BAYESIAN NETWORKS FOR MODELING CORTICAL INTEGRATION

Paul Sajda and Kyungim Baek

36.1 INTRODUCTION

Much of our perceptual experience is constructed by the visual cortex. Architecturally, the visual cortex appears to be designed for integrating multiple sources of information that come through top-down, bottom-up, and horizontal connections, yet the computational mechanisms by which the integration and resulting inference occur remain mostly unexplained. A grand challenge in neural engineering is to reverse engineer the visual cortex to reveal the essence of the computation within this distributed neural network.

Researchers from multiple fields of study have been developing theoretical models to describe the computational mechanisms underlying visual perception. Such efforts date back to the 1850s when Helmholtz described perception as "unconscious inference," a process by which the perceiver progresses from experiencing sensations evoked by an object to recognizing its properties. The information present in the images formed on the retina is incomplete and ambiguous and therefore the visual cortex must integrate previous or independently available knowledge about the environment to perform inferential reasoning and construct a scene. This process is made automatically and unconsciously, and eventually one does not even notice the sensations on which it is based [15, 19, 26].

Recently, much focus has been on developing probabilistic frameworks for understanding the neural mechanisms and computational principles underlying inference within the brain. Information present in raw sensory data is inherently complex and noisy and the challenge is to find models capable of dealing with "complex uncertainty." One particular framework which appears promising uses graphical models for representing, learning, and computing multivariate probability distributions. These "Bayesian" graphical models, also termed Bayesian networks, represent a marriage between probability theory and graph theory and provide a network architecture for dealing with uncertainty and complexity. In Bayesian networks, evidence is accumulated by discrete nodes and "passed" between nodes within the network, resulting in a representation of the joint distribution of the constituent variables. These network models provide the computational means for developing probabilistic representations and associated algorithms for inference and learning [22, 35]. Together with their network architecture, Bayesian networks provide a natural theoretical framework for developing distributed computational models of perceptual inference.

Though attractive as a mathematical framework, an open question is whether neurons in the brain are organized to form Bayesian networks. To address this question one must consider the specifics of the anatomy and physiology in visual cortex. A strong vertical organization is one of architectural features of the cortex, an organization principle uncovered by Mountcastle's discovery of the cortical columns [31]. This columnar structure exists for neurons tuned to different stimulus dimensions with the result being ocular dominance, orientation, and position columns. These findings suggest that columns might form a fundamental functional unit for a variety of visual processes. The sets of columns referred to as cortical "hypercolumns" by Hubel and Wiesel [21] appear to include circuitry necessary to process information from a particular location of visual space across a variety of stimulus dimensions. Hypercolumns are connected via long-range horizontal connections [8, 16, 17, 40] which typically span several hypercolumn lengths, enabling communication between nearby hypercolumns. The span of these connections suggests a locally connected network. In addition, there are extensive feedback projections from extrastriate areas, terminating in multiple layers of V1 [6, 44]. Integration of the bottom-up, horizontal, and top-down inputs into a V1 hypercolumn is therefore likely mediated by geniculate, cortico-cortical, and extrastriate feedback connections which are rather sparse and local in nature.

We argue for the view that the cortical hypercolumn serves as a fundamental functional module for integration and inference in the visual system. Furthermore, we propose that the activity of populations of neurons in cortical hypercolumns explicitly represents probability distributions of attributes presented in the visual stimuli and that the anatomical connections between hypercolumns communicate these probability distributions. This leads to the hypothesis that the inference in the network is performed based on the Bayesian network framework. Advantages of the Bayesian network framework in the context of cortical integration include:

1. An explicit representation of uncertainty in the stimulus
2. A common language of probability for integration across cue and modality
3. A sparse, locally connected network architecture

In this chapter we describe a Bayesian network for integrating spatial cues and streams (form and motion) for inferring intermediate-level representations for constructing the visual scene. We show, through simulations, that the Bayesian network leads to results consistent with human perception. In its current form, our model does not attempt to be biophysically realistic, instead focusing on the network-level computation and isomorphisms with cortical networks. For example, nodes in our Bayesian network are not composed of spiking neurons; instead they represent the presumed collective response of a population of neurons, likely in specific cortical layers.

In the following section we describe the integration problem in more detail. We then briefly describe the cortical circuitry in which the problem arises in order to more clearly demonstrate isomorphisms between Bayesian and cortical networks. We then describe a Bayesian network model consisting of Bayesian hypercolumns, collections of nodes processing a local aperture of the visual input space, computing probability distributions for local scene features and integrating these local features via message passing "beliefs" to neighboring hypercolumns. Following the model description, we present simulation results for several simple example problems related to the construction of intermediate-level visual representations.

36.2   PROBLEM OF CORTICAL INTEGRATION

The classical view of information processing in visual cortex is that of a bottom-up process in a feedforward hierarchy. However, the bottom-up information that encodes physical properties of sensory input is often insufficient, uncertain, and even ambiguous for visual processing (e.g., the picture of a Dalmatian dog [50] and Rubin's vase [42]). Psychophysical, anatomical, and physiological evidence suggests that bottom-up factors interact with top-down effects which play a crucial role in the processing of input stimuli. The top-down factors can modulate the expected output of the visual system.

In addition to the bottom-up and top-down inputs, information flows horizontally between sets of cortical neurons—that is, the hypercolumns. This is due to the fundamental aperture problem that the visual system must deal with; Each hypercolumn "sees" a small patch of the visual field decomposed over the input space (orientation, color, motion etc.). The hypercolumns capture information locally in the input space, which is too small to cover visual objects at a global scale. To form coherent representations of objects, nonlocal dependencies in the image space must be captured and exploited. This implies the need for hypercolumns to communicate information about the local visual field to other hypercolumns distributed throughout the cortex. Communication of such nonlocal dependencies can be thought of as a form of *contextual integration*.

One model for visual processing assumes three stages—early, intermediate-, and high-level vision—involving different types of visual representations. When we see the world, the visual system decomposes the scene into surfaces using various cues. This implies that there exist neural representations of surfaces, and it has been thought to be developed at intermediate-level processing [14]. Previous studies suggest that such intermediate-level representations play an important role in integration [32, 47, 54]. For example, Sajda and Finkel focus on the construction of the neural representations of surfaces in the intermediate-level cortical process and present a network-based model [43]. They propose forming surface representations through ownership of contours and argue that such intermediate-level representations serve as a medium for integration.

The different types of information that flow through these bottom-up, top-down, and horizontal connections must be easily and effectively integrated. Anatomical studies indicate that the architecture of the visual cortex is designed for the integration task; however, the computational mechanisms continue to remain largely unknown. In our current model, we assume that the intermediate-level visual representations such as ownership (or direction of figure [43]) might be a substrate for the integration process and focus on developing a computational model for integration.

36.3   CORTICAL CIRCUITRY

The organization of columnar circuitry in the visual system has been extensively studied. For example, Callaway has proposed a generic model of vertical connectivity connecting layers within columns in primary visual cortex of cats and monkey [8]. Although the majority of connections are within columns, physiological and simulation results indicate that there exist horizontal, long-range connections between sets of columns and that these connections give rise to complex, modulatory neuronal responses.

Since the term *hypercolumn* was coined by Hubel and Wiesel [21], 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 $\sim 1\,\mathrm{mm}^2$ and contains

~50,000 neurons. Current experimental and physiological studies have revealed substantial complexity in neuronal response to multiple, simultaneous inputs, including contextual influence, as early as V1 [17, 24, 25]. Such contextual influences, considered to be mediated by modulatory influences of the extraclassical receptive field, are presumed to arise from the horizontal inputs from neighboring hypercolumns and/or from feedback from extrastriate areas [7, 10, 36, 37, 46].

It appears that in the visual system information computed locally across the input space by individual hypercolumns is propagated within the network and integrated, perhaps in a modulatory way, to influence neural responses which ultimately directly correlate with perception. In the next section we describe how such integration could be done within a probabilistic framework, leading to what we term *Bayesian hypercolumns*.

36.4   BAYESIAN INFERENCE IN NETWORKS

36.4.1   Bayesian Hypercolumns

The Bayesian hypercolumn framework provides a structure for integrating top-down, bottom-up, and horizontally transmitted pulse packets and results in a neural analog of Bayesian inference. In the brain, information is encoded by populations of neurons rather than by a single neuron. Since the initial studies on the statistical properties of population codes that first used Bayesian techniques to analyze them [34], researchers have suggested that it might be probability distributions instead of single values that the population codes actually represent [1, 39, 56]. A computational model proposed by Deneve et al. [11] 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 functions 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.

We argue that the pattern of activity distributed across each hypercolumn instantiates a conditional probability density function (PDF) of some random variables (feature, object, etc.). In striate visual cortex, for example, the PDF might specify an estimate of local orientation as well as the uncertainty in this estimate. Horizontal and top-down inputs modulate this activity, thus expressing the conditional probability of the output given the current links to the population. Anatomical connections communicate the PDFs between hypercolumns. The connection weights are learned through training but capture statistical regularities in the image domain. In visible-domain spatial imagery, these statistical regularities would correspond to the Gestalt laws of perceptual organization.

As Zemel et al. [56] have observed, "a population of neurons is a terrible thing to waste representing a scalar." Accordingly, the PDF representation endows the hypercolumn with computational power far surpassing an equivalent number of "grandmother cells." As environmental conditions change, due to noise or novelty, each hypercolumn recalculates the likelihood of its own distribution based on the state of other hypercolumns. This represents a form of inference, analogous to that carried out in formal Bayesian networks. Recognition and classification are carried out by projecting the current stimulus onto the prior probabilities stored in the cortical architecture.

36.4.2 Bayesian Networks as Model for Cortical Architecture

Bayesian networks are directed graphs in which the nodes represent variables of interest (e.g., stimulus features, target classes) and the links represent conditional probability distributions between possible values of these variables. Bayesian networks are "deliberative" rather than passive reactive agents; thus, unlike traditional artificial neural networks, they respond to novel or changing stimulus conditions instantaneously without relearning [35]. If the environment changes or an input channel is lost, the joint PDF of features and targets is simply recomputed based on the new inputs and stored a priori knowledge. Two features of Bayesian networks account for this advantage over traditional neural networks: The messages passed are probability distributions and message passing is multithreaded so that the influence from different streams is not corrupted. Figure 36.1 illustrates the isomorphism between cortical hypercolumns and the Bayesian network architecture which we propose.

Bayesian networks enable a direct way to integrate different data types (top down, bottom up), since all variables are all mapped to the same probability space. Use of a probabilistic format provides a measure of the degree of confidence together with the most likely classification. Most importantly, Bayesian networks are inference engines, and they can be used to detect relationships among variables as well as a description of these relationships upon discovery.

Over the last several years, substantial progress has been made in Bayesian network theory, probabilistic learning rules, and their application to image analysis [38, 41, 45, 51]. Recent work by Hinton and Brown [20] has been directed at trying to understand the relationship between Bayesian and neural processing. Application of this neuro-Bayesian **Q1**
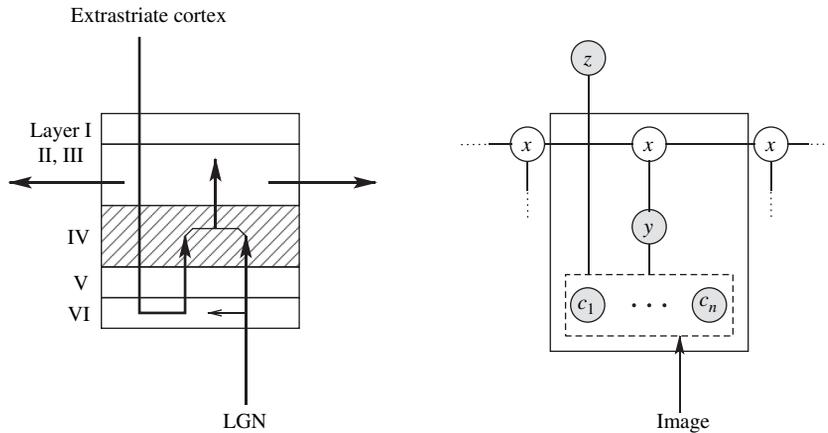

Figure 36.1   Simplified diagram of cortical connections focusing on laminar structure of a hypercolumn (left) and isomorphic architecture of proposed Bayesian network model (right). The lateral geniculate nucleus (LGN) provides bottom-up input via strong connections to layer IV and indirect connections passing through layer VI. The top-down feedback signals from extrastriate cortex pass into layer VI, then are projected onto layer IV. The feedforward connections from layer IV subsequently provide activation to layer II/III, which forms intrinsic horizontal, long-range connections between hypercolumns. In our network model images correspond to the bottom-up LGN input, from which a set of features are extracted (nodes c_i). Those features are combined and also with the top-down, prior knowledge represented by node z to form PDFs for observations (node y). Then the observations are passed into hidden nodes x, corresponding to layer II/III in the hypercolumn structure. The contextual integration is performed by passing PDFs between neighboring hidden nodes. **Q2**

approach to image processing has been significantly advanced by Weiss [52], 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 Marr's dictum [29] 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. In contrast, most current algorithms, which are based on relaxation and regularization methods, do not converge at multiple times this number of iterations. Thus Bayesian methods offer the hope of matching the time constraints posed by human visual recognition.

Weiss has also demonstrated the ability of Bayesian networks to infer intermediate-level representations, for example, to integrate local and more global shape information to determine the direction of figure (DOF) [52]. 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 [53]. These processes have clear ties to cortical networks [18], and such processes have been implemented in retino-morphic and cortex-based chips [4] and in multichip systems with field-programmable connections [5].

36.4.3   Belief Propagation

As mentioned previously, in graphical models (or Bayesian networks, in particular) a node represents a random variable and links specify the dependency relationships between these variables [22]. The states of the random variables can be hidden in the sense that they are not directly observable, but it is assumed that they may have observations related to the state values. The states of hidden variables are inferred from the available observations, and the belief propagation (BP) algorithm solves inference problems based on local message passing. In this section, a BP algorithm is described on undirected graphical models with pairwise potentials. It has been shown that most graphical models can be converted into this general form [55].

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

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

where c is a normalizing constant, x_i represents the state of node i, $T_{ij}(x_i, x_j)$ captures the compatibility between neighboring nodes x_i and x_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 x_i over all x_j other than x_i is called the local *belief* $b(x_i)$.

The BP algorithm iterates a local message computation and belief updates [55]. The message $M_{ij}(x_j)$ passed from a hidden node x_i to its neighboring hidden node x_j represents the probability distribution over the state of x_j. In each iteration, messages and beliefs are updated as follows:

$$M_{ij}(x_j) = c \int_{x_i} dx_i \, T_{ij}(x_i, x_j) E_i(x_i, y_i) \prod_{x_k \in N_i / x_j} M_{ki}(x_i)$$

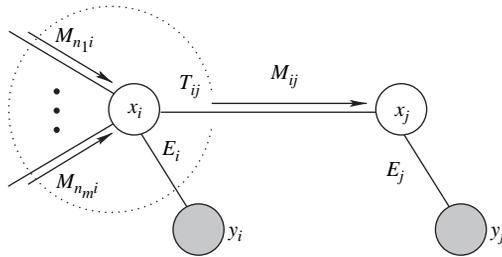$$b(x_i) = c E_i(x_i, y_i) \prod_{x_k \in N_i} M_{ki}(x_i)$$

Figure 36.2   Illustration of local message passing from node x_i to node x_j. Open circles are hidden variables, while shaded circles represent observed variables. The local belief at node x_j is computed by combining the incoming messages from all its neighbors and the local interaction E_j.

where N_i/x_j denotes a set of neighboring nodes of x_i except x_j. Here, M_{ij} is computed by combining all messages received by x_i from all neighbors except x_j in the previous iteration and marginalizing over all possible states of x_i (Fig. 36.2). The current local belief is estimated by combining all incoming messages and the local observations.

It has been shown that, for singly connected graphs, BP converges to exact marginal probabilities [55]. Although how it works for general graphs is not well understood, experimental results on some vision problems, such as motion analysis, also show that BP works well for graphs with loops [13].

36.5   SIMULATION RESULTS

We demonstrate the application of a Bayesian network model, having locally connected nodes each computing a distribution over multiple stimulus dimensions (i.e., cues), for two example problems involving intermediate-level surface representations: computing DOF and integrating form and motion cues. For the DOF example, multiple spatial cues are combined to infer the relative figure–ground relationships. The nodes corresponding to hidden variables represent the beliefs (or probabilities) for the DOF. These beliefs are constructed by probabilistically integrating estimates based on the observed cues and beliefs at neighboring nodes which are integrated via the network of horizontal connections.

The model for computing DOF (form stream) is then extended by adding a second stream for motion processing. Each node in the motion stream computes probability distributions for local direction of motion in the same way as the DOF is estimated, except that it takes into account the influence of the local DOF estimated in the form stream. Simulations show the two network streams can interact, through Bayesian mechanisms, to qualitatively account for motion coherence phenomena uncovered via psychophysical experiments.

36.5.1   Inferring Direction of Figure

"Direction of figure" represents the contour ownership of an object's occluding boundary to a region which represents the object's surface [43]. Therefore DOF specifies a figure–ground relationship directly for occluding contours. In studies on visual perception, the contour ownership is hypothesized to be assigned in early stages of visual processing, which received support from recent neurophysiological findings reported by Zhou et al. [57], showing that more than half of the neurons in extrastriate cortex (areas V2 and V4) are selective to contour ownership. One can consider the assignment of DOF as

essentially a problem in probabilistic inference, with DOF being a hidden variable that is not directly observed but can potentially be inferred from local observations and some form of message passing via horizontal dependencies.

As shown in many perceptual studies, various cues are used in the figure–ground discrimination [23, 33, 43]. In our model [3] 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} 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^{dof} and $y_{i,\mathrm{sim}}$ is computed from the similarity/proximity cue by looking for points having similar local tangent angle (i.e., orientation) that lie in a direction orthogonal to the contour at point i. The interaction $E_{i,\mathrm{cvx}}$ prefers smaller angles and $E_{i,\mathrm{sim}}$ favors shorter distances with similar orientations.

The two local interactions are combined based on a *weak fusion model* [9]. The weak fusion model is a simple scheme which suggests that a property of the environment is estimated separately by each independent cue and then combined in a weighted linear fashion to compute the overall effect. Following the weak fusion scheme, the total local interaction $E_i(x_i^{\mathrm{dof}}, y_i^{\mathrm{dof}})$ is computed by the weighted average of the interactions made by two separate observations $y_{i,\mathrm{cvx}}$ and $y_{i,\mathrm{sim}}$:

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

There are several classic examples in which the discriminating figure from the background is not immediately clear. The first two square spiral figures of Figure 36.3*a* are
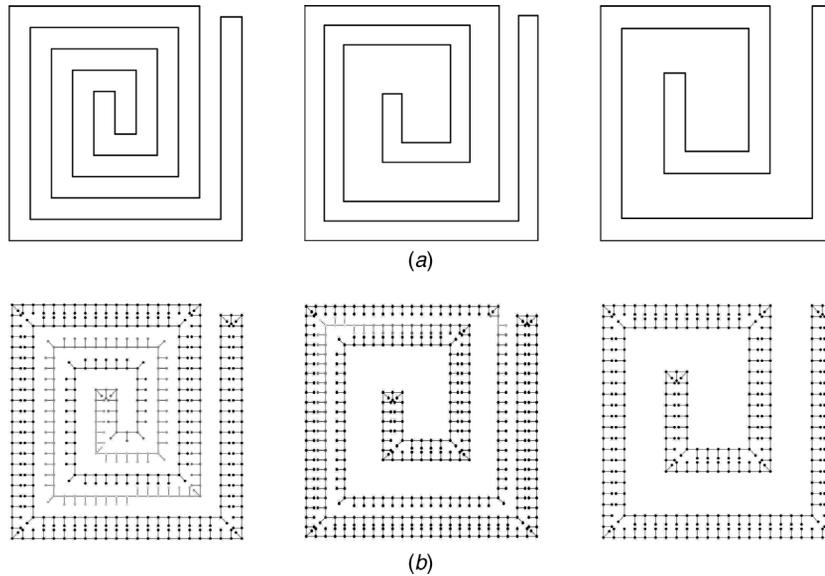

(*a*)

(*b*)

Figure 36.3   (*a*) *Left*: spiral figure in which figure–ground discrimination is ambiguous unless serially tracing the contour. *Middle*: increasing spiral width as it winds around toward center generates incorrect perception of figure surface. *Right*: figure in which correct decision of figure surface can be made immediately. All figures are reproduced from [43]. (*b*) DOF estimate by the model for the three spiral images shown on top. Resulting DOF is represented by small bars with varying intensity, showing the degree of certainty for the corresponding direction. Note the low certainty (bright) for the central part of the first spiral, indicating ambiguity of DOF, consistent with human perception. The high certainty (dark) for the central part of the second spiral shows an incorrect percept, which is also consistent with human interpretation.

such examples [43]. Discrimination of the figure's surface in the first spiral is difficult, with difficulty increasing for regions close to the center of the spiral.[1] On the other hand, discriminating the figure surface in the second spiral seems to be straightforward. Immediately, we tend to perceive the thin strip in the center as the figure, but this is in fact incorrect. In this case, the width of the spiral increases as it winds around toward the center, generating an incorrect perception of the figure–ground. Unlike the first two figures, figure–ground can be correctly inferred almost instantly for the third spiral.

Figure 36.3*b* shows that the network model predicts DOF very close to human perception. The DOF figures illustrate increasing ambiguity and/or incorrect interpretation for the center region of the first two spirals and perfect figure segmentation for the last spiral. Although they are not shown here, results indicate that it takes longer for the network model to converge for the first two spirals, with more oscillations in the DOF assignment, as compared to the third example. The results are slightly different, especially near the periphery of the spirals, from those obtained by the neural computational models described in [43]. This is likely due to the fact that the current model does not exploit observations from closure cues.

Certain figures are perceptually ambiguous in that figure and ground can shift or even oscillate. One of the most famous figures that demonstrates this perceptual ambiguity is Rubin's vase, shown in Figure 36.4*a*. In Figure 36.4, one can perceive either faces or a vase (never both simultaneously), and whenever the perceived object is shifted, the contour ownership is also changed accordingly.

One can bias the interpretation of the ambiguous figures by providing prior information. For example, prior cues might emerge from recognition of distinct visual features (e.g., nose, eyes, chin). In our probabilistic network model, prior information can be considered as another cue for DOF and therefore combined in the same way as described above. In this case, the network model integrates bottom-up (convexity), horizontal (similarity/proximity), and top-down (prior information) cues using a simple combination mechanism.

Figure 36.4*b* shows the DOF estimation for the Rubin's vase image without using prior information. The network assigns the contour to faces mainly because of the influence from the convexity cue; however, the confidence level is low throughout the entire contour. Prior information is added by explicitly specifying figure direction at several locations on the contour (Fig. 36.4*c*). The results of combining prior cues are shown on the right. Even if a small weight is assigned for the prior cue,[2] it strongly biases the DOF.

36.5.2   Integrating Form and Motion Streams

The motion of a homogeneous contour is perceptually ambiguous because a neuron responding to the moving contour sees only a limited extent of the visual scene. Due to this "aperture problem," a single local measurement along the contour cannot determine the object's motion, and therefore motion integration across space is required. There are also several features that may provide unambiguous local motion and dominate the motion integration. Examples of such features are line terminators and junctions.

The degree of certainty in DOF at junctions inferred in the form stream can be used to distinguish between *intrinsic terminators*, which are due to the natural end of a line of an

---

[1]Note that all DOF relationships can be determined if one traces the contours of the spiral and "propagates" high-confidence DOF assignments to low-confidence regions. However, in this discussion we ignore such "tracing," though it could be integrated as another form of prior information/cue.

[2]The ratio of w_{prior} over $(w_{\text{cvx}} + w_{\text{sim}})$ is $1:8$.
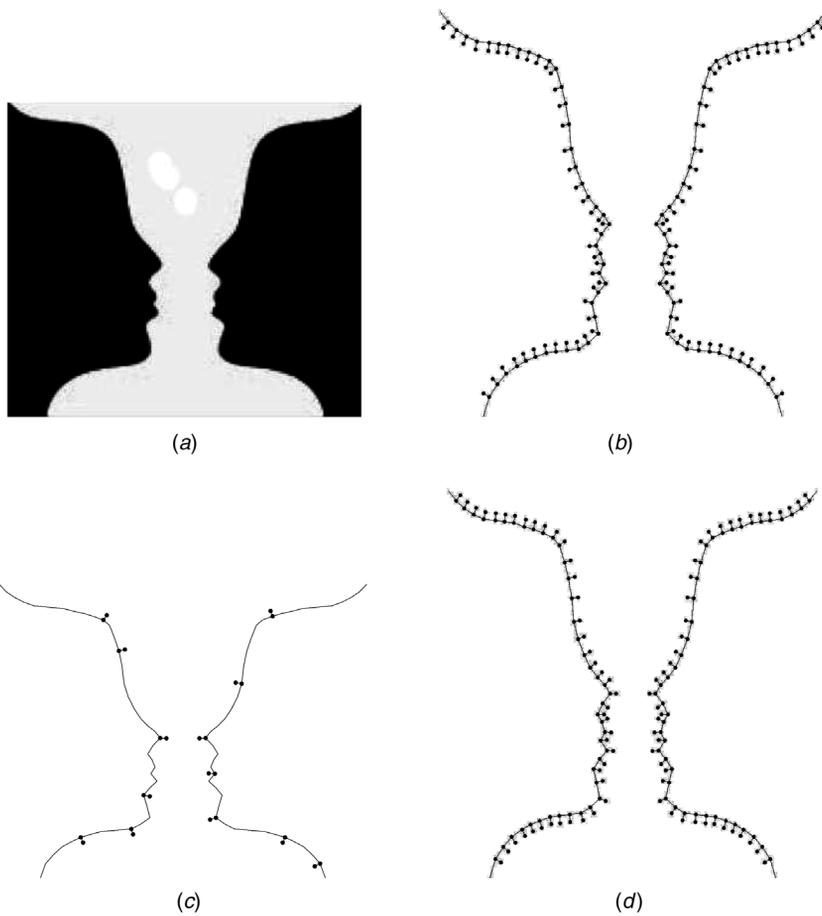
Figure 36.4   (*a*) Rubin's vase (from [2]). (*b*) DOF prediction without using prior information. (*c*) Prior cues for vase features. (*d*) DOF estimate of network model integrating prior information. Figure shows the bias of DOF toward the vase, reflecting strong influence of the prior cue.

object, and *extrinsic terminators*, which are not created by the end of the line itself but rather are caused by occlusion by another surface. Intrinsic terminators provide an unambiguous signal for the true velocity of the line, while extrinsic terminators provide a locally ambiguous signal which must be suppressed for accurate motion computation [32]. The DOF is an explicit surface representation, and therefore the degree of certainty (i.e., belief) in DOF at junctions can be used to represent the strength of the evidence for surface occlusion, which determines the terminator type.

The hidden node in the motion stream represents a velocity of corresponding location along the contour. We assume that both the pairwise compatibility $T_{ij}(x_i^{\mathrm{mot}}, x_j^{\mathrm{mot}})$ and the local interaction $E_i(x_i^{\mathrm{mot}}, y_i^{\mathrm{mot}})$ that models the velocity likelihood at apertures are Gaussian. Also, T_{ij} is set manually and E_i is defined by a mean of the normal velocity at point i and a local covariance matrix Cov_i. Before the BP algorithm starts, the variance at junction points is modulated by a function of DOF belief $b(x_i^{\mathrm{dof}})$ as follows:

$$\mathrm{Cov}_i = e^{\alpha\{b(x_i^{\mathrm{dof}})-0.5\}} \cdot \mathrm{Cov}_i$$

Initially, the covariance matrices of hidden variables are set to represent infinite uncertainty, and mean vectors are set to zero. When the BP algorithm converges, the motion integration is performed by computing the mixture of Gaussians:

$$p(v) = \sum_i p(v \mid i)p(i)$$

where $p(v \mid i)$ is the probability of velocity v from the resulting Gaussian of hidden variable x_i^{mot} and $p(i)$'s are the mixing coefficients.

Figure 36.5a shows the two stimuli used for the experiments. They are $90°$ rotated versions of the diamond stimuli described in McDermott et al. [30], but the perceptual
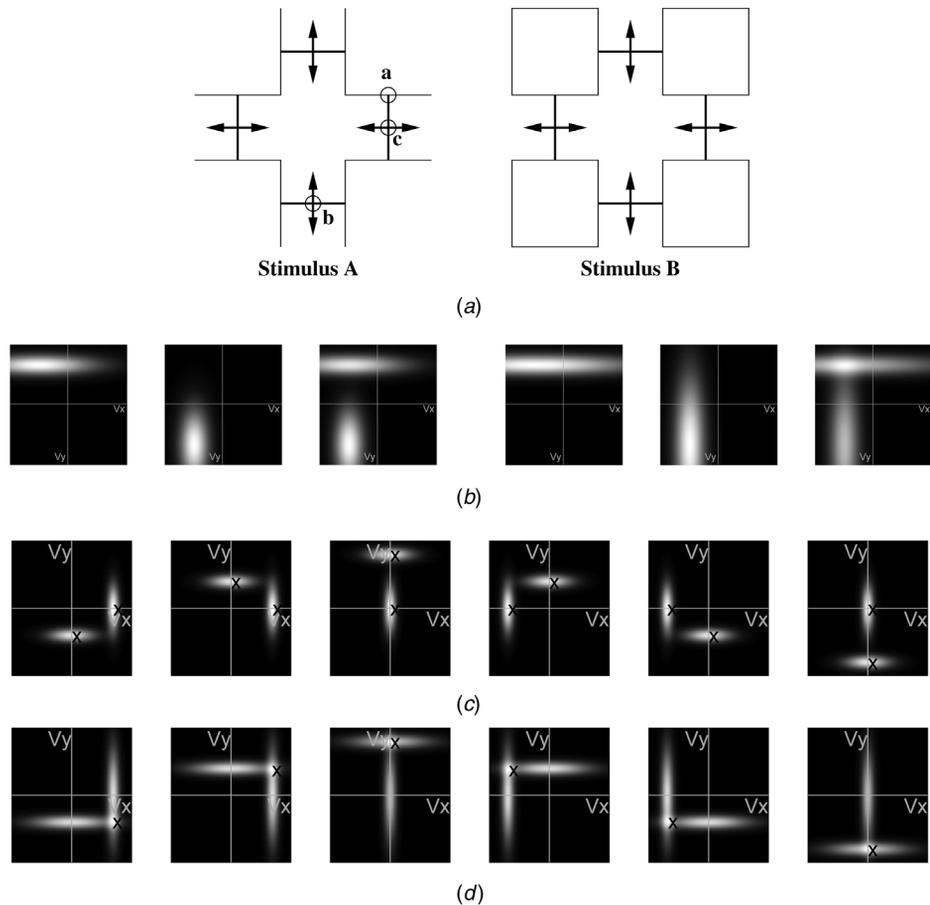

(a)

(b)

(c)

(d)

Figure 36.5   (a) Stimuli generated from four line segments that move sinusoidally, $90°$ out of phase, in vertical and horizontal directions. The line segments are presented with L-shaped occluders (stimulus A) and closed square occluders (stimulus B). The presence of the occluding surface alters the motion perception. (b) Velocity estimation results for stimulus A (first three figures) and stimulus B (last three figures). *First and fourth*: Estimation at location **b**. *Second and fifth*: Estimation at location **c**. *Third and last*: Velocity computed by combining the two estimations using mixture of Gaussian. (c,d) A sequence of resulting velocity estimation for six successive frames sampled from a period of sinusoidal motion with a regular interval. (c) Two separate motions oscillating in the direction normal to the line segment for stimulus A. The sequence in (d), on the other hand, shows a coherent motion forming a circular trajectory.

effect is basically the same. The motion of moving line segments is identical between the two stimuli. A pair of vertical line segments move together sinusoidally in a horizontal direction while the two horizontal lines move simultaneously in a vertical direction. The vertical and horizontal motions are 90° out of phase. The difference between the two stimuli is the shape of the occluders, which alters the perceived motion. A single coherent rotation is perceived for stimulus B, while we are more likely to see two separate motions of the line segments for stimulus A [30].

Figure 36.5*b* shows the resulting velocity estimated by the network model for stimulus A (first three figures) and stimulus B (last three figures). Since the estimated motions along the line segments that move simultaneously are almost identical after convergence, the first two figures for each stimulus show only the motions at locations **b** and **c** (shown in Fig. 36.5*a* in the velocity space). In the third and last figures in Figure 36.5*b*, which display **Q2** the integrated motion, we clearly see the bimodal distribution for stimulus A, while rather a single peak is formed at the intersection of two distributions for stimulus B. This implies that we perceive two separate motions for stimulus A and a single coherent motion for stimulus B. Figures 36.5*c,d* illustrate the resulting velocity estimation for six successive frames sampled from a period of the sinusoidal motion. The maximum a posterior in **Q2** each frame follows a circular trajectory for stimulus B, which is consistent with perceiving rotation.

A prior cue can easily be added for inferring DOF in stimulus A. Psychophysically, this might simulate priming subjects or integrating disparity information to add stereoscopic depth so that the subjects are biased to see the occlusion of L-shaped figures more strongly or weakly. Adding priors for stronger occluders strengthens the belief for DOF in the indicated direction, and consequently more coherent motion would be expected. Although not shown here, as the weights on the prior cue are increased, the solution becomes more unimodal and with a strong prior it produces a single peak at the intersection similar to stimulus B.

36.6 CONCLUSION

In this chapter we have attempted to demonstrate how Bayesian networks might represent an appropriate theoretical framework for understanding the computation performed by cortical hypercolumns for inferring intermediate-level visual representations and integrating cues across form and stream. The hypercolumn architecture is prevalent at all levels of cortical visual processing, not just in primary visual cortex. Thus one would expect the framework for Bayesian integration might also occur at other levels of processing, such as object recognition, occurring in the hypercolumn networks of inferotemporal cortex.

We have not addressed the issue of learning and adaptation, which is critical in cortical networks. The parameters for our model are fixed. We are now exploring methods for learning the emission and transition probabilities in the network given an ensemble of natural spatiotemporal imagery. If the Bayesian network framework is consistent with cortical network processing, we would expect this to provide insight into the organization of receptive fields and contextual interactions seen in the biological networks.

We have only demonstrated the use of Bayesian networks in the inference mode. However, since they represent a generative model, such networks have a much broader utility. For example, one can use the model to synthesize novel "scenes." In the context of cortical networks, this is consistent with the ideas that early visual cortical areas are involved in generating visual mental imagery [27, 28]. The ability to use the framework

to, for example, infer, discriminate, synthesize, and compress makes it attractive as a model of general-purpose visual processing.

The implementations we have presented are not meant to be complete models, instead providing the skeleton of a framework suggesting how spatial cues from multiple streams might be integrated using a common framework to construct a scene. Others have attempted to develop more biologically realistic models of form and motion integration, with mechanisms that are perhaps more neural. However such models are difficult to analyze, particularly in terms of understanding the general role/effect of uncertainty in the observations and reversal or ambiguity in perception. Although we do not claim that our current implementation is a biologically realistic model, it is worth noting that it is constructed as a network model, with no requirements for global connectivity or consistency. Further work by our group is aimed at developing biophysically realistic models of cortical hypercolumns and the cortico-cortical connectivity to investigate what specific neural machinery could be used to locally construct distributions of observations and integrate these via local message passing. Ultimately our goal is a "reverse neuroengineering" of these hypercolumn networks.

ACKNOWLEDGMENTS

REFERENCES

1. C. H. ANDERSON AND D. C. VAN ESSEN, "Neurobiological computational systems," in J. M. ZURADA, R. J. MARKS II, AND C. J. ROBINSON (Eds.), *Computational Intelligence Imitating Life*, pp. 213–222, IEEE Press, New York, 1994.

2. T. J. ANDREWS, D. SCHLUPPECK, D. HOMFRAY, P. MATTHEWS, AND C. BLAKEMORE, "Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion," *NeuroImage*, vol. 17, pp. 890–901, 2002.

3. K. BAEK AND P. SAJDA, "A probabilistic network model for integrating visual cues and inferring intermediate-level representations," Paper presentd at Third International Workshop on Statistical and Computational Theories of Vision, Nice, France, October 2003.

4. K. A. BOAHEN, "Retinomorphic chips that see quadruple images," Paper presented at MicroNeuro'99: 7th International Conference on Neural, Fuzzy, and Bio-Inspired Systems, Granada, Spain, IEEE Computer Society Press, 1999.

5. K. A. BOAHEN, "Point-to-point connectivity between neuromorphic chips using address events," *IEEE Transactions on Circuits and Systems-II*, vol. 47, pp. 416–434, 2000.

6. J. M. L. BUDD, "Extrastriate feedback to primary visual cortex in primates: A quantitative analysis of connectivity," *Proceedings of the Royal Society of London B*, vol. 265, pp. 1037–1044, 1998.

7. J. BULLIER, J. M. HUPÉ, A. JAMES, AND P. GIRARD, "Functional interactions between areas V1 and V2 in the monkey," *Journal of Physiology (Paris)*, vol. 90, pp. 217–220, 1996.

8. E. M. CALLAWAY, "Local circuits in primary visual cortex of the macaque monkey," *Annual Review of Neuroscience*, vol. 21, pp. 47–74, 1998.

9. J. J. CLARK AND A. L. YUILLE, *Data Fusion for Sensory Information Processing Systems*, Kluwer Academic, Boston, 1990.

10. S. MARTINEZ-CONDE, J. CUDEIRO, K. L. GRIEVE, R. RODRIGUEZ, C. RIVADULLA, AND C. ACUÑA, "Effect of feedback projections from area 18 layers 2/3 to area 17 layers 2/3 in the cat visual cortex," *Journal of Neurophysiology*, vol. 82, pp. 2667–2675, 1999.

11. S. DENEVE, P. E. LATHAM, AND A. POUGET, "Reading population codes: A neural implementation of ideal observers," *Nature Neuroscience*, vol. 2, no. 8, pp. 740–745, 1999.

12. M. DIESMANN, M. O. GEWALTIG, AND A. AERTSEN, "Stable propagation of synchronous spiking in cortical neural networks," *Nature*, vol. 402, pp. 529–533, 1999.

13. W. T. FREEMAN, E. C. PASZTOR, AND O. T. CARMICHAEL, "Learning low-level vision," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 25–47, 2000.

14. L. H. FINKEL AND P. SAJDA, "Constructing visual perception," *American Scientist*, vol. 82, pp. 224–237, 1994.

15. W. S. GEISLER AND D. KERSTEN, "Illusions, perception and Bayes," *Nature Neuroscience*, vol. 5, no. 6, pp. 508–510, 2002.

16. C. D. GILBERT AND T. N. WIESEL, "Clustered intrinsic connections in cat visual cortex," *Journal of Neuroscience*, vol. 3, pp. 1116–1133, 1983.

17. C. D. GILBERT, "Horizontal integration and cortical dynamics," *Neuron*, vol. 9, pp. 1–13, 1992.

18. D. J. HEEGER, E. P. SIMONCELLI, AND J. A. MOVSHON, "Computational models of cortical visual processing," *Proceedings of the National Academy of Sciences*, vol. 93, pp. 623–627, January 1996.

19. H. HELMHOLTZ, *Physiological Optics, Vol. III: The Perceptions of Vision* (J. P. Southall, Trans.), Optical Society of America, Rochester, NY, 1925.

20. G. E. HINTON AND A. D. BROWN, "Spiking Boltzman machines," *Advances in Neural Information Processing Systems*, vol. 12, 2000.

21. D. H. HUBEL AND T. N. WIESEL, "Functional architecture of macaque monkey visual cortex," *Proceedings of the Royal Society of London B*, vol. 198, pp. 1–59, 1977.

22. M. JORDAN, *Graphical Models: Foundations of Neural Computation*, MIT Press, Cambridge, MA, 2001.

23. G. KANIZSA, *Organization in Vision*, Praeger, New York, 1979.

24. M. K. KAPADIA, M. ITO, C. D. GILBERT, AND G. WESTHEIMER, "Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys," *Neuron*, vol. 15, pp. 843–856, 1995.

25. M. K. KAPADIA, G. WESTHEIMER, AND C. D. GILBERT, "Spatial distribution of contextual interactions in primary visual cortex and in visual perception," *Journal of Neurophysiology*, vol. 84, pp. 2048–2062, 2000.

26. D. C. KNILL, D. KERSTEN, AND A. YUILLE, "Introduction: a Bayesian formulation of visual perception," in D. C. KNILL AND W. R. RECHARDS (Eds.), *Perception as Bayesian Inference*, pp. 1–21, Cambridge University Press, Cambridge, 1996.

27. S. M. KOSSLYN, "Visual mental images and re-presentations of the world: A cognitive neuroscience approach," in J. S. GERO AND B. TVERSKY (Eds.), *Visual and Spatial Reasoning in Design*, MIT Press, Cambridge, MA, 1999.

28. S. M. KOSSLYN, A. PASCUAL-LEONE, O. FELICIAN, J. P. KEENAN, W. L. THOMPSON, G. GANIS, K. E. SUKEL, AND N. M. ALPERT, "The role of area 17 in visual imagery: Convergent evidence from PET and rTMS," *Science*, vol. 284, pp. 167–170, 1999.

29. D. MARR, *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information*, W. H. Freeman, San Francisco, 1982.

30. J. MCDERMOTT, Y. WEISS, AND E. H. ADELSON, "Beyond junctions: Nonlocal form constraints on motion interpretation," *Perception*, vol. 30, pp. 905–923, 2001.

31. V. MOUNTCASTLE, "Modality and topographic properties of single neurons of cat's somatic sensory cortex," *Journal of Neurophysiology*, vol. 20, pp. 408–434, 1957.

32. K. NAKAYAMA, S. SHIMOJO, AND G. H. SILVERMAN, "Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects," *Perception*, vol. 18, pp. 55–68, 1989.

33. K. NAKAYAMA, "Binocular visual surface perception," *Proceedings of the National Academy of Sciences USA*, vol. 93, pp. 634–639, 1996.

34. M. PARADISO, "A theory for the use of visual orientation information which exploits the columnar structure of striate cortex," *Biological Cybernetics*, vol. 58, pp. 35–49, 1988.

35. J. PEARL, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988.

36. U. POLAT AND A. M. NORCIA, "Neurophysiological evidence for contrast dependent long-range facilitation and suppression in the human visual cortex," *Vision Research*, vol. 36, no. 14, pp. 2099–2109, 1996.

37. U. POLAT, K. MIZOBE, M. W. PETTET, T. KASAMATSU, AND A. M. NORCIA, "Collinear stimuli regulate visual responses depending on cell's contrast threshold," *Nature*, vol. 391, no. 5, pp. 580–584, 1998.

38. J. PORTILLA, V. STRELA, M. J. WAINWRIGHT, AND E. P. SIMONCELLI, "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. September 2002.

39. A. POUGET, P. DAYAN, AND R. ZEMEL, "Information processing with population codes," *Nature Reviews Neuroscience*, vol. 1, pp. 125–132, 2000.

40. K. S. ROCKLAND AND J. S. LUND, "Intrinsic laminar lattice connections in primate visual cortex," *Journal of Comparative Neurology*, vol. 216, pp. 303–318, 1983.

41. J. K. ROMBERG, H. CHOI, AND R. G. BARANIUK, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1056–1068, 2001.

42. E. RUBIN, *Visuell Wahrgenommene Figuren*, Gyldenalske Boghandel, Copenhagen, 1915.

43. P. SAJDA AND L. H. FINKEL, "Intermediate-level visual representations and the construction of surface perception," *Journal of Cognitive Neuroscience*, vol. 7, no. 2, pp. 267–291, 1995.

44. P. SALIN AND J. BULLIER, "Corticocortical connections in the visual system: Structure and function," *Physiological Reviews*, vol. 75, no. 1, pp. 107–154, 1995.

45. C. SPENCE, L. PARRA, AND P. SAJDA, "Hierarchical image probability (HIP) models," in *IEEE International Conference in Image Processing*, vol. 3, pp. 320–323, 2000.

46. D. D. STETTLER, A. DAS, J. BENNETT, AND C. D. GILBERT, "Lateral connectivity and contextual

Q3

Q4

Q5

interactions in macaque primary visual cortex," *Neuron*, vol. 36, pp. 739–750, 2002.

47. G. R. STONER AND T. D. ALBRIGHT, "Image segmentation cues in motion processing: Implications for modularity in vision," *Journal of Cognitive Neuroscience*, vol. 5, no. 2, pp. 129–149, 1993.

Q6 48. S. SUBRAMANIAM, I. BIEDERMAN, AND S. A. MADIGAN, "Accurate identification but no priming and chance recognition memory for pictures in RSVP sequences," *Visual Cognition*, vol. 7, pp. 511–535, 2000.

Q6 49. S. THORPE, D. FIZE, AND C. MARLOT, "Speed of processing in the human visual system," *Nature*, vol. 381, pp. 519–522, June 1996.

50. J. B. THURSTON AND R. G. CARRAHER, *Optical Illusions and the Visual Arts*, Reinhold, New York, 1966.

51. M. J. WAINWRIGHT AND E. P. SIMONCELLI, "Scale mixtures of Gaussians and the statistics of natural images," Paper presented at Conference on Neural Information Processing Systems, Denver, pp. 855–861, 1999.

52. Y. WEISS, "Bayesian belief propagation for image understanding," Paper presented at Workshop on Statistical and Computational Theories of Vision, 1999.

53. Y. WEISS, "Correctness of local probability propagation in graphical models with loops," *Neural Computation*, vol. 12, pp. 1–42, 2000.

54. M. WHITE, "A new effect of pattern on perceived lightness," *Perception*, vol. 3, no. 4, pp. 413–416, 1979.

55. J. S. YEDIDIA, W. T. FREEMAN, AND Y. WEISS, "Understanding belief propagation and its generalizations," in G. LAKEMEYER AND B. NEBEL (Eds.), *Exploring Artificial Intelligence in the New Millennium*, pp. 239–269, **Q7** 2003.

56. R. S. ZEMEL, P. DAYAN, AND A. POUGET, "Probabilistic interpretation of population codes," *Neural Computation*, vol. 10, pp. 403–430, 1998.

57. H. ZHOU, H. S. FRIEDMAN, AND R. VON DER HEYDT, "Coding of border ownership in monkey visual cortex," *Journal of Neuroscience*, vol. 20, no. 17, pp. 6594–6611, 2000.