having identical statistics. We take the ensemble average on both sides of equation 4.4 to get $\langle O_{t+1} \rangle = A \langle \Omega_t \rangle \langle \omega \rangle + \eta A^2 \langle \Omega_t \rangle \langle \omega^2 \rangle \langle \omega \rangle$, where the subscript for the angular speed ω is dropped because the statistics of ω does not change over time. If $\langle \omega \rangle = 0$, then $\langle O_{t+1} \rangle = 0$ for all t. However, taking the ensemble average after squaring equation 4.4 and using $\langle O_t^2 \rangle = A^2 \langle \Omega_t^2 \rangle \langle \omega^2 \rangle$, we can obtain

$$\langle O_{t+1}^2 \rangle = (1 + 2\eta \mathcal{E} + \alpha \eta^2 \mathcal{E}^2) \langle O_t^2 \rangle \tag{4.5}$$

where

$$E := A \langle \omega^2 \rangle = \left\langle \sum_{\mathbf{r}} \mathbf{v} \cdot \mathbf{v} \right\rangle$$
$$\alpha := \langle \omega^4 \rangle / \langle \omega^2 \rangle^2$$

are constants. When ω is drawn from a gaussian distribution of zero

mean, for instance, $\alpha = 3$. Applying equation 4.5 iteratively yields $\langle O_{t+1}^2 \rangle = e^{t/\tau} \langle O_1^2 \rangle$, where constant

$$\tau := 1/\ln(1 + 2\eta\mathcal{E} + \alpha\eta^2\mathcal{E}^2)$$

Because $O_1 = \sum_{\mathbf{r}} \mathbf{w}_1 \cdot \mathbf{v}_1$ and $\mathbf{w}_1 = \mathbf{w}_0 + \eta O_0 \mathbf{v}_0 \approx \eta O_0 \mathbf{v}_0 = \eta(\sum_{\mathbf{r}} \mathbf{w}_0 \cdot \mathbf{v}_0)\mathbf{v}_0$, by similar arguments as above we can get $\langle O_1^2 \rangle = \alpha\eta^2\mathcal{E}^2 \langle O_0^2 \rangle$. Hence

$$\langle O_t^2 \rangle = \alpha\eta^2\mathcal{E}^2 e^{(t-1)/\tau} \langle O_0^2 \rangle \tag{4.6}$$

Consequently, when many MST-like units develop in parallel starting from random initial weights, the responses (either positive or negative) to rotation and to dilation/contraction are expected to grow exponentially in the initial stage of development. The variety of the initial responses leads to a continuous spectrum of selectivity to rotation and dilation/contraction, which is what has actually been found in the neurophysiological experiments (Duffy and Wurtz 1991a; Andersen *et al.* 1991).

5 Discussion

The model provides a unified, albeit simplified, account for several essential properties of MST neurons and how they might develop. These properties include selectivity to rotation, dilation, and contraction, the position independence of the responses (Saito *et al.* 1986; Tanaka and Saito 1989; Tanaka *et al.* 1989; Duffy and Wurtz 1991a,b), the selectivity to spiral velocity fields (Graziano *et al.* 1990; Andersen *et al.* 1991), and the continuous spectrum of selectivity (Duffy and Wurtz 1991a; Andersen *et al.* 1991). The model's response saturates at higher speeds (as a result of the sigmoid function) as does the response of real neurons (Orban *et al.* 1992).

In addition to rotation and dilation/contraction, shear also naturally arises in the optic flow (Koenderink 1986). Since the linear combination of shear fields is still a shear, according to equation 3.8 the weight vector field itself will also have a shear component. Consistent with the model, neurons selective to shear components were also found in the cortical areas including MST (Lagae *et al.* 1991).

This model differs somewhat from the original model in Sereno and Sereno (1991) and from real MST neurons in that it "linearly decomposes" the velocity field—that is, an MST-like unit will respond exclusively to the, say, rotational component of a flow field, regardless of the magnitude of the radial component. Since a cosine tuning curve means that the input unit sees exactly the vector component of the local stimulus movement in the preferred (here rotational) direction, it leads to linear decomposition. With narrower tuning curves, the response of individual MST-like units provides more information about the exact composition of the flow field—for example, the extent to which it approximates a pure

rotation; nevertheless, approximate position independence with narrower tuning curves is still explained by a direction-template mechanism like that described above.

Roughly speaking, learning with a simple Hebb rule tends to maximize the total response by gradient ascent and thus tune the net to the input patterns that frequently occur. Consider the output

$$O = \sigma(I) = \sigma \left(\sum_i w_i I_i \right)$$

The Hebb rule

$$\Delta w_i \propto I_i O$$

is always of the same sign as the gradient of the function $E = \frac{1}{2}O^2$:

$$\frac{\partial E}{\partial w_i} = O \frac{\partial O}{\partial w_i} = I_i O \sigma'(I)$$

because the derivative σ' is always positive. As a consequence, there should be a general tendency for local direction selectivity to be aligned with the direction of the stimulus velocity.

Recently, it was demonstrated that although dilation-sensitive MSTd neurons are basically position invariant in their responses, they often respond best to dilations centered at a particular location in the receptive field (often not the receptive field center) (Duffy and Wurtz 1991c).Similar results were obtained in the simulations in Sereno and Sereno (1991) using MT-like (narrower) input-layer tuning curves. It may be advantageous to retain information about *combinations* of flow field components--- here, dilation and translation—in single units since these combinations can have particular behavioral relevance—for example, in signaling direction of heading (Perrone 1992). More realistic peaked (instead of linear) speed tuning curves (Maunsell and Van Essen 1983)in the MT-like input layer could potentially sharpen the response to particular flow components since local speeds may be changed from the optimum as flow field components are added. Cross-direction inhibition (known to occur in MT: Snowden et al. 1991) could also be incorporated, effectively deleting portions of the flow field containing conflicting local motion signals. This could improve performance with more complex, real-world motion arrays.

The rotation, dilation, and contraction velocity fields required for training are readily produced when an animal is moving around in a rigid environment. Exposure to such velocity fields may be crucial in order for a young animal to develop rotation and dilation cells in its visual system. Human babies, for instance, can distinguish **a** rotation field from a random velocity field only after several months of visual experience (Spitz *et al.* 1988). This could be tested by recording from MST in infant monkeys.

Feedforward networks using Hebb rules have been shown to be capable of producing detectors selective to a hierarchy of features like those found in the successive stages of visual processing: center-surround units like those in the LGN (Linsker 1986a), orientation-selective units like simple cells in V1 (Linsker 1986b), pattern motion units like some cells in MT (Sereno 1989), and finally position-independent rotation and dilation units like cells in dorsal MST (Sereno and Sereno 1991). The visual system may use simple local learning rules and a richly textured environment to build up complex filters in stages. This strategy could drastically reduce the amount of supervision that is required later on (cf. Geman *et al.* 1992) as the visual system learns to recognize objects and direct navigation and manipulation.

Note Added in Proof

Recently, Gallant *et al.* (1993) found that neurons in V4 respond selectively, and in a position-invariant way to static patterns containing concentric, radiating, shearing, or spiral contours. The main outlines of our analysis could be extended to explain the selectivity and development of these neurons by substituting an orientation-selective input layer for the direction-selective input layer considered here.

Acknowledgments

References

Andersen, R., Graziano, M., and Snowden, R. 1991. Selectivity of area MST neurons for expansion/contraction and rotation motions. *Invest. Ophthal. Vis. Sci., Abstr.* **32,** 823.

Duffy, C. J., and Wurtz, R. H. 1991a. Sensitivity of MST neurons to optic flow stimuli. I. **A** continuum of response selectivity to large-field stimuli. *J. Neurophysiol.* **65,** 1329–1345.

Duffy, C. J., and Wurtz, R. H. 1991b. Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *J. Neurophysiol.* **65,** 1346–1359.

Duffy, C. J., and Wurtz, R. H. 1991c. MSTd neuronal sensitivity to heading of motion in optic flow fields. *Soc. Neurosci. Abstr.* **17,** 441.

Felleman, D., and Van Essen, D. C. 1991. Distributed hierarchical processing in primate cerebral cortex. *Cerebral Cortex* **1,** 147.

Gallant, J. L., Braun, J., and Van Essen, D. C. 1993. Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* **259,** 100-103.

Geman, S., Bienenstock, E., and Doursat, R. 1992. Neural networks and the bias/variance dilemma. *Neural Comp.* **4,** 1–58.

Graziano, M. S. A., Andersen, R. A., and Snowden, R. 1990. Stimulus selectivity of neurons in macaque MST. *Soc. Neurosci. Abstr.* **16,** 7.

Koenderink, J. J. 1986. Optic flow. *Vision Res.* **26,** 161–180.

Koenderink, J. J., and van Doorn, A. J. 1975. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Opt. Acta* **22,** 773–791.

Koenderink, J. J., and van Doorn, A. J. 1976. Local structure of movement parallax of the plane. *J. Opt. Soc. Am.* **66,** 717–723.

Lagae, L., Xiao, D., Ralguel, S., Maes, H., and Orban, G. A. 1991. Position invariance of optic flow component selectivity differentiates monkey MST and FST cells from MT cells. *Invest. Ophthal. Vis. Sci., Abstr.* **32,** 823.

Linsker, R. 1986a. From basic network principles to neural architecture: emergence of spatial-opponent cells. *Proc. Natl. Acad. Sci. U.S.A.* **83,** 7508–7512.

Linsker, R. 1986b. From basic network principles to neural architecture: emergence of orientation-selective cells. *Proc. Natl. Acad. Sci. U.S.A.* **83,** 8390-8394.

Longuet-Higgins, H. C., and Prazdny, K. 1980. The interpretation of a moving retinal image. *Proc. R. Soc. London B* **208,** 385–397.

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

Movshon, J. **A.,** Adelson, E. H., Gizzi, M. S., and Newsome, W. T. 1985. Analysis of moving visual patterns. In *Pattern Recognition Mechanisms,* C. Chagas, R. Gattass, and C. Gross, eds., pp. 117-151. Springer-Verlag, New York.

Orban, G. A., Lagae, L., Verri, **A.,** Raiguel, S., Xiao, D., Maes, H., and Torre, V. 1992. First-order analysis of optical flow in monkey brain. *Proc. Natl. Acad. Sci. U.S.A.* **89,** 2595-2599.

Perrone, J. A. 1992. Model for the computation of self-motion in biological systems. *J. Opt. Soc. Am. A* **9,** 177–194.

Poggio, T., Verri, A., and Torre, V. 1990. *Does cortical area MST know Green theorems?* Instituto per la Ricerca Scientifica e Technologica Tech. Rep. No. 9008-07, 1–8.

Poggio, T., Verri, **A.,** and Torre, V. 1991. Green theorems and qualitative properties of the optical flow. MIT *A.I. Memo,* No. 1289, 1–6.

Rodman, H. R., and Albright, T. D. 1987. Coding of visual stimulus velocity in area MT of the macaque. *Vision Res.* **27,** 2035–2048.

Saito, H., Yukie, M., Tanaka, K., Hikosaka, K., Fukada, Y., and Iwai, E. 1986. Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *J. Neurosci.* **6,** 145-157.

Sakata, H., Shibutani, H., Ito, Y., and Tsurugai, K. 1986. Parietal cortical neurons responding to rotary movement of visual stimulus in space. *Exp. Brain Res.* **61,** 658-663.

Sereno, M. I. 1989. Learning the solution to the aperture problem for pattern

motion with a Hebb rule. In *Advances in Neural Information Processing System I,* D. S. Touretzky, ed., pp. 468-476. Morgan Kaufmann, San Mateo, CA.

Sereno, M. I., and Allman, J. M. 1991. Cortical visual areas in mammals. In *The Neural Basis of Visual Function,* A. G. Leventhal, ed., pp. 160-172. Macmillan, London.

Sereno, M. I., and Sereno, M. E. 1990. Learning to discriminate senses of rotation and dilation with a Hebb rule. *Invest. Ophthal. Vis. Sci., Abstr.* **31,** 528.

Sereno, M. I., and Sereno, M. E. 1991. Learning to see rotation and dilation with a Hebb rule. In *Advances in Neural Information Processing Systems 3,* R. P. Lippmann, J. Moody, and D. S. Touretzky, eds., pp. 320–326. Morgan Kaufmann, San Mateo, CA.

Snowden, R. J., Treue, S., Erickson, R. G., and Andersen, R. A. 1991. The response of area MT and V1 neurons to transparent motion. *J. Neurosci.* **11,** 2768–2785.

Spitz, R. V., Stiles-Davis, J., and Siegel, R. M. 1988. Infant perception of rotation from rigid structure-from-motion displays. *Soc. Neurosci. Abstr.* **14,** 1244.

Tanaka, K., and Saito, H.-A. 1989. Analysis of motion of 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. *J. Neurophysiol.* **62,** 626–641.

Tanaka, K., Hikosaka, K., Saito, H.-A., Yukie, M., Fukada, Y., and Iwai, E. 1986. Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *J. Neurosci.* **6,** 134–144.

Tanaka, K., Fukada, Y., and Saito, H.-A. 1989. Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.* **62,** 642-656.