Available online at www.sciencedirect.com
SCIENCE DIRECT
Neural Networks
Neural Networks 17 (2004) 809–821
www.elsevier.com/locate/neunet

2004 Special Issue

Integration of form and motion within a generative model of visual cortex

Paul Sajda*, Kyungim Baek

Department of Biomedical Engineering, Columbia University, New York, NY 10027, USA

Received 31 March 2004; accepted 31 March 2004

Abstract

One of the challenges faced by the visual system is integrating cues within and across processing streams for inferring scene properties and structure. This is particularly apparent in the inference of object motion, where psychophysical experiments have shown that integration of motion signals, distributed across space, must also be integrated with form cues. This has led several to conclude that there exist mechanisms which enable form cues to 'veto' or completely suppress ambiguous motion signals. We describe a probabilistic approach which uses a generative network model for integrating form and motion cues using the machinery of belief propagation and Bayesian inference. We show, using computer simulations, that motion integration can be mediated via a local, probabilistic representation of contour ownership, which we have previously termed 'direction of figure'. The uncertainty of this inferred form cue is used to modulate the covariance matrix of network nodes representing local motion estimates in the motion stream. We show with results for two sets of stimuli that the model does not completely suppress ambiguous cues, but instead integrates them in a way that is a function of their underlying uncertainty. The result is that the model can account for the continuum of bias seen for motion coherence and perceived object motion in psychophysical experiments.

Keywords: Form and motion integration; Generative model; Belief propagation; Aperture problem; Hypercolumn; Direction of figure; Contour ownership; Occlusion

1. Introduction

The classical view of information processing in visual cortex is that of a bottom-up process in a feed-forward hierarchy (Hubel & Wiesel, 1977). However bottom-up information that encodes physical properties of the sensory input is often insufficient, uncertain, and even ambiguous—for example consider the classic demonstrations of the Dalmatian dog (Thurston & Carraher, 1966) and Rubin's vase (Rubin, 1915). Psychophysical (Adelson, 1992; Driver & Spence, 1998), anatomical (Budd, 1998; Callaway, 1998) and physiological (Bullier, Hupé, James, & Girard, 1996; Martinez-Conde et al., 1999) evidence suggests that integration of bottom-up and top-down processes plays a crucial role in the processing of the sensory input. For example top-down factors, such as attention, can result in strong modulation of neural responses as early as primary visual cortex (V1) (McAdams & Read, 2003). Information also flows laterally between populations of cortical neurons within the same level of the processing hierarchy. This is due to the 'generalized aperture problem'[1] with which the visual system is confronted. An individual neuron or local population of neurons 'sees' only a limited patch of the visual field. To form coherent representations of objects, non-local informational dependencies, and their uncertainties, must be integrated across space and time, as well as other stimulus and representational dimensions.

One particularly striking example illustrating the nature of the visual integration problem is that of inferring object motion in a scene. The motion of a homogeneous contour (or edge) is perceptually ambiguous because of the 'aperture problem'—i.e. a single local measurement along an object's bounding contour cannot be used to reliably infer an object's motion. However, this ambiguity can be potentially overcome by measuring locally unambiguous motion signals, tied to specific visual features, and then integrating these to form a global motion percept. An early study by Adelson (Adelson & Movshon, 1982) has suggested that

* Corresponding author. Tel.: +1-212-854-5279.
E-mail address: ps629@columbia.edu (P. Sajda).

[1] We use the term 'generalized aperture problem' to distinguish it from the term 'aperture problem' which is traditionally associated with local motion estimation and which we will discuss in more detail below.

0893-6080/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2004.03.013

Nomenclature

b	local belief
c	random variable for observed cue (network model) or normalizing/dividing constant (equations)
$i;j;k$	indices of the nodes
x	hidden variable
y	observed variable
z	variable for prior input
E	local interaction between hidden and observed nodes
M	message passed from a hidden node to a neighboring hidden node
N	a set of neighboring nodes
T	compatibility/transition matrix between neighboring nodes
w	weight for different form cues
f	weight function modulating local covariance in the motion stream
α	modulation variable for exponent of the weight function
$r_{\max}$	maximum value of the weight function
v	velocity
Cov	covariance matrix for local velocity
μ	mean vector for local velocity
Σ	inverse covariance matrix for local velocity

these two stages—local motion measurement and integration—are indeed involved in visual motion perception.

There are several visual features that have been identified as being unambiguous local motion cues. Examples of such features are line terminators and junctions. Line terminators have been traditionally classified into two different types: *intrinsic terminators*, that are due to the natural end of a line, and *extrinsic terminators*, that are not created by the end of a line itself but rather a result of occlusion by another surface (Nakayama, Shimojo, & Silverman, 1989). Intrinsic terminators are claimed to provide an unambiguous signal for the true velocity of the line, while extrinsic terminators generate a locally ambiguous signal which presumably should have little influence for accurate motion estimation (Shimojo, Silverman, & Nakayama, 1989). One problem is that there is ambiguity in all local measurements and therefore it is not a simple matter of determining which of the motions are ambiguous or unambiguous but the degree of ambiguity—i.e. degree of certainty of the cue.

In this paper we describe a generative network model for integrating form and motion cues which directly exploits the uncertainty in these cues. Generative models are probabilistic models which directly model the distribution of some set of observations and underlying hidden variables or states (Jebara, 2004). The advantages of a generative model are that (1) it uses probabilities to represent the 'state of the world' and therefore directly exploits uncertainties associated with noise and ambiguity, (2) through the use of Bayesian machinery, one can infer the underlying state of hidden variables, (3) it naturally enables integration via multiple sources, including those arising from bottom-up, top-down, and lateral inputs as well as those arising from other streams and (4) it results in a system capable of performing a variety of analysis functions including segmentation, classification, synthesis, compression, etc.

We describe a generative network model that accounts for interaction between form and motion in a relatively simple way, focusing on the influence of 'direction of figure' on local motion at junctions where line terminators are defined. A visual surface can be defined by associating an object's boundary to a region representing the object's surface. The basic problem in this surface assignment is to determine the contour ownership (Nakayama et al., 1989). We represent ownership using a local representation which we call the 'direction of figure' (DOF) (Sajda & Finkel, 1995). In our model the DOF at each point of the contour is represented by a hidden variable whose probability is inferred via integration of bottom-up, top-down and lateral inputs. The 'belief' in the DOF is used to estimate occlusion boundaries between surfaces that are defined by where the DOF changes—the ownership junction (Finkel & Sajda, 1994). In the model, the probability of extrinsic line terminators is a function of the probability of these ownership junctions. Thus rather than completely suppressing the motion signals at the extrinsic terminators, the degree of certainty (i.e. belief) in the DOF is considered as the strength of the evidence for surface occlusion and used to determine the strength of local motion suppression.

The remainder of the paper is organized as follows. Section 2 describes our generative network model, putting it in the context of the organizational structure of visual cortex. Though the model we describe does not use biologically realistic units (e.g. conductance based integrate-and-fire neurons) it is instructive to consider how the generative model maps to the cortical architecture. We next describe the details of the integration process between the form and motion streams, a process that exploits informational uncertainty. We describe how this uncertainty is propagated through the network using Bayesian machinery. We then present two sets of simulation results, illustrating the interaction of the form and motion streams. The first set of simulations shows how the model can generate results for motion coherence stimuli consistent with the psychophysical experiments of McDermott, Weiss, and Adelson (2001). We show how form cues change the model's inference of perceived motion, in particular a gradual transition from incoherent to coherent motion. We then demonstrate results

for the classic barber-pole stimulus (Wallach, 1935), showing how occlusion influences the certainty in the perceived object motion through form (DOF) cues.

2. Generative network model

2.1. Hypercolumns in the visual cortex

Since the term 'hypercolumn' was coined by Hubel and Wiesel (1977) it has been used to describe the neural machinery necessary to process a discrete region of the visual field. Typically, a hypercolumn occupies a cortical area of ~ 1 mm^2 and contains tens of thousands of neurons. Current experimental and physiological studies have revealed substantial complexity in neuronal response to multiple, simultaneous inputs, including contextual influence, as early as V1 (Gilbert, 1992; Kapadia, Ito, Gilbert, & Westheimer, 1995; Kapadia, Westheimer, & Gilbert, 2000). Such contextual influences, considered to be mediated by modulatory influences of the extraclassical receptive field, are presumed to arise from the lateral inputs from neighboring hypercolumns and/or from feedback from extrastriate areas (Bullier et al., 1996; Martinez-Conde et al., 1999; Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998; Polat & Norcia, 1996; Stettler, Das, Bennett, & Gilbert, 2002). It appears that in the visual system, information computed locally across the input space by individual hypercolumn is propagated within the network and integrated, perhaps in a modulatory way, to influence neural responses which ultimately correlate with perception.

The organization of columnar circuitry in the visual system has been extensively studied (Bosking, Zhang, Schofield, & Fitzpatrick, 1997; Crowley & Katz, 1999; Horton & Hocking, 1996; Hubel & Wiesel, 1977; Tsunoda, Yamane, Nishizaki, & Tanifuji, 2001). Callaway has proposed a generic model of vertical connectivity connecting layers within columns in primary visual cortex of cat and monkey (Callaway, 1998). Although the majority of connections are within columns, anatomical and physiological results indicate that there exist horizontal, long-range connections between sets of columns and that these connections give rise to complex, modulatory neuronal responses (Gilbert, 1992; Polat & Norcia, 1996; Stettler et al., 2002). In addition, a well-defined laminar structure has been identified with input and output projections to a column being a function of specific layers. Grossberg and colleagues have proposed that this laminar architecture is a key functional organizational principle of visual cortex and that particular neural circuits integrating information bottom-up, top-down and laterally may account for a wide variety of perceptual phenomena in static and dynamic vision (Grossberg & Williamson, 2001; Raizada & Grossberg, 2003).

The generative network model we construct is organized around the laminar hypercolumn structure of the visual cortex. In our model, a hypercolumn 'module' consists of a set of nodes representing random variables which are either observations to the column or hidden. The observations are limited by the effective aggregate field of the column—i.e. the aperture over visual space in which the hypercolumn constructs its bottom-up input. Though observations and representations are local, integration via belief propagation (described below) enables information to be passed more globally. Fig. 1 compares the generative network hypercolumn structure to that of a cortical hypercolumn.

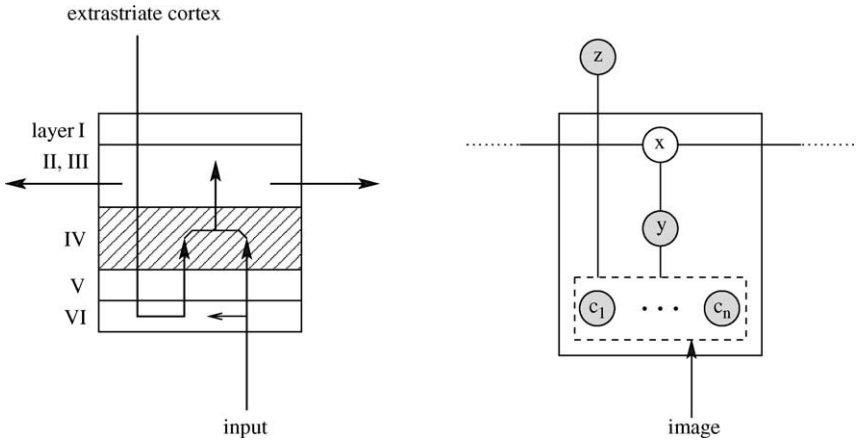

Fig. 1. A simplified diagram of cortical connections focusing on the laminar structure of a hypercolumn (left) and the isomorphic architecture used in our generative network model (right). For hypercolumns in primary visual cortex (V1) the LGN provides bottom-up input via strong connections to layer IV and indirect connections passing through layer VI. The top-down, feedback signals from extrastriate cortex pass into layer VI then project into layer IV. Feedforward connections from layer IV project to layer II/III, which forms intrinsic lateral, long-range connections between hypercolumns. In our generative network model cues corresponding to bottom-up input are represented as random variables c_i in a set of nodes. Those cues are combined, together with top-down prior knowledge, represented by the variable z, to form distributions for the observation variables y. Observations are passed to the nodes representing hidden variables x, corresponding to layers II/III in the hypercolumn structure. Spatial integration is performed by passing probabilities between neighboring hidden nodes.

Given a set of observations, including top-down and lateral inputs, inference is used to compute the most likely states of the hidden nodes. For example, in our model of the form stream, the hidden variables in each column are the DOF at each point along a contour. The DOF is inferred by considering the bottom-up cues, such as curvature and similarity, as well as top-down cues which can be interpreted as prior beliefs see (Baek & Sajda (2003) for details of the form generative network model). A critical feature of the network model is the use of *belief propagation* to transmit information regarding the states of the random variables, including their underlying uncertainties.

2.2. Belief propagation

In a generative network model a node represents a random variable and links specify the de-pendency relationships between variables (Jebara, 2004). The states of a random variable can be hidden in the sense that they are not directly observable. A hidden variable state is inferred from the states of other hidden variables and the available observations, through application of a local message passing algorithm called *belief propagation* (BP) (Pearl, 1988). In this section, a BP algorithm is described for undirected network models with pairwise potentials. It has been shown that most generative network models (i.e. graphical models) can be converted into this general form (Yedidia, Freeman, & Weiss, 2003).

Let x be a set of hidden variables and y a set of observed variables. The joint probability distribution of x given y is given by,

$$P(x_1, \ldots, x_n|y) = c \prod_{i,j} T_{ij}(x_i, x_j) \prod_i E_i(x_i, y_i) \tag{1}$$

where c is a normalization constant, x_i represents the state of the node positioned at i, $T_{ij}(x_i, x_j)$ captures the compatibility between neighboring nodes i and j, and $E_i(x_i, y_i)$ is the local interaction between the hidden and observed variables at location i. An approximate marginal probability of this joint probability at node i over all x_j other than x_i is called the local *belief*, $b(x_i)$.

The BP algorithm iterates local message passing and belief updates (Yedidia et al., 2003). The message $M_{ij}(x_j)$ passed from a hidden node i to its neighboring hidden node j represents the probability distribution over the state of x_j. In each iteration, messages and beliefs are updated as follows:

$$M_{ij}(x_j) = c \int_{x_i} \mathrm{d}x_i T_{ij}(x_i, x_j) E_i(x_i, y_i) \prod_{k \in N_i/j} M_{ki}(x_i) \tag{2}$$

$$b(x_j) = c E_j(x_j, y_j) \prod_{k \in N_j} M_{kj}(x_j) \tag{3}$$

where N_i/j denotes a set of neighboring nodes of i except node j. M_{ij} is computed by combining all messages received by node i from all neighbors except node j in the previous iteration and marginalizing over all possible states of x_i (Fig. 2). The current local belief is estimated by combining all incoming messages and the local observations.

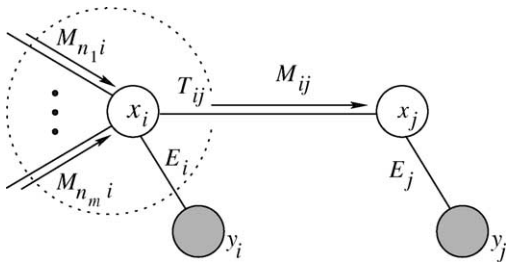

Fig. 2. Illustration of local message passing from node i to node j. Open circles are hidden variables, while shaded circles represent observed variables. The local belief at node j is computed by combining the incoming messages from all its neighbors and the local interaction E_j.

It has been shown that, for singly connected graphs (networks), BP converges to exact marginal probabilities (Yedidia et al., 2003). Although how it works for general graphs is not well understood, empirical results have demonstrated that BP can converge to correct solutions even with networks having loops (Freeman, Pasztor, & Carmichael, 2000).

2.3. Generative network model for form and motion streams

Fig. 3 shows the generative network model used for estimating perceived motion through integration of form information. The model consists of two structurally identical but functionally different processing streams for form and motion. The form and motion properties are inferred from local observations using BP for integrating local measurements across space. In the current model, the two streams interact unidirectionally so that influence flows from the form stream to motion stream only, i.e. form constrains motion perception.

The network model first infers DOF in the form stream. As described in detail in Baek and Sajda (2003), the local interaction between hidden variable x_i^{dof} and observed variable y_i^{dof} is specified by local figure convexity and similarity/proximity cues. The local convexity is determined by the local angle of the contour at a given location, while the similarity/proximity cues are estimated by considering points having similar local orientation that lie in orthogonal directions to the contour. These cues are computed separately and combined in a weighted linear fashion to form the total local interaction. When prior information is available, it is multiplied with this local interaction.[2] The hidden variable x_i^{dof} has two possible states, specifying the DOF relative to the direction of local convexity. When the form stream converges, both the direction of figure as well as the certainty

[2] Note that in our previous work (Baek & Sajda, 2003), prior information was treated as an additional cue and combined as a weighted linear sum. The current mechanism of multiplying the prior with the local likelihood results in a more localized posterior which is also more consistent with a Bayesian framework. Examples shown in Fig. 4 were generated using this multiplication scheme.

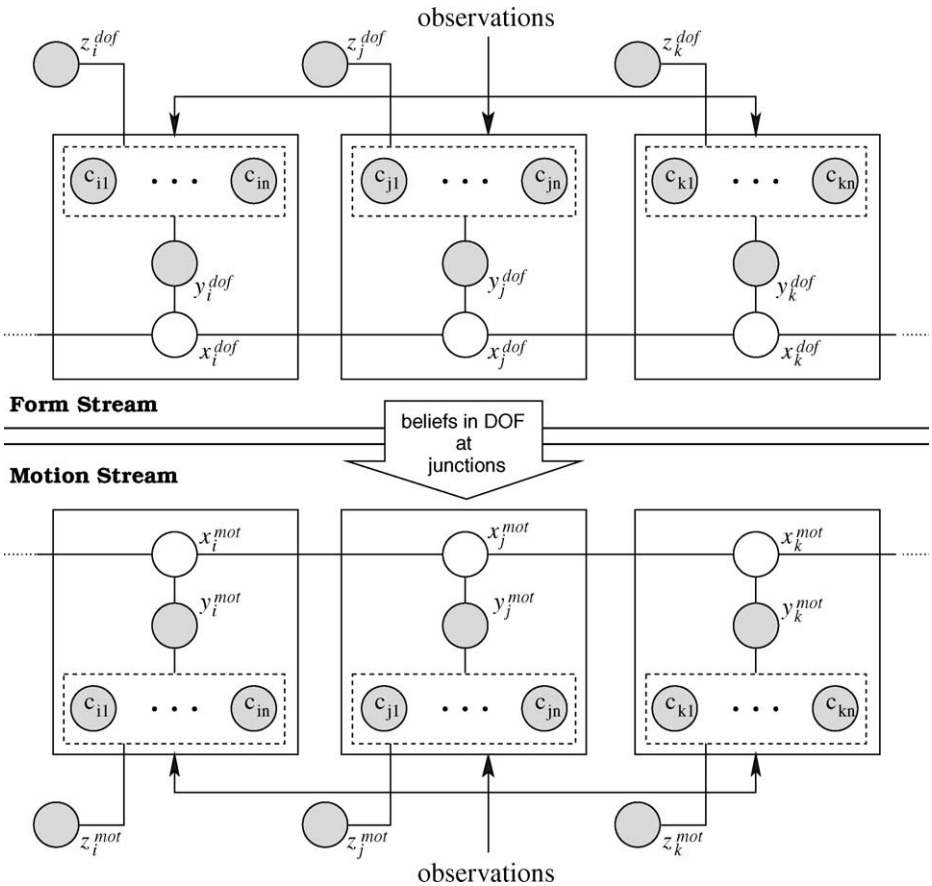

Fig. 3. Generative network model for form and motion streams. Each stream is modeled by a network of hypercolumn nodes which integrate inputs via bottom-up, top-down, and lateral connections. The DOF probability computed in the form stream integrates into the motion stream by modulating the covariances of distributions used to model the velocity at apertures located at junction points. In the current model, the interaction between form and motion streams is unidirectional.

(belief) of this direction is represented at each point along the contour as the state of the corresponding hidden node. The form stream has been applied to several examples, including perceptually ambiguous figures, with results that are in agreement with human perception (Baek & Sajda, 2003). Fig. 4 shows some examples of DOF estimates inferred by the form stream.

A hidden node in the motion stream represents the velocity at the corresponding location along the contour. We assume a prior that specifies a preference for slower motions, modeled using a symmetric Gaussian centered at the origin. Both the pairwise compatibility $T_{ij}(x_i^{\text{mot}}, x_j^{\text{mot}})$ and the local interaction $E_i(x_i^{\text{mot}}, y_i^{\text{mot}})$ that model the velocity likelihood at the apertures are also Gaussian. Currently, T_{ij} is set manually and E_i is defined by the mean of the normal velocity at point i and a local covariance matrix Cov_i. Before the BP algorithm begins, the variance at junction points is modulated by a function of the DOF belief $b(x_i^{\text{dof}})$:

$$Cov_i = f(b(x_i^{\text{dof}}),\ \alpha,\ r_{\max}) \cdot Cov_i \tag{4}$$

$$= (c \cdot (e^{\alpha\{b(x_i^{\text{dof}})-0.5\}} - 1) + 1) \cdot Cov_i \tag{5}$$

The weighting function $f(b(x_i^{\text{dof}}); \alpha, r_{\max})$ is an exponential defined on $[0, 0.5\alpha]$ and ranges from 1 to $r_{\max}$. c is defined by the maximum weight $r_{\max}$ as $c = (r_{\max} - 1)/(e^{0.5\alpha} - 1)$.

Initially, the covariance matrices of hidden variables are set to represent infinite uncertainty, and mean vectors are set to zero. Local motions are then propagated and integrated across space. When the BP algorithm converges, the global motion percept is estimated by a mixture of Gaussians:

$$p(v|y) = c \sum_i p_i(v|y) p(i) \tag{6}$$

where $p_i(v|y)$ is the probability distribution of velocity v at location i resulting from the Gaussian of hidden variable x_i^{mot}, $p(i)$ are the mixing coefficients, and c is the normalization constant. The peak location of distribution $p(v|y)$ is the maximum *a posteriori* (MAP) estimate of velocity, which is the most probable motion given the observations. In addition, as described in the following sections, the posterior distribution provides a means for interpreting perceived motion coherence.

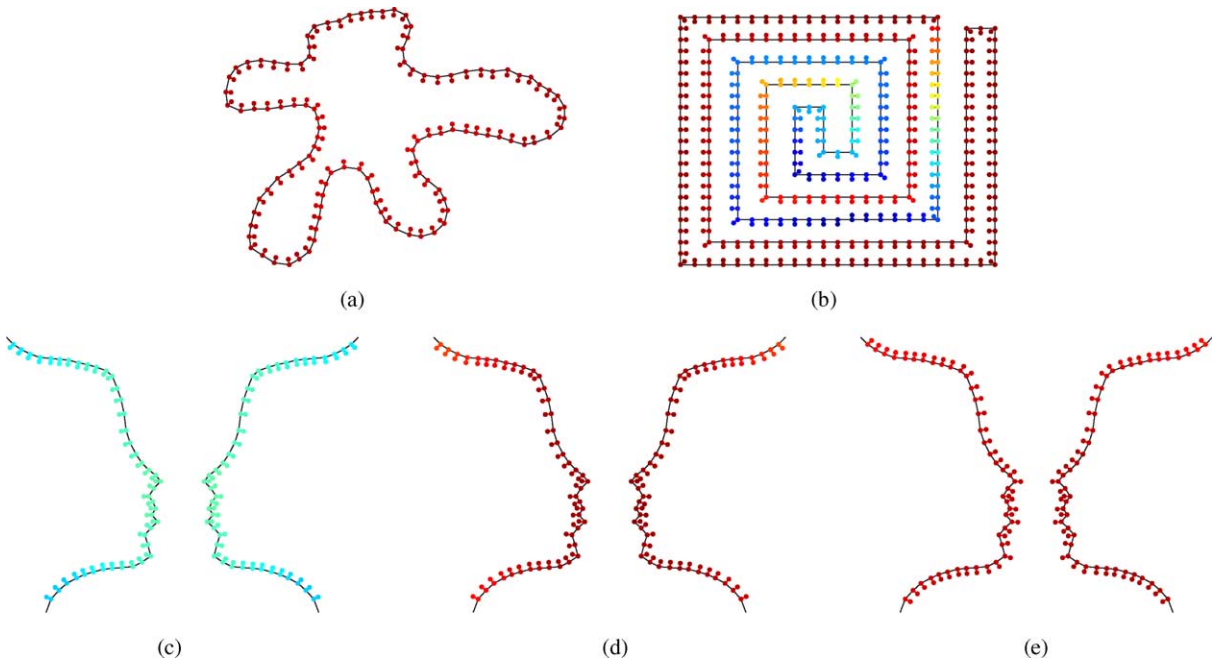

Fig. 4. Examples of DOF estimates by the form stream. 'Dumbbell' bars indicate the DOF, as they point to the surface locally owned by the contour. Degree of certainty (belief) in the DOF increases as the color changes (shown in web version) from blue ($P = 0.5$) to red ($P = 1.0$). Note that since the assignment of DOF is binary, $P = 0.5$ implies complete uncertainty in the assignment. (a) DOF estimate for arbitrary shaped figure. (b) A spiral figure in which figure-ground discrimination is ambiguous unless one serially traces the contour. The low certainty (blue) for the central region of the spiral indicates ambiguity of DOF, consistent with human perception. Figures (c)–(e) show the perceptual shift induced by prior information. In the form stream, prior information is multiplied with the local likelihood computed by combining interactions from different local cues. (c) DOF estimate for Rubin's vase without using prior information. (d) DOF estimate biased toward face figures by the prior cue for face features (nose, chin, brow). (e) DOF estimate shifts to the vase figure via a change in prior cues for vase figures (base, top, center cup).

3. Simulation results

All simulations were done with the same network architecture and parameter values, except the covariance of the motion prior. See Appendix for parameter values used in the simulations.[3]

3.1. Circular motion of a square modulated by occluders

The first row in Fig. 5 shows the four stimuli used for this experiment. They are 90° rotated versions of the diamond stimuli described in McDermott et al. (2001), but the perceptual effect is basically the same. The motion of moving line segments is identical in all four stimuli. A pair of vertical line segments move together sinusoidally in a horizontal direction while the two horizontal line segments move simultaneously in a vertical direction. The vertical and horizontal motions are 90° out of phase. The difference between the stimuli is the degree of 'completeness' of the L-shaped occluders, which alters the perceived motion. When there are no occluders as in (a), we are more likely to see two separate motions of the line segments, while the perceived motion coherence of a square rotating behind the occluders increases as the occluders' completeness increases (McDermott et al., 2001). The colored 'dumbbell' bars at the junctions between the occluders and moving bars represents the DOF inferred by the form stream. The belief in DOF increases, with color changing from blue to red, because additional form cues (convexity cues) are observed for the more 'complete' occluding figures.

The second row in Fig. 5 shows the resulting velocity estimates by the network model for the four stimuli. For stimulus (a), all line terminators are intrinsic, hence there is no modulation of local motion signals by DOF. However, the presence of occluders in stimuli (b), (c), and (d) classifies the line terminators as extrinsic and the local motion signals are suppressed. The strength of the suppression increases and becomes highest for stimulus (d) since the belief in DOF at the junctions becomes stronger as the degree of completion of the occluders increases. In the first two figures, we clearly see the bimodal distribution which implies that we perceive two separate motions for stimuli (a) and (b). The last two figures show that a single peak is formed at the intersection of two distributions for stimuli (c) and (d), which indicates a stronger perception of single coherent motion compared to stimuli (a) and (b).

Fig. 6 shows the ratio of the peak value at the intersection of the horizontal and vertical motion distributions normalized by the maximum value of horizontal motion for the four stimuli in Fig. 5. The values are scaled relative to the ratio for stimulus (d). This figure can be interpreted as

[3] For Flash demos of the types of stimuli used in the simulations, see http://web.mit.edu/persci/demos/Motion&Form/master.html

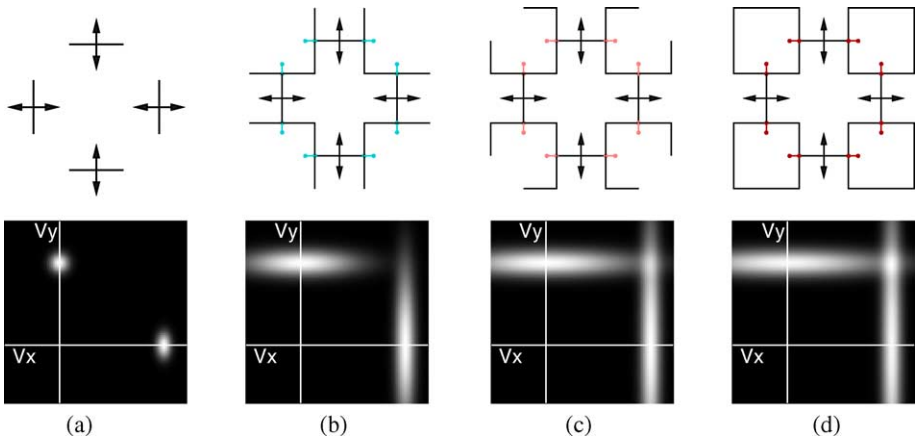

Fig. 5. First row shows stimuli generated from four line segments that move sinusoidally, 90° out of phase, in vertical and horizontal directions. The line segments are presented without occluders (a) and with occluders, with an increasing degree of completeness (b)–(d). The colored dumbbell bars at junctions show the DOF inferred by the form stream. The color of the bar (shown in web version) represents the strength of the beliefs in DOF which increases from blue ($P = 0.5$) to red ($P = 1.0$). The presence of the occluding surface alters the motion perception. Velocity estimation results for the stimulus presented in the second row were computed by combining the estimates at all locations using a mixture of Gaussians.

the model's estimate of the degree of motion coherence, which in fact represents a continuum as one moves from stimulus (a) to (d). Note that the difference in strength of the coherence for stimuli (c) and (d) is clearly seen in this figure. These results are qualitatively consistent and quantitatively similar to the psychophysical results presented in McDermott et al. (2001).

Fig. 7 illustrates the resulting velocity estimate for stimuli (b) and (d) for six successive frames sampled from a period of sinusoidal motion. On the top row, both horizontal and vertical motion are almost equally likely in each frame for stimulus (b) and the two separate motions oscillate in the direction normal to the line segment orientation. On the other hand, for stimulus (d) shown at the bottom, the maximum of the distribution is at the intersection of the two motions in each frame and it follows a circular trajectory which is consistent with perceiving rotation.

A prior cue shown in Fig. 8 can be added for inferring DOF in the form stream. Psychophysically, this can simulate priming subjects or using stereo to add depth information so that subjects are biased to see the occlusion more strongly. Adding priors to the form stream in the model strengthens the belief for DOF in the indicated direction, which increases local motion suppression, and consequently more coherent motion would be expected. Fig. 8 shows the results with weak and strong prior cues for stimulus (b). As the prior is made stronger, the estimated motion produces a single peak at the intersection, similar to stimulus (d) shown in Fig. 5.

3.2. Barber-pole motion modulated by occluders

One of the classic examples illustrating the aperture problem and the importance of occlusion cues in motion perception is the *barber-pole illusion* (Wallach, 1935). In its basic form, the barber-pole stimulus consists of a diagonal drifting grating viewed through a rectangular shaped aperture, with the grating being perceived as moving in the direction of elongated dimension of the aperture. The general explanation for the barber-pole illusion is that the unambiguous local motion signals at the terminators of gratings are integrated, disambiguating the local motion and thereby dominating the global motion percepts (Nakayama & Silverman, 1988). Since longer dimensions of the aperture provide more terminators that move along the axis of the dimension, the perceived motion is determined by the direction of that aperture dimension.

Fig. 9 shows the velocity estimates of the model for the barber-pole stimulus for several different aspect ratios. The velocity likelihood at terminators in the two aperture dimensions and along the gratings of the barber-pole stimulus are shown in Fig. 10. Since all terminators are intrinsic, there is no modulation of the local motion signals

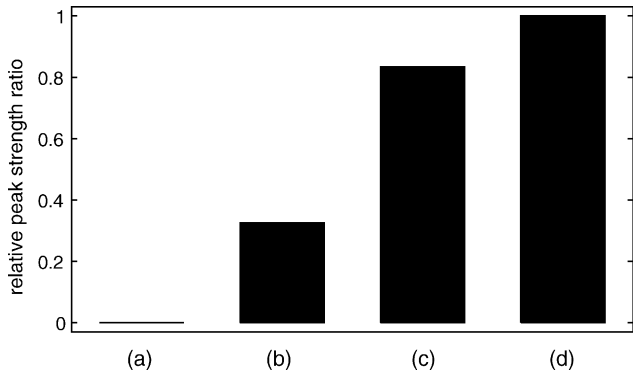

Fig. 6. The probability at the intersection of the distributions in velocity space, representing horizontal and vertical motions, relative to the maximum value of the horizontal motion distribution for stimuli (a)–(d) in Fig. 5. The values are scaled relative to the ratio for stimulus (d). This 'peak strength probability' ratio can be seen as the model's estimate of motion coherence, with values near 1 having high coherence and those near 0 having low coherence. Compare to the psychophysical results (Fig. 6) of McDermott, Weiss, and Adelson (2001).

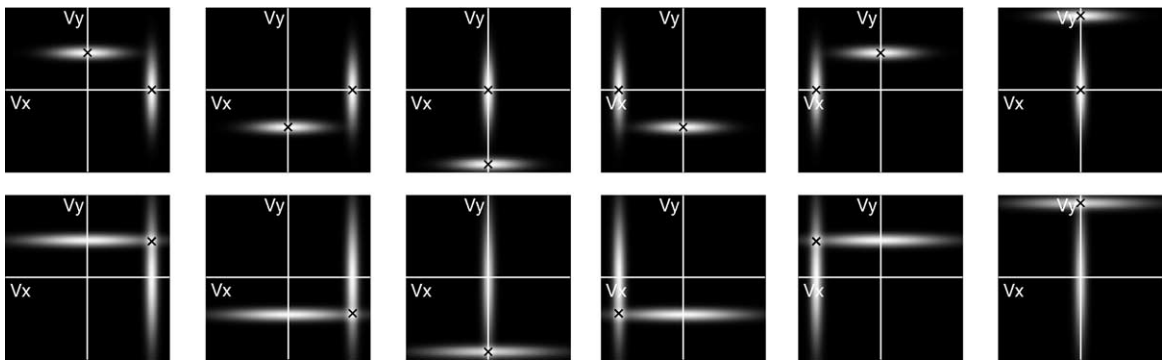

Fig. 7. The sequence of resulting velocity estimates for stimuli (b) and (d) in Fig. 5 for six successive frames sampled from a period of sinusoidal motion. First row shows results for stimulus (b), indicating two separate motions oscillating in the direction normal to the line segment orientation. The sequence in the second row for stimulus (d) shows a coherent motion forming a circular trajectory, indicating coherent rotating motion. 'X' indicates MAP location.

by the form stream. The resulting velocity estimates clearly illustrate that the dominant motion follows the direction of elongated dimension of the rectangular aperture, consistent with the explanation described above. These results can be seen more clearly in Fig. 14, which illustrates the continuum of horizontal motion bias as a function of different aspect ratios.

Variations of the classic barber-pole illusion have been developed for investigating the influence of occlusion in visual motion perception (Duncan, Albright, & Stoner, 2000; Lidén & Mingolla, 1998; Shimojo et al., 1989). Occlusion cues alter the intrinsic/extrinsic visual features and consequently change the perceived motion. Lidén and Mingolla (1998) performed a series of psychophysical experiments on a variety of barber-pole stimuli with occluding patches. They found that (1) monocular occlusion cues play a more critical role than binocular depth cues in motion perception and (2) the local motion signal at extrinsic terminators is not completely discounted.

We used stimuli similar to those used by Lidén and Mingolla (1998) to investigate the models response to different occlusion configurations. Fig. 11 shows these configurations, with occluders placed along either vertical or horizontal sides of the barber-pole. The difference with those used in Lidén and Mingolla (1998) is the absence of depth cues and texture on the occluders. The arrows indicate the perceived direction of barber-pole motion. Presumably, the grating terminators formed at the occluding boundary are considered extrinsic and therefore their motion signal is suppressed. Thus the perceived motion tends to be biased in the direction orthogonal to the occlusion boundary.

Fig. 12 shows the velocity estimates of the model for the three stimuli in Fig. 11. The stimulus with no occluders has a longer side in the horizontal direction so the resulting motion is horizontally biased, as already shown. The occluders placed along the top and bottom of the barber-pole causes horizontally moving terminators to have high probability of being extrinsic. The strong belief in the DOF at the occlusion boundary suppresses their influence in the motion stream. As a result, the intrinsic terminators moving vertically dominate the resulting estimated motion. In the same way, the model estimates horizontal motion when occluders are placed along the vertical sides of the barber-pole, with 'perceived' horizontal motion being even stronger than that in the first figure.

To see whether the influence of extrinsic terminators on the perceived motion is completely discounted or is just decreased, Lidén and Mingolla varied the aspect ratio of the barber-pole stimuli while keeping the size of the occluders fixed (Lidén and Mingolla, 1998). The resulting motion percepts show that the elongation of the barber-pole along the occlusion boundary still influences the perceived direction of motion. They argue that their results suggest that the influence of extrinsic terminators is attenuated rather than completely suppressed. They quantify the relative attenuation of the extrinsic terminators with respect to the intrinsic terminators based on the changes in perceived motion.

We generate similar stimuli as shown in the first and third row of Fig. 13. The aspect ratio was varied from 3:1 to 1:3 and the occluders were placed in a way that altering aspect ratio changes the number of extrinsic terminators. Below each stimulus is the velocity estimate of the model for the stimulus. Although the changes between different configurations may not be as clear as in the psychophysical results reported in Lidén and Mingolla (1998), the model estimates match the psychophysical results qualitatively (see Fig. 3 of Lidén and Mingolla (1998)). A summary of these results is shown in Fig. 14. These results indicate that the influence of

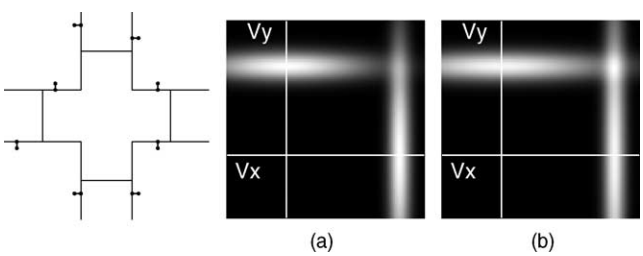

Fig. 8. Left figure is the prior cue for stimulus (b) in Fig. 5 used for inferring DOF in the form stream. Velocity estimation results with weak and strong prior cues are shown in (a) and (b), respectively.

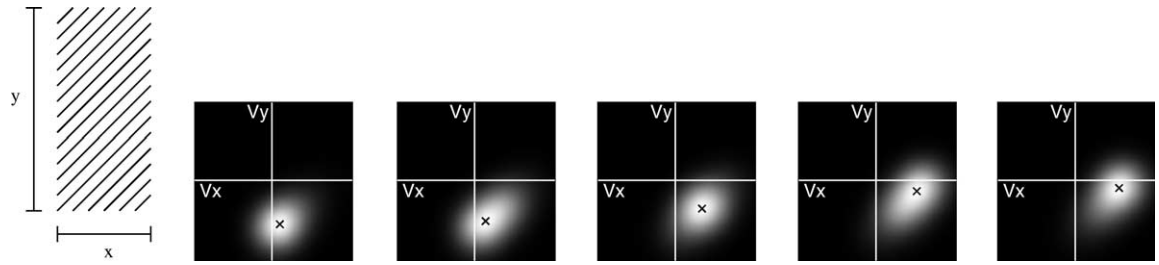

Fig. 9. The barber-pole illusion. In its classic form shown in the far left, a diagonal drifting grating viewed through a rectangular shaped aperture is perceived as moving in the direction of the elongated dimension of the aperture (Wallach, 1935). The prevailing explanation for this is that the unambiguous motion signals at line terminators are integrated to produce the perceived direction of motion. Since there are more line terminators along the elongated dimension of the aperture, they have stronger influence. Second through the last figures are the estimated velocity by the model for different barber-pole aspect ratios.

extrinsic terminators is not abolished in the model, rather it is attenuated resulting in a continuum of horizontal motion bias, consistent with the results from human subjects. The quantitative differences between psychophysical results and our results may be due to stronger suppression by the form stream in our model.

4. Discussion

In this paper we describe a generative network model for integrating form and motion cues. The model can account for a number of perceptual phenomena related to how form information is used to distinguish between intrinsic and extrinsic terminators in the motion integration process. Previous neural network models on segmentation and integration of motion signals have studied the influence of motion signals at terminators and occlusion cues (Grossberg, Mingolla, & Viswanathan, 2001; Lidén & Pack, 1999). The advantage of our model over these previous models is that uncertainties in the underlying observations and inferred representations are used in characterizing terminators as intrinsic vs. extrinsic and therefore these terminators do not fall into one class or the other but instead have some smooth transition between the two classes. A second, more general advantage of our model is that it provides a natural way to integrate different data types (top-down, bottom-up), since all variables are mapped to probability space. The advantage of the previous models, however, is that they are perhaps more neural and thus more closely related to the underlying biological circuitry. In addition, our model does not address which visual cortical areas are best suited for this specific type of cortical integration of form and motion. Nonetheless, we have attempted to demonstrate that the architecture of our model does in fact exploit several of the same organizational principles seen in visual cortex. The laminar hypercolumn architecture of the model, together with the relatively local connectivity, are all reasonable constraints which make such an architecture biologically feasible. In addition, Weiss points out that the updating rules used in Bayesian networks require three building blocks: a distributed representation, weighted linear summation of probabilities, and normalization (Weiss, 2000). These processes have clear ties to cortical networks (Heeger, Simoncelli, & Movshon, 1996).

Many vision researchers have adopted a Bayesian or probabilistic approach to vision. Kersten and Schrater (2000) describe vision within the context of the Bayesian principles of 'least commitment' and 'modularity'. Geisler and Diehl (2002) describe a Bayesian based selection framework which generally applies to the evolution of all perceptual systems. Lee and Mumford (2003) begin to address network and architectural issues, by considering hierarchical Bayesian inference within visual cortex, relating a general hierarchical model to neurophysiological experimental data. Zemel (2004) has described 'cortical belief networks' as a model for orientation tuning in V1 to motion discrimination in MT. Though our approach is consistent with these other theories/model, it differs in that we propose a specific model and architecture which attempts to account for how different occlusion relationships,

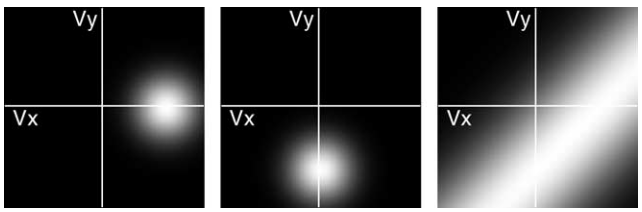

Fig. 10. Velocity likelihood for three apertures in the barber-pole stimulus with no occluders. (a) Likelihood at a line terminator along the horizontal side, (b) likelihood at a line terminator along the vertical side, and (c) likelihood at a location on a diagonal line segment.

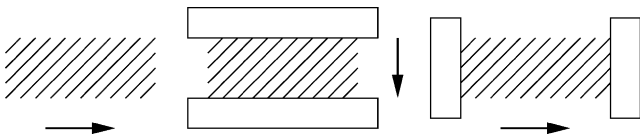

Fig. 11. Barber-pole with occlusion cues. The arrows indicate the perceived direction of motion of the barber-pole. *Left*: Barber-pole with no occluders. *Middle*: Occluders placed on the top and bottom cause the perceived motion to be mostly vertical. The line terminators aligned with the occluders are considered as extrinsic terminators, therefore their motion signals tend to become more ambiguous and have less influence on the perceived motion. *Right*: Occluders placed on the left and right sides of the barber-pole bias the perceived motion toward horizontal (Lidén and Mingolla, 1998).

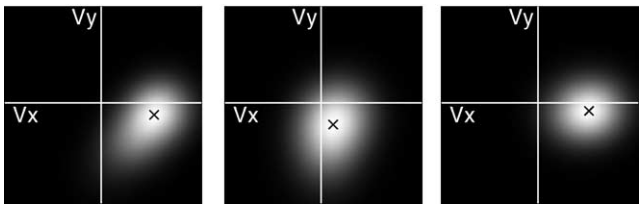

Fig. 12. Resulting velocity estimates for the three barber-pole stimuli shown in Fig. 11. These results are obtained by performing motion integration across all locations on the barber-pole. Left figure shows that the barber-pole stimulus with no occluders is perceived as moving horizontally. Occluders placed along the top and bottom of the barber-pole bias the perceived motion toward vertical (middle). Similarly, strong horizontal motion is perceived by placing occluders on both sides of the barber-pole (right). Note that the horizontal motion becomes stronger with occluders compared to the first figure (no occluders). 'X' indicates the MAP estimate.

represented via certainty in figure-ground representations (DOF), are integrated within the motion integration process and result in a continuum of perceived motion bias.

Since the initial studies on the statistical properties of population codes that first used Bayesian techniques to analyze them (Paradiso, 1988), researchers have suggested that it might be probability distributions instead of single values that a neural population encodes (Anderson & Van Essen, 1994; Pouget, Dayan, & Zemel, 2000; Zemel, Dayan, & Pouget, 1998). A computational model proposed by Deneve, Latham, and Pouget (1999) uses a biologically plausible recurrent network to decode population codes by maximum likelihood estimation. Therefore the network essentially implements optimal inference and the simulation results suggest that cortical areas may function as ideal observers. If it is true that information is represented as probability distributions in cortical areas, it means that the brain may perform Bayesian inference that effectively deals with uncertainty commonly arising in visual tasks. A recent modeling study by Rao (2004) has shown how integrate-and-fire neurons can be used to build networks for carrying out Bayesian inference. Our current work is focusing on using such integrate-and-fire models to build more realistic columnar networks.

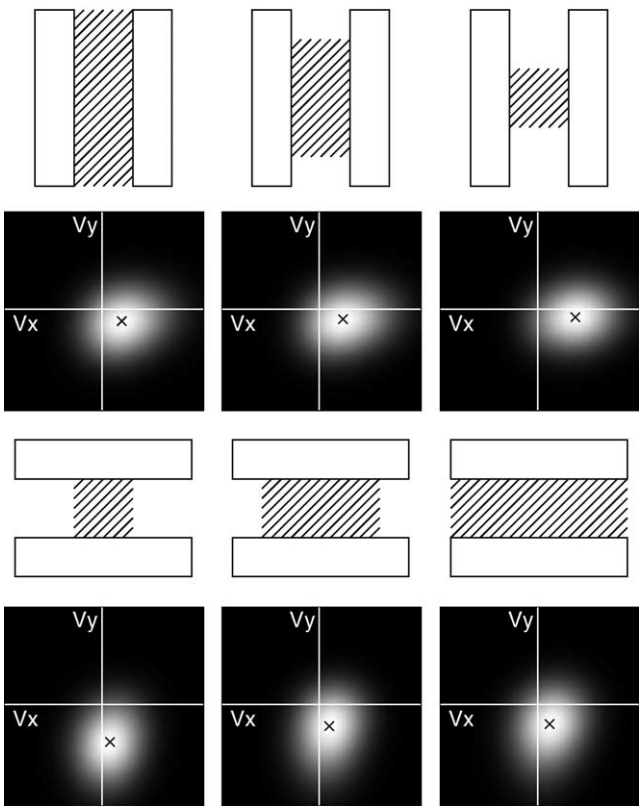

Fig. 13. Occluded barber-pole stimuli with varying aspect ratios and the velocity estimates of the model for each corresponding stimulus. From the top-left, aspect ratios of the barber-pole stimuli are 3:1, 2:1, 1:1, 1:1, 1:2, and 1:3, respectively. The number of extrinsic terminators created along the occlusion boundaries changes according to the aspect ratio. The model estimates of the velocity show an increase for the motion in the direction of the occlusion boundary as more extrinsic terminators are generated along the boundary. This results suggest that the local motion of the extrinsic terminators are not completely suppressed in the model, but instead suppressed relative to the certainty of DOF in the from stream.

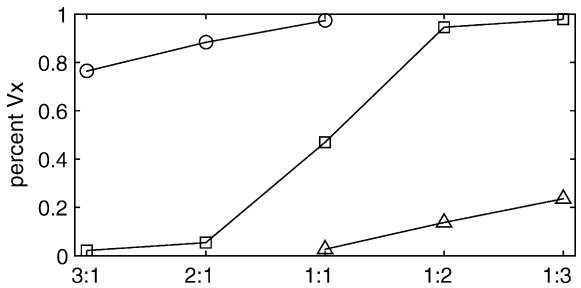

Fig. 14. The percent horizontal velocity of the model estimates is plotted against the aspect ratio for the barber-pole stimuli without occluders (line with squares) and for the six occluded configurations shown in Fig. 13 (line with circles for horizontal occluders and line with triangles for vertical occluders). The results for the unoccluded configuration matches well with the sigmoid estimation for the psychophysical response curve illustrated in Lidén and Mingolla (1998) while the degree of change in horizontal motion for occluded configurations is less dramatic. Compare to Fig. 3 of Lidén and Mingolla (1998).

Over the last several years, substantial progress has been made in generative network theory, probabilistic learning rules, and their application to image analysis (Portilla, Strela, Wainwright, & Simoncelli, 2002; Romberg, Choi, & Baraniuk, 2002; Spence, Parra, & Sajda, 2000; Wainwright & Simoncelli, 1999). Recent work by Hinton (Hinton & Brown, 2000) has been directed at trying to understand the relationship between Bayesian and neural processing. Application of this 'neuro-Bayesian' approach to image processing has been significantly advanced by Weiss (1997), who has demonstrated that when applied to difficult image processing problems, Bayesian networks converge orders of magnitude more rapidly than current relaxation-based algorithms. Processing time is only limited by the time (iterations) required for information to propagate between all units focused on the target. This is in line with David Marr's dictum (Marr, 1982) that visual processing should only take as long as required for all relevant information about the image to be transmitted across cortex,

and no further iterations should be necessary once the information has arrived at the appropriate hypercolumn. Thus Bayesian methods offer the hope of matching the time constraints posed by human visual recognition.

Ultimately a model's value is in the predictions it can generate. Our model predicts that contour ownership is represented as a probabilistic, local representation, and that the DOF serves as a substrate for motion integration. Our previous work has argued for DOF serving as such a substrate for perceptual integration (Finkel & Sajda, 1994), however, the probabilistic representation is an additional prediction from this current model. Though the concept of border ownership was first put forth by Nakayama et al. (1989), the DOF is a local vector representation which is ideally suited for neural coding, for example via a population vector scheme. Interestingly, von der Heydt and colleagues (Zhou, Friedman, & von der Heydt, 2000) have discovered that neurons in V2 appear to code for ownership, in fact through a local representation of the side of the contour that represents the figure. In addition, he finds firing rates of these neurons appear to be modulated in a continuous way, based on occlusion and other figure-ground cues, consistent with a probabilistic representation. The results of our model would predict that the strength of these neuronal firing rates in V2 would modulate the distribution of responses of motion selective neurons, perhaps in area MT, leading to a shift in the distribution of firing rates and resulting in a change in perceived motion.

Acknowledgements

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research under grant N00014-01-1-0625, and a grant from the National Imagery and Mapping Agency, NMA201-02-C0012.

Appendix A

The graphical structure of the form and motion streams in the current network model is an undirected chain. The equations for updating messages and beliefs in the model are described below.

A.1. Belief propagation in the form stream

In the form stream, local figure convexity and similarity/proximity cues are combined to form initial observations. The local interaction $E_{i,\mathrm{cvx}}$ between hidden variable x_i^{dof} (the superscript will be dropped for notational convenience) and observed variable $y_{i,\mathrm{cvx}}$ specified by the convexity at point i is determined by the local angle of the contour at the location. At the same time, the local interaction $E_{i,\mathrm{sim}}$ between x_i and $y_{i,\mathrm{sim}}$ is computed from the similarity/proximity cue by identifying points having a similar local tangent angle (i.e. orientation) that lie in a direction orthogonal to the contour at point i. $E_{i,\mathrm{cvx}}$ prefers smaller angles and $E_{i,\mathrm{sim}}$ favors shorter distances with similar orientations. These two local interactions are combined to compute the overall local interaction:

$$E_i(x_i, y_i) = w_{\mathrm{cvx}} E_{i,\mathrm{cvx}} + w_{\mathrm{sim}} E_{i,\mathrm{sim}} \tag{A1}$$

The hidden variable x_i represents a two dimensional binary DOF vector at location i along the boundary. The vector $x_i = (1,0)^T$ specifies that the DOF is in the direction of local convexity, while $x_i = (0,1)^T$ assigns the opposite direction to the DOF.

Since the graph is a chain, every node has two neighbors from which it receives messages. Furthermore, the hidden variables are discrete and have two possible states, with incoming messages and the belief at x_j in the form stream shown in Fig. 3 computed as follows:

$$M_{ij}(x_j) = c \sum_{x_i} T_{ij}(x_i, x_j) E_i(x_i, y_i) M_{hi}(x_i) \tag{A2}$$

$$M_{kj}(x_j) = c \sum_{x_k} T_{kj}(x_k, x_j) E_k(x_k, y_k) M_{lk}(x_k) \tag{A3}$$

$$b(x_j) = c E_j(x_j, y_j) M_{ij}(x_j) M_{kj}(x_j) \tag{A4}$$

where T_{ij} specifies the pairwise compatibility as described in Section 2.2, and the sum is over two possible states of the hidden variables. The vector multiplication is performed element by element. Each element of $b(x_j)$ represents the degree of confidence in the DOF for the corresponding direction at location j. When the algorithm converges the states of the hidden variables are determined by taking the direction having larger certainty.

A.2. Belief propagation in the motion stream

We assume a Gaussian generating process in the motion stream with the probabilities in velocity space represented by two dimensional mean vectors and 2×2 covariance matrices. The update rules for the parameters in the one dimensional Gaussian case are described in Weiss (1997).

Let μ_j and Σ_j be the mean and inverse covariance matrix defining the probability distribution of hidden variable x_j. Also, let μ_j^{α} be the mean passed along one direction of the chain and μ_j^{β} the mean passed along the opposite direction, with Σ_j^{α} and Σ_j^{β} being the corresponding inverse covariance matrices. Similarly, the mean and inverse covariance matrix passed from the local observation y_j to the hidden variable x_j are represented by μ_j^{l} and Σ_j^{l}. The parameters in the motion stream, shown in Fig. 3, are then updated as follows:

$$\mu_j = (\Sigma_j^{l} + \Sigma_j^{\alpha} + \Sigma_j^{\beta})^{-1} (\Sigma_j^{l} \mu_j^{l} + \Sigma_j^{\alpha} \mu_j^{\alpha} + \Sigma_j^{\beta} \mu_j^{\beta}) \tag{A5}$$

$$\Sigma_j = (\Sigma_j^{l} + \Sigma_j^{\alpha} + \Sigma_j^{\beta})^{-1} \tag{A6}$$

$$\mu_j^{\alpha} = (\Sigma_i^{\alpha} + \Sigma_i^{l})^{-1} (\Sigma_i^{\alpha} \mu_i^{\alpha} + \Sigma_i^{l} \mu_i^{l}) \tag{A7}$$

$$\Sigma_j^{\alpha} = (C + (\Sigma_i^{\alpha} + \Sigma_i^{l})^{-1})^{-1} \quad \text{(A8)}$$

$$\mu_j^{\beta} = (\Sigma_k^{\beta} + \Sigma_k^{l})^{-1}(\Sigma_k^{\beta}\mu_k^{\beta} + \Sigma_k^{l}\mu_k^{l}) \quad \text{(A9)}$$

$$\Sigma_j^{\beta} = (C + (\Sigma_k^{\beta} + \Sigma_k^{l})^{-1})^{-1} \quad \text{(A10)}$$

where C is the covariance matrix of a zero mean Gaussian distribution describing noise in the observations. The global motion percept is estimated by combining the resulting Gaussians across all locations.

A.3. Parameter values used in the simulations

Parameter	Value	Description
α	5	Constant multiplying exponent of weight function $f(b(x_i^{\text{dof}})$; $\alpha,\ r_{\max})$
$r_{\max}$	350	Maximum value of the weight function $f(b(x_i^{\text{dof}})$; $\alpha,\ r_{\max})$
$w_{\text{cvx}} : w_{\text{sim}}$	8:2	Weights for convexity cue and similarity cue
T_{dof}	*diag* (0.995; 0.005)	Matrix defining pairwise compatibility in form stream
σ_{sqrs}	34.64	Standard deviation of prior for circular motion of square simulation
σ_{bpole}	2.24	Standard deviation of prior for barber-pole simulation

References

Adelson, E. H. (1992). Perceptual organization and the judgment of brightness. *Science*, *262*, 2042–2044.

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *300*, 523–525.

Anderson, C. H., & Van Essen, D. C. (1994). Neurobiological computational systems. In J. M. Zurada, R. J. Marks, II, & C. J. Robinson (Eds.), *Computational intelligence imitating life* (pp. 213–222). New York: IEEE Press.

Baek, K., & Sajda, P. (2003). A probabilistic network model for integrating visual cues and inferring intermediate-level representations. *Proceedings of IEEE Workshop on Statistical and Computational Theories of Vision, Nice, France.* (http://www.stat.ucla.edu/~yuille/meetings/papers/sctv03_10.pdf).

Bosking, W. H., Zhang, Y., Schofield, B., & Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of Neuroscience*, *17*(6), 2112–2127 (http://www.stat.ucla.edu/~yuille/meetings/papers/sctv03_10.pdf).

Budd, J. M. L. (1998). Extrastriate feedback to primary visual cortex in primates: a quantitative analysis of connectivity. *Proceedings of the Royal Society of London B*, *265*, 1037–1044.

Bullier, J., Hupé, J. M., James, A., & Girard, P. (1996). Functional interactions between areas V1 and V2 in the monkey. *Journal of Physiology (Paris)*, *90*, 217–220.

Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience*, *21*, 47–74.

Crowley, J. C., & Katz, L. C. (1999). Development of ocular dominance columns in the absence of retinal input. *Nature Neuroscience*, *2*(12), 1125–1130.

Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: a neural implementation of ideal observers. *Nature Neuroscience*, *2*(8), 740–745.

Driver, J., & Spence, C. (1998). Cross-modal links in spatial attention. *Philosophical Transactions of the Royal Society of London B: Biological Science*, *353*, 1319–1331.

Duncan, R. O., Albright, T. D., & Stoner, G. R. (2000). Occlusion and the interpretation of visual motion: perceptual and neuronal effects of context. *The Journal of Neuroscience*, *20*(15), 5885–5897.

Finkel, L. H., & Sajda, P. (1994). Constructing visual perception. *American Scientist*, *82*, 224–237.

Freeman, W. T., Pasztor, E. C., & Carmichael, O. T. (2000). Learning low-level vision. *International Journal of Computer Vision*, *40*(1), 25–47.

Geisler, W. S., & Diehl, R. L. (2002). Bayesian natural selection and the evolution of perceptual systems. *Philosophical Transactions of the Royal Society of London B: Biological Science*, *357*, 419–448.

Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. *Neuron*, *9*, 1–13.

Grossberg, S., Mingolla, E., & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, *41*, 2521–2553.

Grossberg, S., & Williamson, J. R. (2001). A neural model of horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual grouping and learning. *Cerebral Cortex*, *11*(1), 37–58.

Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (1996). Computational models of cortical visual processing. *Proceedings of the National Academy of Sciences*, *93*, 623–627.

Hinton, G. E., & Brown, A. D. (2000). Spiking Boltzman machines. In S.A. Solla, T.K. Leen, & K.-R. Müller (Eds.), *Advances in Neural Information Processing Systems 12*, Cambridge, MA: MIT Press, pp. 122–128.

Horton, J. C., & Hocking, D. R. (1996). Intrinsic variability of ocular dominance column periodicity in normal macaque monkeys. *Journal of Neuroscience*, *16*(22), 7228–7339.

Hubel, D. H., & Wiesel, T. D. (1977). Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London B*, *198*, 1–59.

Jebara, T. (2004). *Machine learning: Discriminative and generative*. Dordrecht: Kluwer Academic Publishers.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron*, *15*, 843–856.

Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology*, *84*, 2048–2062.

Kersten, D., & Schrater, P. (2000). Pattern inference theory: A probabilistic approach to vision. In R. Mausfeld, & D. Heyers (Eds.), *Perception and the physical world*. Chichester: Wiley.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America (A)*, *20*(7), 1434–1448.

Lidén, L., & Mingolla, E. (1998). Monocular occlusion cues alter the influence of terminator motion in the barber pole phenomenon. *Vision Research*, *38*, 3883–3898.

Lidén, L., & Pack, C. (1999). The role of terminators and occlusion cues in motion integration and segmentation: a neural network model. *Vision Research*, *39*, 3301–3320.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.

Martinez-Conde, S., Cudeiro, J., Grieve, K. L., Rodriguez, R., Rivadulla, C., & Acuña, C. (1999). Effect of feedback projections from area 18 layers 2/3 to area 17 layers 2/3 in the cat visual cortex. *Journal of Neurophysiology*, *82*, 2667–2675.

McAdams, C. J., & Read, R. C. (2003). Effects of attention on the spatial and temporal structure of receptive fields in macaque primary visual cortex. *Society for Neuroscience Annual Meeting, Washington, DC*, 5519.

McDermott, J., Weiss, Y., & Adelson, E. H. (2001). Beyond junctions: nonlocal form constraints on motion interpretation. *Perception*, *30*, 905–923.

Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, *18*, 55–68.

Nakayama, K., & Silverman, G. H. (1988). The aperture problem II: spatial integration of velocity information along contours. *Vision Research*, *28*, 747–753.

Paradiso, M. (1988). A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biological Cybernetics*, *58*, 35–49.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Los Altos, CA: Morgan Kaufmann.

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., & Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature*, *391*(5), 580–584.

Polat, U., & Norcia, A. M. (1996). Neurophysiological evidence for contrast dependent long-range facilitation and suppression in the human visual cortex. *Vision Research*, *36*(14), 2099–2109.

Portilla, J., Strela, V., Wainwright, M. J., Simoncelli, E. P., (2002). Image denoising using gaussian scale mixtures in the wavelet domain. Technical report TR2002-831. Computer Science Department, Courant Institute of Mathematical Sciences, New York University.

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience*, *1*, 125–132.

Raizada, R. D. S., & Grossberg, S. (2003). Towards a theory of the laminar architecture of cerebral cortex: computational clues from the visual system. *Cerebral Cortex*, *13*(1), 100–113.

Rao, R. P. N. (2004). Bayesian computation in recurrent neural circuits. *Neural Computation*, *16*, 1–38.

Romberg, J. K., Choi, H., & Baraniuk, R. G. (2002). Bayesian tree-structured image modeling using wavelet-domain hidden Markov models. *IEEE Transactions on Image Processing*, *10*(7), 1056–1068.

Rubin, E. (1915). *Visuell Wahrgenommene Figuren*. Copenhagen: Gyldenalske Boghandel.

Sajda, P., & Finkel, L. H. (1995). Intermediate-level visual representations and the construction of surface perception. *Journal of Cognitive Neuroscience*, *7*(2), 267–291.

Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, *29*, 619–626.

Spence, C., Parra, L., & Sajda, P. (2000). Hierarchical image probability (hip) models. *IEEE International Conference in Image Processing*, *3*, 320–323.

Stettler, D. D., Das, A., Bennett, J., & Gilbert, C. D. (2002). Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, *36*, 739–750.

Thurston, J. B., & Carraher, R. G. (1966). *Optical Illusions and the Visual Arts*. New York: Reinhold Pub Corp.

Tsunoda, K., Yamane, Y., Nishizaki, M., & Tanifuji, M. (2001). Complex objects are rep-resented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, *4*(8), 832–838.

Wainwright, M. J., & Simoncelli, E. P. (1999). Scale mixtures of gaussians and the statistics of natural images. In S.A. Solla, T.K. Leen, & K.-R. Müller (Eds.), *Advances in Neural Information Processing Systems 12*, pp. 855–861.

Wallach, H. (1935). Üeber visuell whargenommene bewegungrichtung. *Psychologische Forschung*, *20*, 325–380.

Weiss, Y. (1997). Interpreting images by propagating Bayesian beliefs. In M.C. Mozer, M.I. Jordan, & T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9*, pp. 908–915.

Weiss, Y. (2000). Correctness of local probability propagation in graphical models with loops. *Neural Computation*, *12*, 1–42.

Yedidia, J. S., Freeman, W. T., & Weiss, Y. (2003). Understanding belief propagation and its generalizations. In G. Lakemeyer, & B. Nebel (Eds.), *Exploring artificial intelligence in the new millennium* (pp. 239–269). Amsterdam: Elsevier Science and Technology Books.

Zemel, R. (2004). Cortical belief networks. In R. Hecht-Neilsen, & T. McKenna (Eds.), *Computational models for neuroscience*. New York: Springer.

Zemel, R., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation*, *10*, 403–430.

Zhou, H., Friedman, H. S., & Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, *20*(17), 6594–6611.