Lecture 5

This lecture describes how groups of neurons can perform edge detection, edge grouping, stereo and motion.
We also introduce weak methods for cue combination.
This lecture include the demos: (4c) Hopfield Network for Binocular Stereo.
(5) Cue Combination.

The Line Process Model (I)

Our first example is the classic *line process* model [40][12][122] which was developed as a way to segment images. It has explicit *line process* variables which "break" images into regions where the intensity is piecewise smooth. Our presentation will follow the work of ([85]) who translated it into neural circuits. The model takes intensity values $\vec{I}$ as input and outputs smoothed intensity values. But this smoothness is broken at places where the intensity changes are too high, see figure (28). The model has continuous variables $\vec{J}$ representing the intensity and binary-valued variables $\vec{l}$ for the line processes (or edges). The model is formulated as performing *maximum a posteriori* (MAP) estimation. The algorithm for estimating MAP is a neural network model which can be derived from the original Markov Model [40] by mean field theory [36]. Note that in this model the variables do not have to represent intensity. Instead they can represent texture, depth, or any other property which is spatially smooth except at sharp discontinuities.

The Line Process Model (II)

For simplicity we present the weak membrane model in one-dimension. The input is $\vec{I} = \{I(x) : x \in \mathcal{D}\}$, the estimated, or smoothed, image is $\vec{J} = \{J(x) : x \in \mathcal{D}\}$, and the line processes are denoted by $\vec{l} = \{l(x) : x \in \mathcal{D}\}$, where $l(x) \in \{0, 1\}$.
The model is specified by a posterior probability distribution:

$$P(\vec{J}, \vec{l} | \vec{I}) = \frac{1}{Z} \exp\{-E[\vec{J}, \vec{l} : \vec{I}]/T\},$$

where

$$E[\vec{J}, \vec{l} : \vec{I}] = \sum_x (I(x) - J(x))^2 + A \sum_x (J(x+1) - J(x))^2 (1 - l(x)) + B \sum_x l(x).$$

The Line Process Model (III)

The first term ensures that the estimated intensity $J(x)$ is close to the input intensity $I(x)$. The second encourages the estimated intensity $J(x)$ to be spatially smooth (e.g., $J(x) \approx J(x+1)$), unless a line process is activated by setting $l(x) = 1$. The third pays a penalty for activating a line process. The result encourages the estimated intensity to be piecewise smooth unless the input $I(x)$ changes significantly, in which case a line process is switched on and the smoothness is broken. The parameter T is the variance of the probability distribution and has a default value $T = 1$.

The Line Process Model Illustration

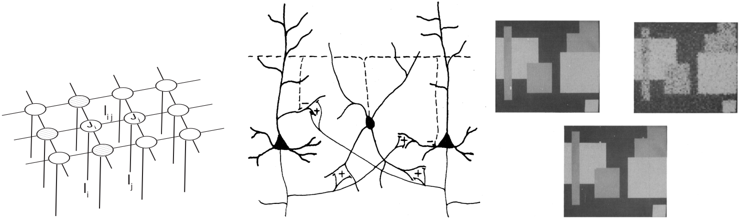

Figure 28: A representation of the Line-process model (far left) compared to a real neural network (left). On the right we show the original image (upper left), the image corrupted with noise (upper right) and the image estimated using the line-process model (bottom).

The Line Process Model and Neural Circuits (I)

This model can be implemented by a neural circuit [85]. The connections between these neurons is shown in figure (28). To implement this model [85] proposed a neural net model which is equivalent to doing mean field theory on the weak membrane MRF (as discussed earlier) by replacing the binary-valued line process variables $l(x)$ by continuous variables $q(x) \in [0,1]$ (corresponding roughly to the probability that the line process is switched on).
This gives an algorithm which updates the regional variables $\vec{J}$ and the line variables $\vec{q}$ in a coupled manner. It is helpful, as before, to introduce a new variable $\vec{u}$ which relates by $q(x) = \frac{1}{1+\exp\{-u(x)/T\}}$ and $u(x) = T \log \frac{q(x)}{1-q(x)}$.

The Line Process Model and Neural Circuits (II)

$$\frac{dJ(x)}{dt} = -2(J(x) - I(x))$$

$$= -2A\{(1 - q(x))(J(x) - J(x+1)) + (1 - q(x-1))(J(x) - J(x-1))\}, \quad (31)$$

$$\frac{dq(x)}{dt} = \frac{1}{T} q(x)(1 - q(x))\{A(J(x+1) - J(x))^2 - B - T \log \frac{q(x)}{1 - q(x)}\}, \quad (32)$$

$$\frac{du(x)}{dt} = -u(x) + A(J(x+1) - J(x))^2 - B. \quad (33)$$

The update rule for the estimated intensity $\vec{J}$ behaves like non-linear diffusion which smooths the intensity while keeping it similar to input $\vec{I}$. The diffusion is modulated by the strength of the edges $\vec{q}$. The update for the lines $\vec{q}$ is driven by the differences between the estimated intensity, if this is small then the lines are not activated.

The Line Process Model and Neural Circuits (III)

This algorithm has a Lyaponov function $L(\vec{J}, \vec{q})$ (derived using mean field theory methods) and so will converge to a fixed point, with

$$L(\vec{J}, \vec{q}) = \sum_x (I(x) - J(x))^2 + A \sum_x (J(x+1) - J(x))^2 (1 - q(x)) + B \sum_x q(x) + T \sum_x \{q(x) \log q(x) + (1 - q(x)) \log(1 - q(x))\}. \quad (34)$$

Relations to Electrophysiology (I)

There is some evidence that a generalization of this models roughly matches the electrophysiological findings for those types of stimuli shown in figure (33). The generalization is performed by replacing the intensity variables $I(x), J(x)$ by a filterbank of Gabor filters so that the weak membrane model enforces edges at places where the texture properties change [95]. The experiments, and their relation to the weak membrane models are reviewed in [96]. The initial responses of the neurons, for the first 80 msec, are consistent with the linear filter models described earlier. But after 80 msec the activity of the neurons change and appear to take spatial context into account.

While the weak membrane model is broadly consistent with the perceptual phenomena of segmentation and "filling-in", the types of filling-in, their dynamics, and the neural representations of contours and surface is complicated [167, 86]. Exactly how contour and surface information is represented and processed in cortex is an active topic of research [54, 142].

Relations to Electrophysiology (II)

The findings of the electrophysiological experiments are summarized as follows: (1) There are two sets of neurons where one set encodes regional properties (such as average brightness) and the other set codes boundary location (in agreement with J and l variable in the model respectively). (2) The processes for computing the region and the boundary representations are tightly coupled, with both processes interacting with and constraining each other (as in the dynamical equations above). (3) During the iterative process, the regional properties diffuse within each region and tend to become constant, but these regional properties do not cross the region (in agreement with the model). (4) The interruption of the spreading of regional information by boundaries results in sharp discontinuities in the responses across two different regions (in agreement with the model). The development of abrupt changes in regional responses also results in a gradual sharpening of the boundary response, reflecting increased confidence in the precise location of the boundary. These findings are roughly consistent with neural network implementations of the weak membrane model. But other explanations are possible. For example, the weak membrane model requires lateral (sideways) interaction and it is possible that the computations are done hierarchically using feedback from V2 to V1.

Relations to Electrophysiology Illustration

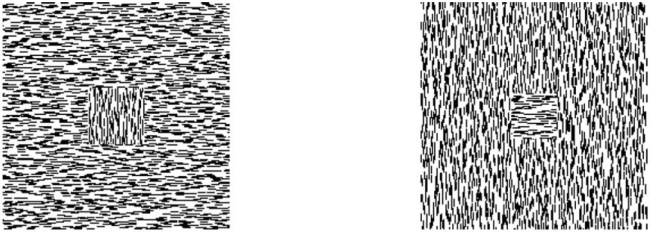

Figure 29: The stimuli for the experiments in the experiments by TS Lee and his collaborators [96].

Edge Detection with Spatial Context (I)

Our second example is to develop a model for detecting edges using spatial context. This relates to the phenomena known as association fields, see figure (26)(left panel), where Gabor filters which are spatially aligned (in orientation and direction) get grouped into a coherent form.
For this model, we have a set of neurons at every spatial position x, each tuned to a different angle $\theta_i : i = 1, ..., 8$, and a default cell at angle θ_0. The first cells are designed to detect edges at each orientation – i.e. they can be driven by the log-likelihood ratio of an edge detector at orientation θ_i at this position. The default cell is a dummy that is intended to fire if there is no edge present at this position. This organization forms a population of cells arrayed according to orientation (similar to a hypercolumn in V1). See figure (27)(right panel).

Edge Detection with Spatial Context (II)

We define a Gibbs distribution for the activity s_{x,θ_i} of the cells. The energy function $E(\vec{s})$ contains four types of terms: (I) A term $\sum_x \sum_{i=0}^{8} s_{x,i} \phi(f_1, ..., f_M)$. This term represents the local evidence for an edge at each point and for its orientation. (II) A term $\sum_x (\sum_{i=0}^{8} s_{x,i} - 1)^2$. This term is intended to ensure that only one cell is active at any spatial position. This corresponds to an inhibitory interaction between cells in the same hypercolumn. The cells in the hypercolumn give alternative, and inconsistent, interpretations of the input – hence only one of them can be correct. (III) A term that encourages edges to be continuous and for their directions to change smoothly. To define this term, we let $\vec{\theta_i} = (\cos\theta_i, \sin\theta_i)$ and $\vec{\theta_i^T} = (-\sin\theta_i, \cos\theta_i)$ denote the tangent to the edge and the normal. This term encourages there to be edges in the tangent direction, while the next term discourages them in the normal direction, see figure (27)(far right panel). This term is motivated by the intuition that curves are spatially smooth and can be justified by the statistics of natural images [38],[30].

Edge Detection with Spatial Context (III)

We write it as $\sum_{x,y} \sum_{i,j=1}^{8} W^{T}_{(x,\theta_i),(y,\theta_j)} s_{x,i} s_{y,j}$, where

$$W^{T}_{(x,\theta_i),(y,\theta_j)} = -\exp\{-|\vec{\theta}_i - \vec{\theta}_j|/K_1\} \exp\{-|x-y|/K_2\} \exp\{-|\hat{x}y - \vec{\theta}_i|/K_3\} \quad (35)$$

and ($\hat{x}y$ is the unit vector in direction $x - y$). This term encourages edges which are in similar directions (first term), nearby in position (second term), and where the edge orientation is similar to the difference $x - y$ between the two points. This term is excitatory. (IV) The final terms is inhibitory and discourages edges to be parallel to each other (if they are nearby). It is written as $\sum_{x,y} \sum_{i,j=1}^{8} W^{N}_{(x,\theta_i),(y,\theta_j)} s_{x,i} s_{y,j}$. Here

$$W^{N}_{(x,\theta_i),(y,\theta_j)} = \exp\{-|x-y|/K_4\} \exp\{-|\hat{x}y - \vec{\theta}_i^{T}|\} \quad (36)$$

Edge Detection with Spatial Context (IV)

The first term says this interaction decreases with distance. The second term discourages edges which are parallel to each other.
This gives an overall energy:

$$E(\vec{s}) = \sum_{x} \sum_{i=0}^{8} s_{x,i} \phi(f_1, ..., f_M) + \hat{K}_0 \sum_{x} (\sum_{i=0}^{8} s_{x,i} - 1)^2$$

$$+ \hat{K}_1 \sum_{x,y} \sum_{i,j=1}^{8} W^T_{(x,\theta_i),(y,\theta_j)} s_{x,i} s_{y,j} + \hat{K}_2 + \sum_{x,y} \sum_{i,j=1}^{8} W^N_{(x,\theta_i),(y,\theta_j)} s_{x,i} s_{y,j}. \qquad (37)$$

This yields a probability:

$$P(\vec{s}|\vec{f}) = \frac{1}{Z} \exp\{-E(\vec{s})\}.$$

This model can be implemented in neural networks by defining either stochastic or deterministic neural dynamics (i.e. either Gibbs sampling or mean field theory). The resulting update equations are more complex that those defined for our earlier examples but have the same basic ingredients. Models of this type can qualitatively account for associative field phenomena.

Stereo Models

This section introduces computational models for estimating depth by binocular stereo. The key problem is to solve the *correspondence problem* between the inputs in the two eyes to determine the *disparity*. Then the depth of the points in space can be estimated by trigonometry. (This pre-supposes that the eyes are *calibrated*, meaning that the distance between the eyes and the direction of gaze are known, which is beyond the scope of this chapter). Julesz [71] showed that humans could perceive depth from stereo if the images consisted of random dot stereograms which minimize the effect of feature similarity cues, suggesting that human vision can solve this task by relying mainly on geometric regularities (assumed about the structure of the world). Other researchers [18] have studied human estimation of surface shape quantitatively and showed, among other things, bias towards fronto-parallel surfaces.

Stereo: The Correspondence Problem

Most stereo algorithm address the correspondence problem by assuming that: (i) image features in the two eyes are more likely to correspond if they have similar appearance, (ii) the surface being viewed obeys prior knowledge such as being piecewise smooth (e.g., like the weak membrane model). The first assumption depends on local properties of the images while the second assumption uses non-local context. In an earlier section we discussed how a population of Gabor filters could be used to match local image features. In this section we describe how context can be used to impose prior knowledge about the geometry of the scene. We will study classic models which assume that the surface is piecewise smooth. This leads to a markov field model which includes excitatory connections, imposing the geometric constraints, with inhibitory connections which prevent points from one eye having more than one match in the second eye. This yields an algorithm which involves cooperation, to implement the excitatory constraints, and competition to deal with the inhibitory constraints. This is consistent with findings from recent electrophysiological experiments [146],[145]. These complement experiments [125] which tested the local stereo models described earlier.

A Cooperative Stereo Model (I)

We now specify a computational model for stereo which, for simplicity, we formulate in one-dimension. There is a long history of this type of model starting with the cooperative stereo algorithm [28, 110] and current computer vision stereo algorithms are mostly designed on similar principles.
We specify the left and right images by $\vec{I}_L, \vec{I}_R$ and denote features extracted from them by $\vec{f}(\vec{I}_L) = \{f(x_L) : x_L \in \mathcal{D}_L\}$, $\vec{f}(\vec{I}_R) = \{f(x_R) : x_R \in \mathcal{D}_R\}$. We define a discrete-valued correspondence variable $V(x_L, x_R)$, so that $V(x_L, x_R) = 1$ means that the features at x_L, x_R in the two images correspond, and hence the disparity is $x_L - x_R$. If the features do not match then we set $V(x_L, x_R) = 0$. We encourage all data-points to match one, but allow some datapoints to be unmatched and others to match more than once (by paying a penalty).

A Cooperative Stereo Model (II)

We specify a distribution $P(\vec{V}|\vec{f}(\vec{I}_L), \vec{f}(\vec{I}_R)) = \frac{1}{Z}\exp\{-E(\vec{V}; \vec{f}(\vec{I}_L), \vec{f}(\vec{I}_R))/T\}$, where the energy $E(\vec{V}; \vec{f}(\vec{I}_L), \vec{f}(\vec{I}_R))$ is given by:

$$\begin{aligned} E(\vec{V}; \vec{f}(\vec{I}_L), \vec{f}(\vec{I}_R)) = \sum_{x_L, x_R} V(x_L, x_R) M(f(x_L), f(x_R)) \\ + A \sum_{x_L} (\sum_{x_R} V(x_L, x_R) - 1)^2 + A \sum_{x_R} (\sum_{x_L} V(x_L, x_R) - 1)^2 \\ + C \sum_{x_L, x_R} \sum_{y_L \in N(x_L)} \sum_{y_R \in N(x_R)} V(x_L, x_R) V(y_L, y_R) \{(x_R - x_L) - (y_R - y_L)\}^2. \end{aligned} \tag{38}$$

A Cooperative Stereo Model (III)

The first term imposes that there are matches between image points with similar features, here $M(.,.)$ is a measure which takes small values if $f(x_L), f(x_R)$ are similar and large values if they are different. We will discuss at the end of this section how $M(f(x_L), f(x_R))$ relates to model for local stereo discussed earlier. The second two terms penalize image points which are either unmatched, or are matched more than once. The third term encourages the disparities, $x_L - x_R$, to be similar for neighboring points (here $N(.)$ defines a spatial neighborhood as before). These models can be applied to two-dimensional images by solving the correspondence problem for each epipolar line separately (by maximizing $P(\vec{V}|\vec{f}(\vec{I_L}), \vec{f}(\vec{I_R})))$. This is shown in figure (30)(right panel). The parameter T is the variance of the model, as for the line process model, and has default value $T = 1$.

A Cooperative Stereo Model Illustration

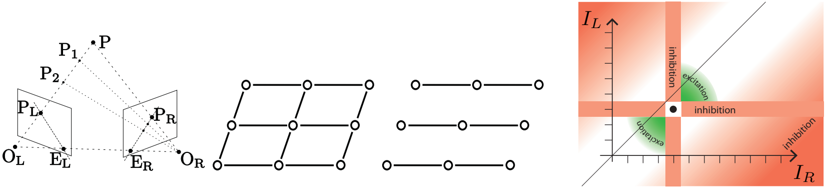

Figure 30: Stereo. The geometry of stereo (left). A point P in 3-D space is projected onto points PL; PR in the left and right images. The projection is specified by the focal points OL,OR and the directions of gaze of the cameras (the camera geometry). The geometry of stereo enforces that points in the plane specified by P,OR OL, must be projected onto corresponding lines EL;ER in the two images (the epipolar line constraint). If we can find the correspondence between the points on epipolar lines then we can use trigonometry to estimate their depth, which is (roughly) inversely proportional to the disparity, which is the relative displacement of the two images. Right Panel: binocular stereo requires solving the correspondence problem which involves excitation (to encourage matches with similar depths/disparities) and inhibition (to prevent points from having multiple matches).

A Cooperative Stereo Model (IV)

We obtain a neural circuit model by performing mean field theory on $P(\vec{V}|\vec{f}(\vec{I_L}), \vec{f}(\vec{I_R}))$. This replaces $V(x_L, x_R) \in \{0, 1\}$ by continuous-valued $q(x_L, x_R) \in [0, 1]$ and an associated variable $u(x_L.x_R) = T \log \frac{q(x_L,x_R)}{1-q(x_L,x_R)}$ with $q(x_l, x_R) = \frac{1}{1+\exp\{-u(x_L,x_R)\}}$.
The update equation is:

$$\frac{du(x_L, x_R)}{dt} = -u(x_L, x_R) - M(f(x_L), f(x_R))$$
$$-2A(\sum_{y_R \neq x_R} q(x_L, y_R) - 1) - 2A(\sum_{y_L \neq x_L} q(y_L, x_R) - 1),$$
$$-2C \sum_{y_L \in N(x_L)} \sum_{y_R \in N(x_R)} q(y_L, y_R)\{(x_R - x_L) - (y_R - y_L)\}^2. \tag{39}$$

This update includes the standard integration term (first term) and the second term encourages matches where the features agree. There is also inhibition between competing matches (the third and fourth term), and excitation for matches which are consistent with a smooth surface (last term).

A Cooperative Stereo Model: Interactive Demo.

There is a variant of this algorithm which is used in interactive demo (4c). This algorithm is a discrete Hopfield network which attempts to minimize the energy $E(\vec{V}; \vec{f}(\vec{I}_L), \vec{f}(\vec{I}_R))$ in equation (38). The algorithm starts by assigning initial values, 0 or 1, to each state variable $V(x_L, x_R)$. The algorithm proceeds by selecting a state variable, changing its value (e.g., changing $V(x_L, x_R) = 1$ to $V(x_L, x_R) = 0$), calculating if this change reduces the energy $E(\vec{V}; \vec{f}(\vec{I}_L), \vec{f}(\vec{I}_R))$, and keeping the change if it does. This process repeats until the algorithm converges (i.e., all possible changes raise the value of the energy).

A Cooperative Stereo Model and the Local Model

How does the cooperative stereo algorithm relate to our earlier algorithm for computing stereo disparity locally? Recall that the algorithm estimated the disparity at a single point by having a set of neurons which were tuned to different disparities $\{D_i : i = 1, ..., N\}$, summing the votes $v(D_i)$ for each disparity by equation (14),and selecting the disparity with most votes. Using the cyclopean coordinate system [71], we express the disparity by $D(x) = \frac{1}{2}(x_R - x_L)$ where $x = \frac{1}{2}(x_R + x_L)$. At each point x we specify a population of neurons which encode the votes $v(D(x))$ for the different disparities. Then, instead of using winner-take-all to make a local decision, we feed the responses $v(D(x))$ back into cooperative stereo algorithm by defining $M(f(x_L), f(x_R)) = \exp\{-v(\frac{1}{2}(x_R - x_L))\}$ (the negative exponential $\exp\{-\}$ is required so the $M(f(x_L), f(x_R))$ is small if the vote for disparity $D(x) = \frac{1}{2}(x_R - x_L)$ is large).

A Cooperative Stereo Model and Electrophysiology

Analysis of electrophysiological studies [146],[145] were in general agreement with the predictions of this type of stereo algorithm. In particular, studies showed that neural populations responses included excitation between cells tuned to similar disparities at neighboring spatial positions and inhibition between cells tuned to different disparities at the same position, see figure (31). In addition, Samonds *et al.* [147] implemented a variant of the stereo algorithm described above and showed that it could account for additional phenomena such as sharper tuning to the disparity for larger stimuli and performance on anti-correlated stimuli (where the left and right images have opposite polarity).

A Cooperative Stereo Model and Electrophysiology Illustration

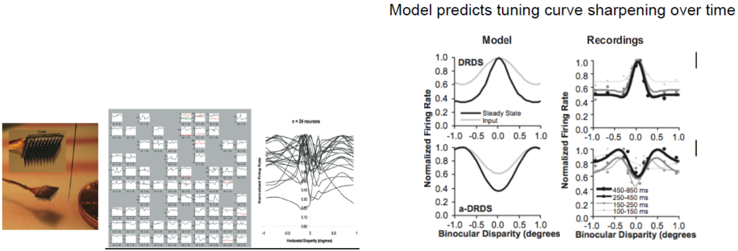
Model predicts tuning curve sharpening over time

Figure 31: Experiments for testing stereo algorithms [146],[145]. Left Panel: the experimental setup. Right Panel: the experiments give evidence for excitation between similar disparity and inhibition to prevent multiple matches.