Contents

13
Cortical networks of visual recognition

Christian Thériault, Nicolas Thome, Matthieu Cord

13.1
Introduction

Human visual recognition is a feat of nature which is far from trivial. The light patterns projected on the human retina are always changing, and never an object will create twice the exact same pattern. Objects move and transform constantly, appearing under an unlimited number of aspects. And yet, the human visual system can recognize objects in milliseconds. It is thus natural for computer vision models to draw inspiration from the human visual cortex.

Most of what is understood about the visual cortex, and particularly how it is able to achieve object recognition, comes from neuro-physiological and psychophysical studies. A global picture emerges from over six decades of studies — the visual cortex is mainly organized as a parallel and massively distributed network of self-repeating local operations. Neuro-physiological data and models of cortical circuitry have shed light on the processes by which feedforward (bottom-up) activation can generate the early neural response of visual recognition [1, 2]. Contextual modulation and attentional mechanisms, through lateral and cortical feedback (top-down) connections, are clearly essential to the full visual recognition process [3, 4, 5, 6, 7, 8]. Nevertheless, basic feedforward models [9, 10, 11], without feedback connections, already display interesting levels of recognition, and provide a simple design around which the full functioning of the visual cortex can be studied. The circuitry of the visual cortex has also been studied in the language of differential geometry, which provides a natural connection between local neural operations, global activation and perceptual phenomena [12, 13, 14, 15, 16].

This chapter introduces basic concepts on visual recognition by the cortex and some of its models. The global organization of the visual cortex is presented in section 13.2. Local operations are introduced in section 13.3, followed in section 13.4 by a special emphasis on operations in the *primary visual cortex*. Object recognition models are presented in section 13.5, with a detailed description of a general model in section 13.6. Section 13.7 focuses on a mathematical abstraction which corresponds to the structure of the primary visual cortex and which provides a model of contour emergence. Section 13.8 presents psychophysical and biological bases suporting such a model.The importance of feedback connection are discussed in section 13.9, and the chapter concludes with the role of transformations (i.e., motion) in learning invariant representations of objects.

13.2
Global organization of the visual cortex

The brain is a dynamical system in which specialized areas receive and send connections to multiple other areas — brain functions emerge from the interaction of subpopulations of specialized neurons. As illustrated in figure 13.1, the visual cortex makes no exception, and contains interacting subpopulations of neurons tuned to process visual information (shape, color, motion, etc.).

The laminar structure of the cortex enables researchers to distinguish between *lateral connections* inside each area, *feedback connections* between areas, and *feedforward connections* streaming up from retinal inputs [17]. Up to now, the two most studied cortical visual areas have been the V1 area, the first area receiving extracortical inputs from the lateral geniculate nucleus (LGN), and the V5/MT (Medial Temporal) area.

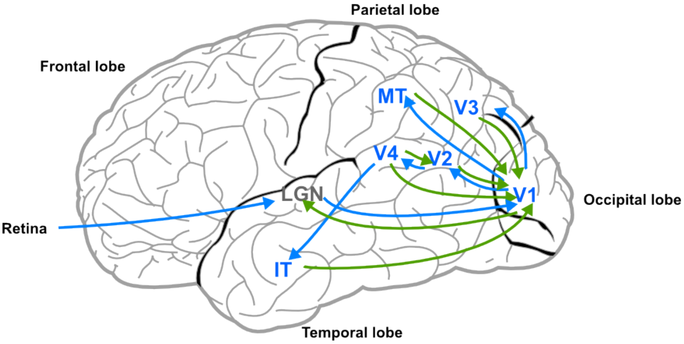

Figure 13.1 Basic organization of the visual cortex. Arrows illustrate the connections between various specialized subpopulations of neurons. Green and blue arrows indicate feedback and feedforward connections respectively. The lateral geniculate nucleus (LGN), an extracortical area, is represented in gray.

A general consensus is that neurons in V1 behave as spatio-temporal filters, selective to local changes in color, spatial orientations, spatial frequencies, motion direction and speed [18, 19]. The V5/MT area shares connections with V1 and also contains populations of neurons sensitive to motion speed and direction (see chapter **??**). Its contribution to motion perception beyond what is already observed in V1 is still not fully determined [20] and may involve the processing of motion over a broader range of spatio-temporal patterns compared to V1 [21, 22, 18].

Beginning at V1 there are interacting streams of visual informations — the *ventral pathway*, which begins at V1 and goes into the temporal lobe, and the *dorsal pathway*, which also begins at V1, but goes into the parietal lobe [23]. The latter is traditionally associated with spatial information (i.e., where) while the former is traditionally associated with object recognition (i.e.,what). However, mounting evidences indicate significant connections between the two pathways and the dynamics of their interaction is now recognized [17, 24, 25, 26, 27].

Feedforward activation in the ventral pathway is correlated with the rapid (i.e.,$\sim$100-200ms) ability of humans to recognize visual objects [1, 28], but recurrent feedback activity between these areas should not be ruled out, even during rapid recognition [29]. The role of later stages beyond V1, (i.e.,V2,V3 and V4) remains less understood, although neurons in the IT (Inferior Temporal) area have been shown to respond to more complex and global patterns, independently of spatial position and pose (i.e., full objects, faces, etc.) [23, 1].

13.3
Local operations: receptive fields

Neurons in the visual cortex follow an *retinotopic* organization — local regions of the visual field, called *receptive fields*, are mapped to corresponding neurons [30]. In some areas, such as V1, the retinotopic mapping is continuous [31] — adjacent neurons have adjacent overlapping receptive fields. At different stages in the visual pathway neurons have receptive fields which vary in size and complexity [23]. As presented in section 13.4, neurons in V1 responds to local changes (i.e., derivative of light intensity) over orientations and spatio-temporal frequencies. These small local receptive fields are integrated into larger receptive fields by neurons in the later stages of the V1$\rightarrow$IT pathway (figure 13.2). This gives rise to representations of more complex patterns along the pathway, such as in the V4 [32, 33]. In the latest stages, such as the IT area, neurons have receptive sizes which cover the entire visual field and respond to the identity of objects rather than their position [34, 1].

13.4
Local operations in V1

Together with the work of [35], neuro-physiological studies [36, 37, 38] have established a common finding about the spatial domain profile of receptive fields in the V1 area. This visual area is organized into a tilling of *hypercolumns*, in which the columns are composed of cells sensitive to spatial orientations with ocular dominance. Brain imaging techniques [39] have since revealed that orientation columns are spatially organized into a pinwheel crystal as illustrated in figure 13.3.

Cells inside orientation columns are referred to as *simple cells*. In chapter **??**,

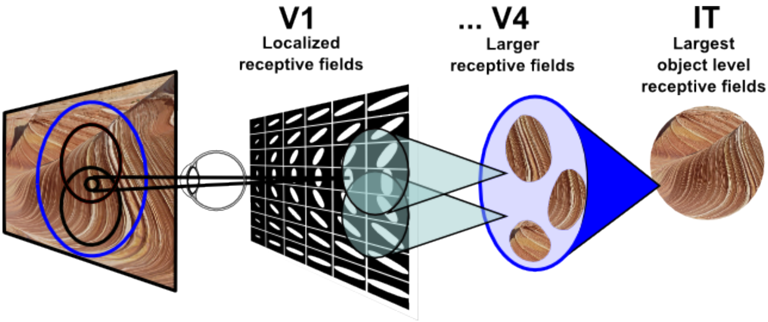
V1
Localized receptive fields
... V4
Larger receptive fields
IT
Largest object level receptive fields

Figure 13.2 Receptive field organization. Neurons along the visual pathway V1→IT have receptive fields which vary in size and complexity.

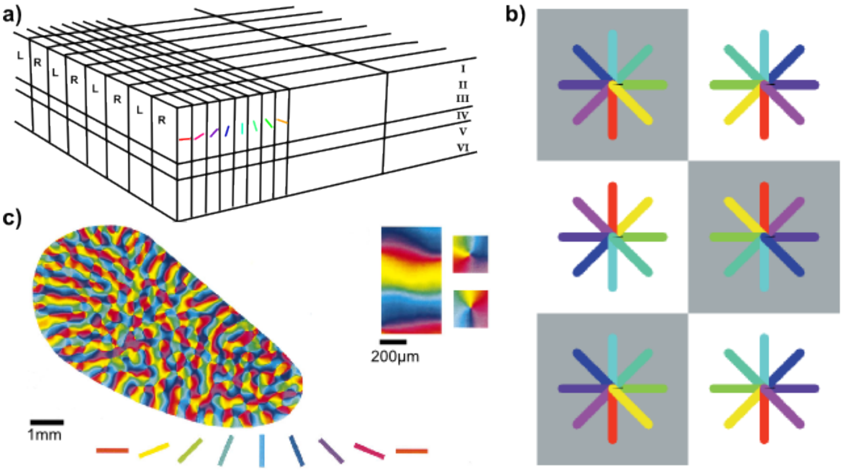

Figure 13.3 Primary visual cortex organization. a) Hypercolumn structure showing the orientation and ocular dominance axes (image reproduced from [40]). b) Idealized crystal pinwheel organization of hypercolumns in visual area V1 (image taken from [14]). c) Brain imaging of visual area V1 of the Tree Shrew showing the pinwheel organization of orientation columns (image adapted from [39]).

the conditions under which simple cells profile naturally emerge from unsupervised learning are presented. The profile of simple cells can be modeled by Gaussian modulated filters — Gaussian derivatives [41, 42] or Gabor filters [43, 44, 45]. Gaussian derivatives describe the V1 area in terms of differential geometry [41] whereas Gabor filters present the V1 area as a spatio-temporal frequency analyzer [46, 37]. Both filters are mathematically equivalent and differ mostly in terminology. As illustrated in figure 13.4, the 1^{st} and 2^{th} order Gaussian derivatives give good approximations of Gabor filters with odd and even phase respectively. In fact, Gaussian derivatives are asymptotically equal (for high order derivatives) to Gabor filters [41]. The one-dimensional profile for the Gaussian derivatives and the (odd phase) Gabor filter are respectively given by

$$f(x) = \frac{\partial^n}{\partial x^n} G(x) \quad \text{and} \quad g(x) = G(x)\sin(2\pi\omega x) \tag{13.1}$$

where $G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$ is the Gaussian envelope of scale σ, and where ω gives the spatial frequency of the Gabor filter. In [47], derivatives are further normalized with respect to σ to obtain scale invariance — filters at different scales will give the same maximum output. The even phase Gabor filters is simply defined by a cosine function instead of a sine. In [42], cortical receptive field with order of differentiation as high as $n = 10$ are reported, with the vast majority being $n \leq 4$.

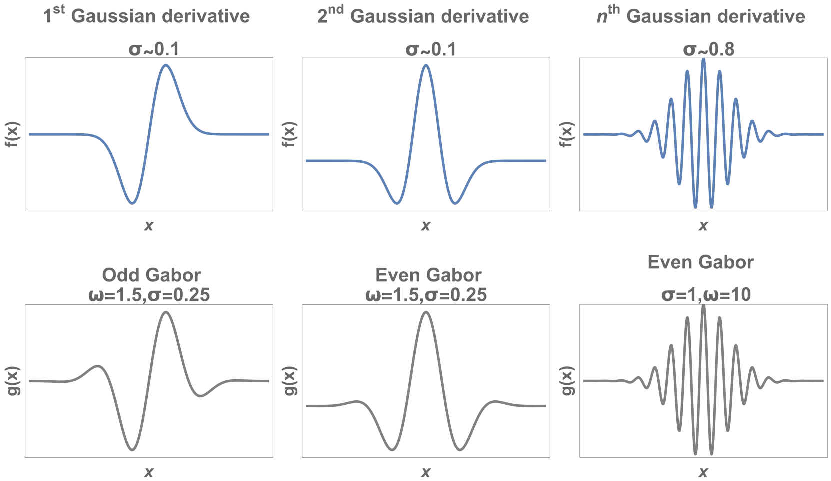

Figure 13.4 One dimensional simple cells profiles. The curves shown are the graphs of local operators modeling the sensitivity profile of neurons in the primary visual cortex. Gabor filter and Gaussian derivative are local approximations of each other and are asymptotically equivalent for high order of differentiation.

Two-dimensional filters for the first order Gaussian derivative and the odd phase Gabor are respectively given by

$$f(x, y) = \frac{\partial}{\partial y} G(x, y) \quad \text{and} \quad g(x, y) = G(x, y)\sin(2\pi\omega y). \tag{13.2}$$

where the Gaussian envelope is $G(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}}$. For $u = x\cos\theta + y\sin\theta$ and $v = -x\sin\theta + y\cos\theta$, the filters $f_{\sigma,\theta}(u, v)$ and $g_{\sigma,\theta}(u, v)$ correspond to a rotation of the axes by an angle θ at scale σ, and model the orientation selectivity of simple cells in the hypercolumns. Specifically, when applied to an image I, the Gaussian derivative $f_{\sigma,\theta}(u, v)$ gives the directional derivative, in the direction $v = (-\sin\theta, \cos\theta)$, of the image smoothed by the Gaussian G at scale σ

$$f_{\sigma,\theta} * I = [-\sin\theta \frac{\partial}{\partial x} + \cos\theta \frac{\partial}{\partial y}] G * I \tag{13.3}$$

where $*$ denotes the convolution product. The odd phase Gabor $g_{\sigma,\theta}$ gives the same directional derivative up to a multiplicative constant. Figure 13.5 illustrates examples of 1^{st} order Gaussian derivatives and even phase Gabor filters at various scales and orientations. When representing cell activations in V1, the outputs of these filters can then be propagated synchronously or asynchronously [48] through a multilayer architecture simulating the basic feedforward principles of the visual cortex, as in section 13.6.

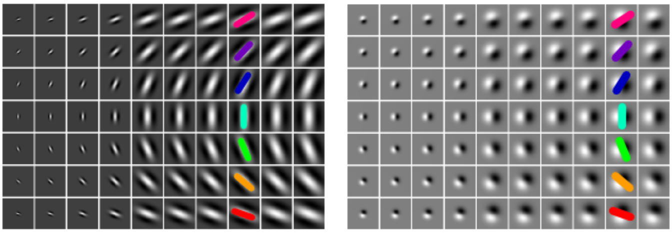

Figure 13.5 Two dimensional simple cells profiles. The figure illustrates two-dimensional Gabor filters (left) and 1^{st} order Gaussian derivatives (right) at various scales and orientations. Colors can be put in relation with figure 13.3.

Neuro-physiological studies also identified *complex cells* in the V1→V4 pathway. These cells also respond to specific frequencies and orientations, but their spatial sensitivity profiles are not localized like the simple cells, as illustrated in figure 13.6. Complex cells display invariance (tolerance) to the exact position of the visual patterns at the scale of the receptive field. By allowing local shifts in the exact position of patterns, they may play a role in our ability to recognize objects invariantly with respect to transformations. Complex cells can be modeled by a MAX or soft-MAX operations applied to incoming simple cells [49]. Section 13.6 presents a model of such simple and complex cells network using the MAX operation in a multilayer architecture.

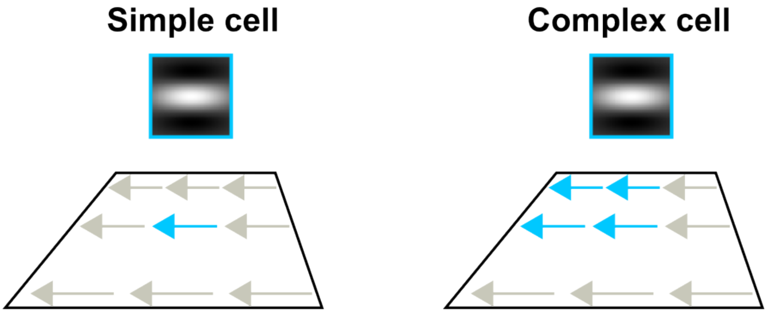

Figure 13.6 Complex cells. On the left, the simple cell responds selectively to a local spatial derivative (blue arrow). The complex cells responds selectively to the same spatial pattern, but displays invariance to its exact position. Complex cells are believed to gain their invariance by integrating (summing) over simple cells of the same selectivity.

13.5 Multilayer models

Based on the above considerations, the vast majority of biologically inspired models are multilayer networks where the layers represent the various stages of processing corresponding to physiological data obtained about the mammalian visual pathways.

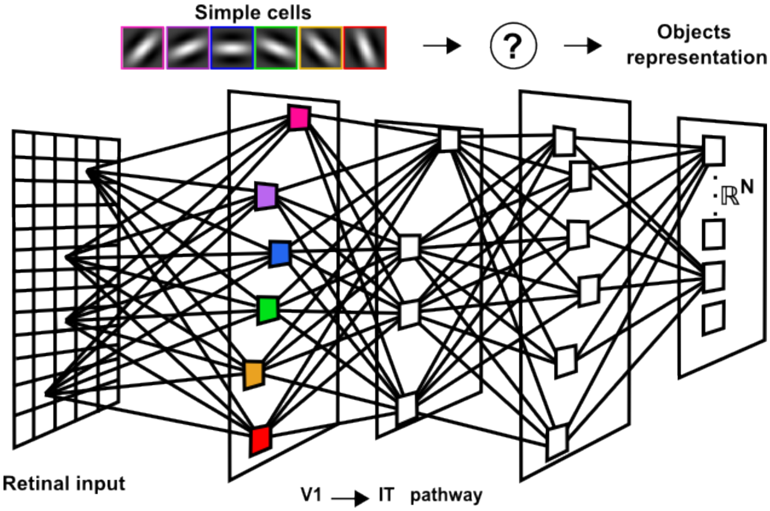

Figure 13.7 Multilayer architecture. The cortical visual pathway is usually modeled by multilayer networks. Synaptic connections for the early stages, such as V1 (first layer), are usually sensitive to local changes in spatial orientations and frequencies. Connections in the latter stages is an active research topic.

For most multilayer networks, the first layers are modeled by the operations of *simple cells* and *complex cells* in the primary visual cortex V1, as presented in the preceding section. However, the bulk of neuro-physiological data clearly indicate important areas for visual recognition beyond area V1. These stages of processing are less understood. Multilayer networks usually model these stages with a series of layers which generate increasing complexity and invariance of representations [9, 11] (figure 13.7). These representations are most often learned from experience by defining learning rules which modify the connections between layers.

Supervised learning can be used with the objective to minimize error in object representations while retaining some degree of invariance to basic visual transforms [50, 51]. Networks using supervised learning are most often based on the error gradient *backpropagation* through the layers [52].

Unsupervised learning uses statistical regularities of the visual world to learn useful representations. One basic principle is to use temporal correlations between views of a transforming object. The hypothesis, in this case, is that we don't change the world by moving in it. Or put differently, neural representations should change on a slower time scale than the retinal input [53]. Such networks [23, 54, 55] are often based on the so called *trace rule* which minimizes variations of neural response to objects undergoing transformations. Other authors [48] have shown that the temporal aspect of neural spikes across layers can be used to define unsupervised learning of relevant object features. Unsupervised learning has also been used in *generative models* to find the appropriate representation space for each layer as the initial condition for supervised learning [56, 57].

13.6 A basic introductory model

A simple and yet efficient model of the basic feedforward mechanisms of the visual pathway V1→IT is given by the HMAX network and its variations [9, 10, 11]. In its most basic form, this network does not model the full connectivity of the visual cortex such as, for instance, the long-range horizontal connections described in the next section — the basic HMAX model is in essence a purely feedforward network. Nevertheless, it is based on a hypercolumns organization and provides a starting point to model the visual cortical pathway.

The layers of the HMAX model are decomposed into parallel sublayers called *feature maps*. The activation of one feature map corresponds to the mapping of one filter (one feature) at all positions of the visual field. On the first layer, the features are defined by the ouputs of simple cells given by equation 13.2. Each simple cell calculates a directional derivative, and corresponds geometrically to an orientation and an amplitude (i.e., a vector). The feature maps on the first layer of the HMAX are therefore vector fields, corresponding to cross-sections of the hypercolumns in V1 — the selection of one directional derivative per position is, by definition, a vector field expressing the local action of a transformation.

As illustrated in figure 13.8, the first layer in HMAX enables the expression of various *group actions* and particularly those corresponding to basic transformations imposed on the retinal image, and for which visual recognition is known to be invariant. Pooling over the parameters of a transformation group (i.e., pooling over translations) generates representations which are invariant to such transformation [58]. In the basic HMAX, mapping and pooling over translation fields of multiple orientations gives the model some degree of invariance to local translations, and consequently to shape deformations.

The overall HMAX model follows a series of convolution/pooling steps as in [9, 59] and illustrated in figure 13.9. Each convolution step yields a set of feature maps and each pooling step provides tolerance (invariance) to variations in these feature maps. Each step is detailed below.

Layer 1 (simple cells). Each feature map $\mathbf{L1}_{\sigma,\theta}$ is activated by the convolution of the input image with a set of simple cell filters $g_{\sigma,\theta}$ with orientations θ and scales σ as defined in equation 13.2. Given an image I, Layer 1 at orientation θ and scale σ is given by the absolute value of the convolution product

$$\mathbf{L1}_{\sigma,\theta} = |g_{\sigma,\theta} * I|. \tag{13.4}$$

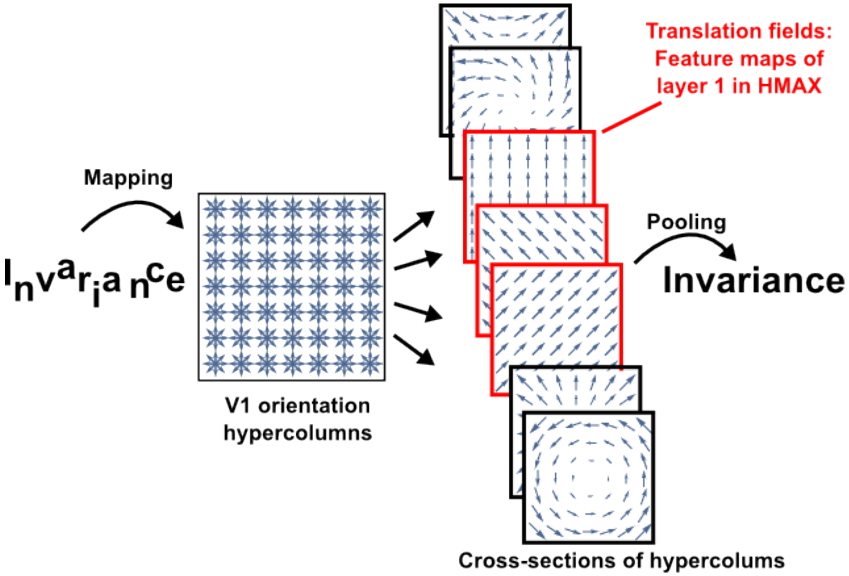

Figure 13.8 Sublayers in HMAX. The figure illustrates cross-sections of the orientation hypercolumns in V1. Each cross-section selects one simple cell per position, thereby defining a vector field. The first layer of HMAX is composed of sublayers defining translation fields (i.e., same orientation at all positions). By pooling over translation fields at various orientations, the HMAX model displays tolerance to local translations, which results in degrees of invariance to shape distortions. For illustration, the figure shows the image of the word *Invariance*, with its component translated. By pooling the local maximum values of translation fields mapped onto the image, the HMAX model tolerates local translations, and will produce a representation which is invariant to those local transformations.

The layer can further self-organized [60] through a process called *lateral inhibition* which appear through out the cortex. Lateral inhibition is the process by which a neuron suppresses the activation of its neighbors through inhibitory connections. This competition between neurons creates a form of refinement or sharpening of the signal by filtering out components of smaller amplitude relatively to their surrounding. It is also considered as a biological mechanism for neural sparsity and corresponds to a *subtractive normalization* [61]. Sparse neural firing has been shown to improve the original HMAX architecture on classification simulations [62]. The effect of inhibitory connections between neighboring neurons can be implemented by taking the convolution product of maps in layer 1 with the filter defined in equation 13.5 — an inhibitory surround with an excitatory center. A form of *divisive normalization* [63, 64] can also used (see layer 3 below). More refinement can be obtained by applying inhibition (or suppressing) the weaker orientations at each position [11, 10]. This can be accomplished by a one dimensional version of the lateral inhibition filter (figure 13.10) in equation 13.5.

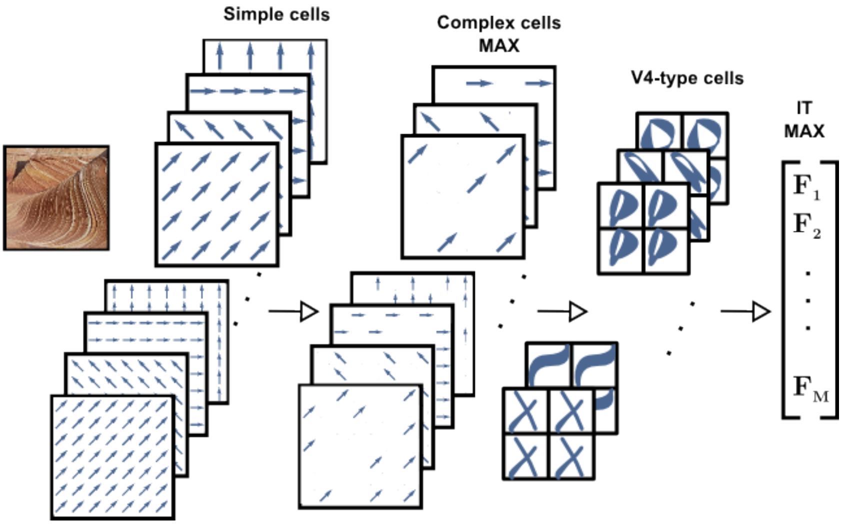

Figure 13.9 General HMAX network. The network alternates layers of feature mapping (convolution) and layers of feature pooling (MAX function). The convolution layers generate specific feature information whereas the pooling layers results in degrees of invariance by relaxing the configuration of these features.

$$I(x,y) = \begin{cases} \delta e^{-\frac{x^2+y^2}{\sigma^2}} & : x \neq 0, y \neq 0 \\ 1 & : x = 0, y = 0 \end{cases} \tag{13.5}$$

where σ defines the width of inhibition and δ defines the contrast.

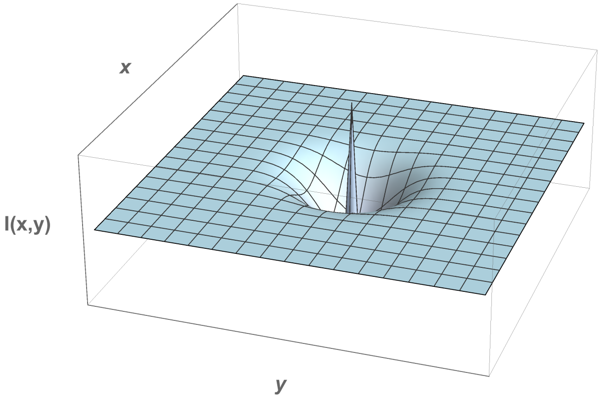

Figure 13.10 Lateral inhibition filter. A surround inhibition and an excitatory center.

Layer 2 (complex cells). Each feature map $\mathbf{L2}_{\sigma,\theta}$ models the operations of complex cells in the visual cortex, illustrated in figure 13.6. The output of complex cell selects the maximum value on a local neighborhood of simple cells. Maximum pooling over local neighborhoods results in invariance to local translations and thereby to global deformations [50] — It provides *elasticity* (illustrated in figure 13.8) to the configuration of features of layer 1. Specifically, the second layer partitions each $\mathbf{L1}_{\sigma,\theta}$ map into small neighborhoods $\mathbf{u}_{i,j}$ and selects the maximum value inside each $\mathbf{u}_{i,j}$ such that

$$\mathbf{L2}_{\sigma,\theta}(i,j) = \max_{\mathbf{u}_{i,j} \in \mathbf{L1}_{\sigma,\theta}} \mathbf{u}_{i,j}. \tag{13.6}$$

Some degree of local scale invariance is also achieved by keeping only the maximum output over two adjacent scales at each position (i, j).

Layer 3 (V4-type cells). Layer $\mathbf{L3}$ at scale σ is obtained by applying filters $\boldsymbol{\alpha}^m$ to layer 2.

$$\mathbf{L3}^m_\sigma = \boldsymbol{\alpha}^m * \mathbf{L2}_\sigma. \tag{13.7}$$

Each $\boldsymbol{\alpha}^m$ filter represents a V4-type cell — a configuration of multiple orientations representing more elaborated visual patterns than simple cells. In [11], the $\boldsymbol{\alpha}^m$ also cover multiple scales, which give each filter the possibility of responding selectively to more complex patterns.

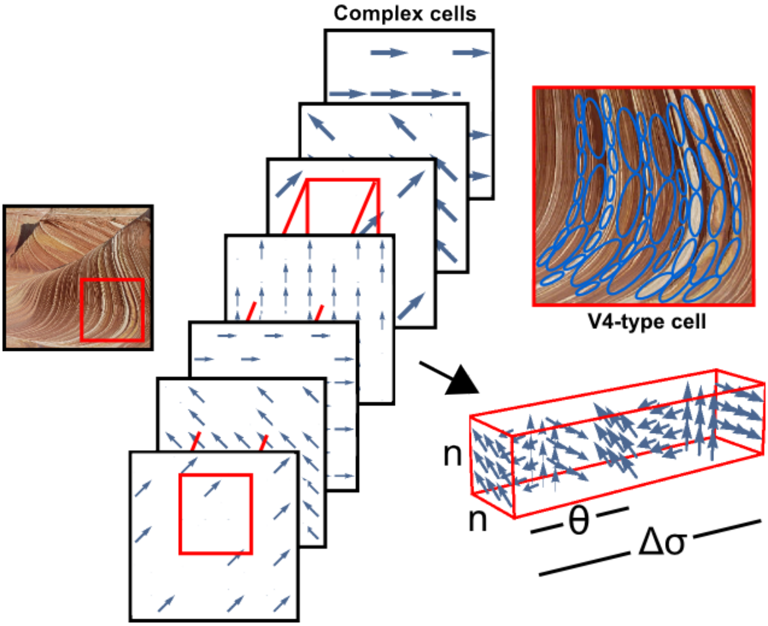

Figure 13.11 V4-type cells. On layer 3 the cells represent configurations of complex cells, of various orientations and scales, sampled during training. The ellipses represent the receptive fields of simple cells on the first layer.

To implement a form of divisive normalization [63], the filters, and its input, can be normalized to unit length. This normalization will ensure that only the geometrical aspect of the inputs are considered, and not the contrast or the luminance level. The filters $\boldsymbol{\alpha}^m$ are learned from training images by sampling subsets of layer $\mathbf{L2}$. This simple learning procedure consists in sampling subparts of layer $\mathbf{L2}$ and storing them as the connections weights of the filters $\boldsymbol{\alpha}^m$. The training procedure, as done in [11], is illustrated in figure 13.11. In the figure, a local subset of spatial size $n \times n$ is selected from layer $\mathbf{L2}$. The sample covers all orientations θ and a range of scales $\Delta\sigma$. To make the filter more selective, one possibility is to keep only the strongest coefficient in scales and orientations at each position, setting all other coefficients to zero. This sampling is repeated for a total of M filters. In [58], it is shown that for a system invariant to various geometrical transformations, such as the HMAX, this type of learning requires less samples in order to display classification properties.

Layer 4 (IT-type cells). To gain global invariance, and to represent patterns of neural activation in higher visual areas such as IT, the final representation is activated by pooling the maximum output of $\mathbf{L3}^m_\sigma$ across positions and scales. Pooling over the entire visual space, spatially and over all scales, guaranties invariance to position and size. However, important spatial and scale information for recognition might be lost by such global pooling.

In [11], to maintain some spatial and scale information in the final representation, a set of concentric pooling regions is established around the spatial position and the scale at which each filter was sampled during training. As shown in figure 13.12, each pooling region is defined by a radius R_i centered on the training position (x_m, y_m) and covers scales in the range $\sigma^m \pm 1$.

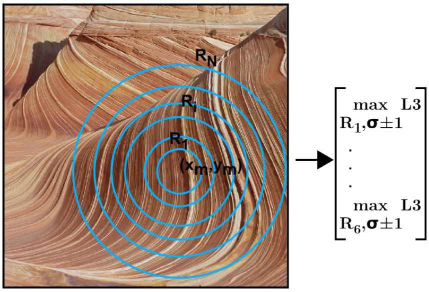

Figure 13.12 Multi-resolution pooling. The training spatial position and scale of each filter is stored during training. For each new image, concentric pooling regions are centered on this coordinate. The maximum value is pooled from each pooling radius and across scales $\sigma^m \pm 1$. This ensures that some spatial and scale information are kept in the final representation.

This pooling procedure is applied for all M filters and the results are concatenated into a final $\mathbf{L4}$ vector representation (equation 13.8). Each element of the $\mathbf{L4}$ vector represents the maximum activation level of each filter inside each search region. One hypothesis discussed in [1], is that local selectivity, combined with invariant pooling, may be a general principle by which the cortex performs visual classification. A combination of selectivity and invariance (i.e., relaxation of spatial configurations by maximum pooling), gained locally on each layer of the network, progressively maps the inputs to a representation space (i.e., a vector space), where the visual classes are more easily separable—where the class manifolds are *untangled*. As such, the final vector representation of equation 13.8 generates a space inside which classification of complex visual inputs can be performed relatively well with a vector classifier [11].

$$\mathbf{L4} = \begin{bmatrix} \max\limits_{\mathbf{R}_1,\, \sigma^1 \pm 1} \mathbf{L3}^1 \\ \vdots \\ \max\limits_{\mathbf{R}_1,\, \sigma^M \pm 1} \mathbf{L3}^M \\ \max\limits_{\mathbf{R}_6,\, \sigma^1 \pm 1} \mathbf{L3}^1 \\ \vdots \\ \max\limits_{\mathbf{R}_6,\, \sigma^M \pm 1} \mathbf{L3}^M \end{bmatrix}. \qquad (13.8)$$

The operations of layer 1 of the basic HMAX network presented above is a highly simplified, and not entirely faithful, model of the hypercolumns organization in the primary visual cortex. For one, the actual hypercolumns structure is not explicitly represented, it is only implied by its decomposition into translation vector fields at multiple orientations. Also, there are no long-range horizontal connections between the hypercolumns corresponding to known neuro-physiological data. The next section presents an idealized mathematical model of V1 which represents the hypercolumns explicitly. This model gives the hypercolumns of V1 a one-to-one correspondence with a mathematical structure and gives a formal expression of its horizontal connections.

13.7
Idealized mathematical model of V1: Fiber Bundle

There exist a mathematical model of visual area V1 which gives a theoretical formulation of the way the local operations of simple cells defined by equation 13.2 can merge into global percepts, and more precisely, into visual contours. As seen in the previous section, the HMAX is founded on a simplified model of the hypercolumns of V1. However, neuro-physiological studies, clearly identified the important role, in the perception of shapes, of horizontal long-range connections between the hypercolumns [65]. When taken individually, the hypercolumns define the local orientations in the visual field. But how can these local orientations dynamically interact such that a global percept emerge ? The language of differential geometry provides a natural answer to this question. Indeed, there is a mathematical structure which, at a certain level, gives an abstraction of the physical structure of the primary visual cortex. It also provides, in the spirit of Gestalt psychology, a top-down definition of the way global shapes are generated by the visual system (13.13).

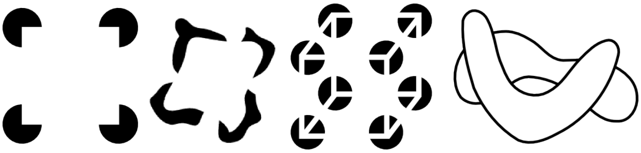

Figure 13.13 Contour completion. Understanding the principles by which the brain is able to spontaneously generate or complete visual contours, sheds light on the importance of top-down and lateral processes in shape recognition (image reproduced from [66]).

To understand the abstract model of V1 described below, one can first note the similarity between the retinotopic organization of overlapping receptive fields on the retinal plane and the mathematical structure of a *differentiable manifold* — a smooth structure which locally looks like the euclidean plane $\mathbb{R}^2$. One can also note that the repetition of hypercolumns in V1 provide a copy of the space of orientations over each position of the retinal plane. As introduced in the seminal work of Hoffman [67, 68, 12] and illustrated in figure 13.14, this structure is the physical realization of a *fiber bundle*.

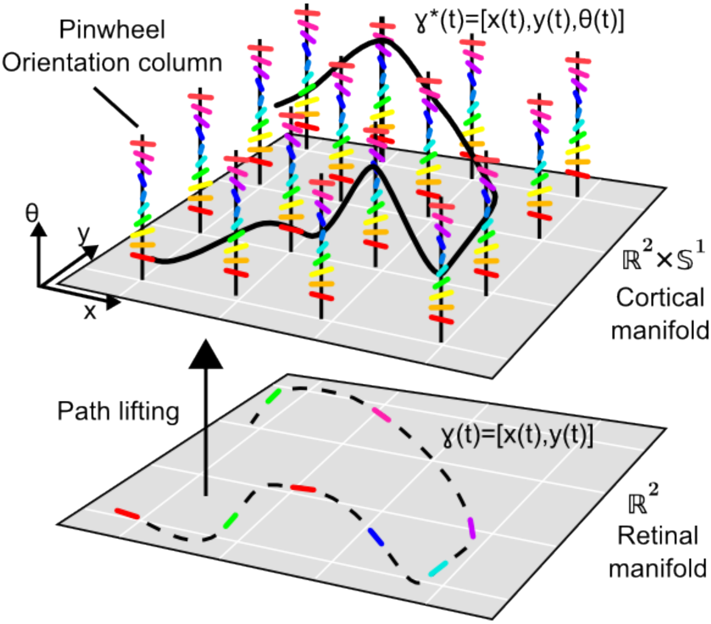

Figure 13.14 A fiber bundle abstraction of V1. Contour elements (doted line) on the retinal plan ($\mathbb{R}^2$) are lifted in the cortical fiber bundle ($\mathbb{R}^2 \times \mathbb{S}^1$) by making contact with simple cells (figure inspired from [40, 16]).

A fiber bundle $E = M \times F$ is a manifold M on which is attached, at every position, the entire copy of another manifold called the fiber F. The analogy with the visual cortex is direct — M is the retinal plane and the fibers F are the hypercolumns of simple cells. If simple cells are directional derivatives (equation 13.3), then all orientations $[0, 2\pi]$ are distinguishable, and V1 is abstracted as $\mathbb{R}^2 \times \mathbb{S}^1$ [15, 40, 16, 69] where $\mathbb{R}^2$ corresponds to the retinal plane, and where the unit circle $\mathbb{S}^1$ corresponds to the group of rotations in the plane. This fiber bundle, also called the unit tangent bundle, is isomorphic to the *Euclidean motion group SE*(2), also called the *Roto-translation group*. If the simple cells are second order derivatives (i.e., even phase Gabors), then only the angles $[0, \pi]$ are distinguishable, and V1 is represented by $\mathbb{R} \times \mathbb{P}^1$, where $\mathbb{P}^1$ is the space of all lines through the origin [70].

As illustrated in figure 13.14, when a visual contour, expressed as a regular parameterized curve $\gamma(t) = [x(t), y(t)]$, makes *contact* with the simple cells, it is lifted in the hypercolumns to a curve $\gamma^*(t) = [x(t), y(t), \theta(t)]$. In the new space $\mathbb{R}^2 \times \mathbb{S}^1$, the orientations θ are explicitly represented. Because of this, the problem of contour representation and contour completion in V1 is different than if it is expressed directly on the retinal image. Indeed, the type of curves in $\mathbb{R}^2 \times \mathbb{S}^1$ that can represent visual contours is very restricted. Every curve on the retinal plane corresponds to a lifted curve in $\mathbb{R}^2 \times \mathbb{S}^1$, but the converse is not true — only special curves in $\mathbb{R}^2 \times \mathbb{S}^1$ can be curves in $\mathbb{R}^2$. For instance, the reader can verify that a straight line in $\mathbb{R}^2 \times \mathbb{S}^1$ is not a curve in $\mathbb{R}^2$. The *Frobenius Integrability Theorem* says that for a curve in $\mathbb{R}^2 \times \mathbb{S}^1$ to be a curve in $\mathbb{R}^2$ we must have

$$tan(\theta) = \dot{y(t)}/\dot{x(t)} = dy/dx. \tag{13.9}$$

It follows that a curve in V1 is admissible to represent a visual contour only if condition 13.9 is satisfied. Admissible curves in $\mathbb{R}^2 \times \mathbb{S}^1$, illustrated in figure 13.15, can be defined as the family of integral curves of $\mathbf{X}_1 + k\mathbf{X}_2$, where

$$\mathbf{X}_1 = (\cos\theta, \sin\theta, 0)\ ,\ \mathbf{X}_2 = (0, 0, 1) \tag{13.10}$$

are vector fields generating a space in $\mathbb{R}^2 \times \mathbb{S}^1$ which is orthogonal to the directional derivative expressed by simple cells (equation 13.2), and where k gives the curvature of the curve projected on the retinal plane [15, 40].

In [15, 40], individual points at each hypercolumn are connected into global curves through the fan of integral curves, which together give a model of the lateral connectivity in V1, described in section 13.8. This pattern of connectivity over the columns of V1 is illustrated in figure 13.16.

In [16, 69], following earlier works [71, 72, 73], contour completion of the type shown in figure 13.13 is defined by selecting, among the admissible curves of equation 13.9, the shortest path (i.e., geodesic) in $\mathbb{R}^2 \times \mathbb{S}^1$ between two visible boundary points. This gives a top-down formalism for perceptual phenomena such as contour completion and saliency (pop-out). As suggested by Hoffman in [68, 12], contour completion and saliency is more the exception than the rule in the visual world. When the premises of the Frobenius Integrability Theorem are not satisfied (when orientations of simple cells do not line up), texture is perceived instead of contours. In other

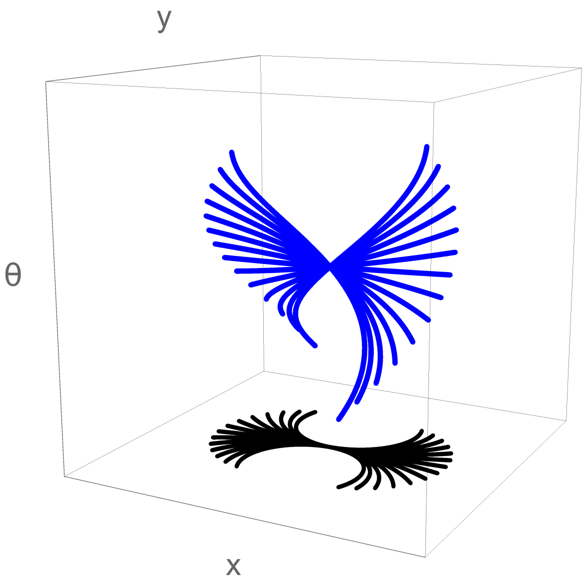

Figure 13.15 Family of integral curves in V1. The figure shows admissible V1 curves projected (in blue) on the retinal plane over one hypercolumn in $\mathbb{R}^2 \times \mathbb{S}^1$ (image adapted from [40]).

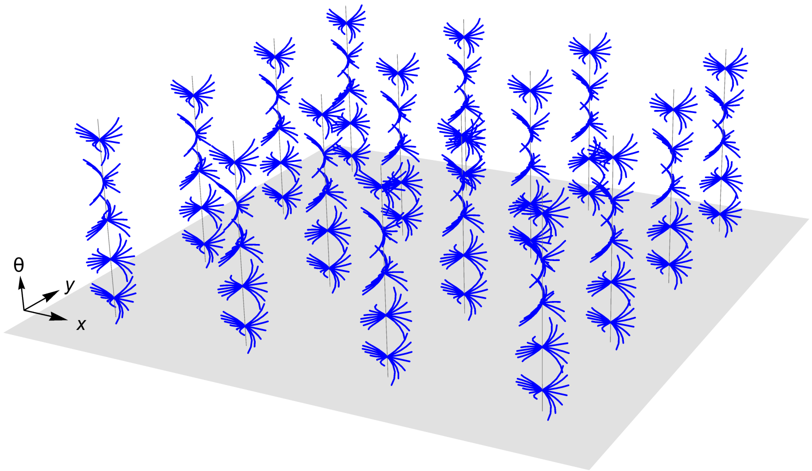

Figure 13.16 Horizontal connections. The fan of integral curves of the vector fields defined by equation 13.10 gives a connection between individual hypercolumns. It connects local tangent vectors at each hypercolumn into global curves (image adapted from [40]).

words, contour formation is not possible in V1 when the above integrability condition is not satisfied—in this frequent situation, it is not a contour that is perceived, but a texture. Using these principles, Hoffman defines visual contours of arbitrary complexity as the integral curves of an algebra of vector fields (i.e., Lie algebra) satisfying the above integrability conditions.

Of particular interest is the fact that the integral curves defined above are shown to correspond with long-range horizontal connections on the visual cortex. These connections are known as the *association field* [65] of perceptual contours and are described in the next section.

13.8 Horizontal connections and the Association field

Neuro-physiological studies have shown that the outputs of simple cells (i.e., the shape of the filters in figure 13.5) are modulated by contextual spatial surrounding — the global visual field modulates the local responses through long-range horizontal feedback in the primary visual cortex [74, 75]. In particular, in [65], psychophysical results suggest that long-range connections exist between simple cells oriented along particular paths, as illustrated in figure 13.17. These connections may define what is referred to as an *association field*—grouping (associating) visual elements according to alignment constraints on position and orientation. By selecting configurations of simple cells, the V4-type cells modeled in section 13.6 share some relations with the association field. However, these configurations are not explicitly defined by principles of grouping and alignment. As shown in [40, 14, 15], the association field can be put in direct correspondence with the integral curves (see figure 13.15) of the cortical fields expressed by simple cells and may well be the biological substrate for perceptual contours and phenomena (i.e., Gestalt) such as illusory contours, surface fill-in, and contour saliency.

13.9 Feedback and attentional mechanisms

Studies also show that attentional mechanisms, regulated through cortical feedback, modulate neural processing in regions as early as V1. Spatial pattern of modulation in [76] reveals an attentional process which is consistent with an object-based attention window but is inconsistent with a simple circumscribed attentional processes. This study suggests that neural processing in visual area V1 is not only driven by inputs from the retina but is often influenced by attentional process mediated by top-down connections. Although rapid visual recognition has sometimes been modeled by purely feedforward networks (section 13.6), studies and models [3] suggest that attentional mechanisms, mediated by top-down feedback, are essential for a full understanding of visual recognition mechanisms in the cortex. The question remains an open discussion, but it can be argued that there is indeed enough time for recurrent

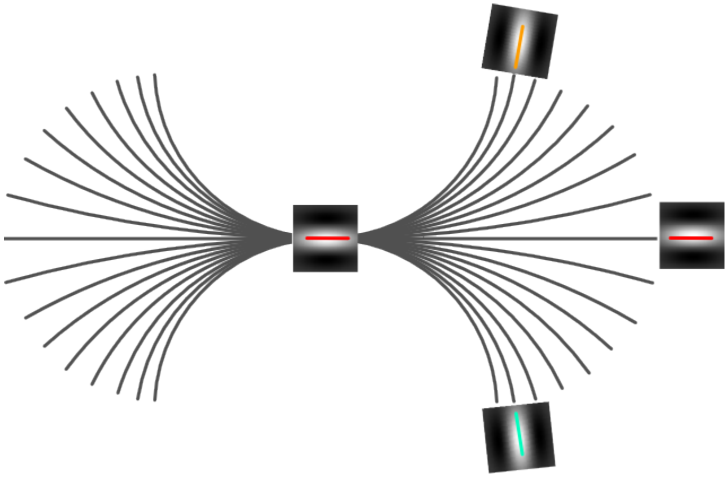

Figure 13.17 Association field. Long-range horizontal connections are the basis for the association field in [65]. The center of the figure represents a simple cell in V1. The curve displayed represent the visual path defined by the horizontal connections to other simple cells. These possible contours are shown to correspond to the family of integral curves as defined in section 13.7, where the center cell gives the initial condition.

activation (feedback) to occur during rapid visual recognition ($\sim$150ms) [29]. For a description of bottom-up attentional visual models see chapter **??**.

13.10 Temporal considerations, transformations and invariance

Object recognition may appear as a static process which involves the identification of visual inputs at one given instant. However, recognizing a visual object is everything but a static process and is fundamentally involved with transformations. The retinal image is never quite stable and visual perception vanishes in less than 80ms in the absence of motion [77]. Ecologically speaking, invariance to transformations generated by motion is essential for visual recognition to occur, and transformations of the retinal image are present at every instant since birth. Self-produced movements, such as head and eyes movements, are correlated with transformations of the retinal image and must have profound effect on the neural coding of the visual field [78].

The involvement of motion in the perception of shapes is displayed clearly by phenomena such as *structure from motion* [79], and the segmentation of shapes from motion [20]. The connectivity of the visual cortex also suggests a role for motion in object recognition. The ventral pathway, traditionally associated with object recognition, and the dorsal pathway, often associated with motion perception and spatial localization, are known to be significantly interacting [26, 27].

In neural network modeling of visual recognition, motion and the temporal dimen-

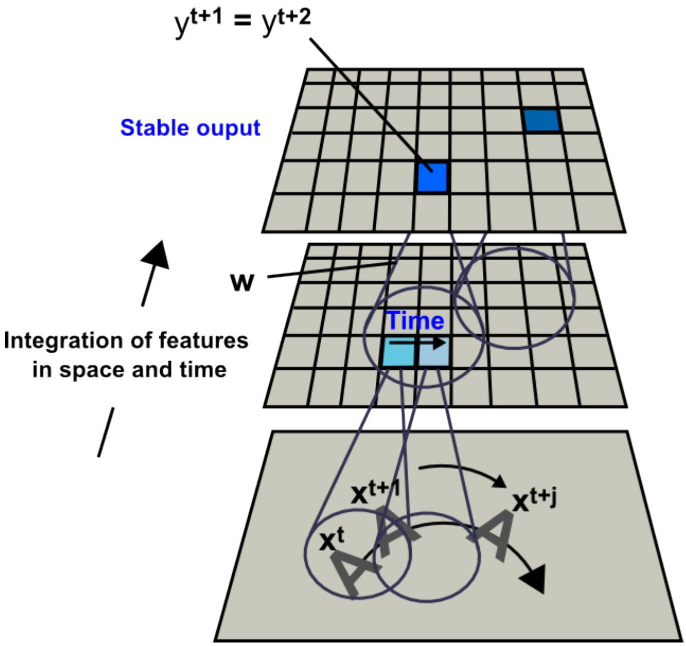

Figure 13.18 Temporal association. By keeping a temporal trace of recent activation, the trace learning rule enables neurons to correlate their activation with a transformation sequence passing through their receptive fields. This generates a neural response which is stable or invariant to transformation of objects inside their receptive fields.

sion of the neural response to an object undergoing transformations has been extensively modeled [80, 54, 23, 53, 81]. One unsupervised learning principle, common to all of these models, is to remove temporal variations in the neural response in the presence of transforming objects, as done by [80] with a *differential synapse*. Another unsupervised learning rule based on the same principle is known as the *trace rule* [55]. This is a *Hebbian* learning rule which keeps a temporal trace of the neuron's output in response to a moving object passing inside its receptive field (see figure 13.18). The inclusion of an activation trace in the learning rule drives neurons to respond invariantly to temporally close inputs (i.e., positions of a moving object). Inside an architecture with overlapping receptive fields, this allows a smooth mapping of transformations. This learning rule has also been used for the last layer of the HMAX model to gain translation invariance [81]. Interestingly, the trace learning rule can be mathematically related [82] to temporal difference learning (*TD*) [83], which minimizes temporal differences in response to successive inputs. For instance, one version of the trace learning rule modifies the connection weight vector $\mathbf{w}$ of output neuron $y^t = f(\mathbf{w}\mathbf{x}^t)$ at time t in response to input $\mathbf{x}^t$ (figure 13.18) such that

$$\Delta\mathbf{w} = \alpha(\beta\bar{y}^{t-1} - y^t)\mathbf{x}^t \tag{13.11}$$

where $\bar{y}^{t-1}$ is an exponential trace of the neuron's past activation.

A similar principle is also the basis of *Slow Feature Analysis* [53] in which neurons learn to output a signal at a lower frequency than the incoming retinal signal, producing a stable response to transforming objects. The Slow Feature Analysis has been integrated in the HMAX architecture in the context of dynamic scenes recognition [84, 85]. The authors applied the Slow Feature Analysis to the output of simple cells responding to image sequences in videos. This generates more stable representations of motion in comparison with the raw simple cell outputs. All these unsupervised learning rules, from the point of view of visual recognition, implicitly suggest that, for obvious ecological reasons, the identity of objects should not change because of relative motion between the observer and the objects, and that the brain uses spatio-temporal correlations in the signal to maintain a coherent representation.

13.11 Conclusion

The average number of synapses per neuron in the brain is close to 10000, with a total estimate of 86 billion neurons in the human brain [86]. In particular, the primary visual cortex averages 140 millions neurons [87]. This roughly amounts to 1.4 trillion synaptic connections in and out of V1 alone. Interestingly, there is a large connection ratio between V1 and the retina. In particular, with an estimate of one million retinal ganglion cells [88], there is a connectivity ratio well over one hundred V1 neurons for each retinal ganglion cell.

With such massive connectivity, it is clear that the representation, the sampling and the filtering (i.e., averaging) capacity of the visual cortex is far beyond neural network models which may be simulated on today's basic computers. The large number of V1 neurons allocated on each retinal fiber suggests a significant amount of *image processing*. For instance, the simple calculation of derivatives by the local operators defined in equation 13.2 is plagued with discretization noise when applied directly on pixelized images, unless very large images are used with a large Gaussian envelope defining the filters. This makes theoretical principles, such as differentiation, difficult to express in practice. It is tempting to suggest that visual cortex evolved with such an astronomical number of connections to deal with noise and irregularities of the visual world and to create a nearly smooth signal from the discretized retinal receptors. A massive number of horizontal connections between V1 cells, as presented in sections 13.7 and 13.8, seem to be part of nature's solution to clean, to segment, and to amplify coherent visual structures in an otherwise cluttered environment.

Models such as the HMAX illustrate the basic functioning of simple and complex cells in V1. Without extensive horizontal and feedback connections, this model does not represent the full capacity of the visual cortical pathway. Nevertheless, it seems quite certain by now that one axiom of processing in V1 is based on local operations such as the ones described in sections 13.4 and 13.6. A code for the model presented in section 13.6 is available at http://www-poleia.lip6.fr/ cord/Projects/Projects.html. The next step in such a model could be to integrate feedback and horizontal connections, as defined in a more theoretical context in section 13.7. It remains a challenge to implement such formal theoretical models on full scale natural scenes. Research would most likely progress in the right direction by generating effort in bridging the gap between what is observed in the brain and models of visual recognition.

References

1 DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012) How does the brain solve visual object recognition? *Neuron*, **73**, 415–34.

2 Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., and Poggio, T. (2005), A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex, AI Memo 2005-036/CBCL Memo 259, Massachusetts Institute of Technology.

3 Lee, T.S. and Mumford, D. (2003) Hierarchical Bayesian Inference in the Visual Cortex. *Journal of the Optical Society of America*, **20**, 1434–1448.

4 Borji, A. and Itti, L. (2013) State-of-the-Art in Visual Attention Modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35** (1), 185–207.

5 Itti, L. and Koch, C. (2001) Computational Modelling of Visual Attention. *Nature Reviews Neuroscience*, **2** (3), 194–203.

6 Siagian, C. and Itti, L. (2007) Rapid Biologically-Inspired Scene Classification Using Features Shared with Visual Attention. *IEEE Trans. Pattern Anal. Mach. Intell.*, **29** (2), 300–312.

7 Zhang, Y., Meyers, E.M., Bichot, N.P., Serre, T., Poggio, T.A., and Desimone, R. (2011) Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences USA*, **108** (21), 8850–8855.

8 Chikkerur, S., Serre, T., Tan, C., and Poggio, T. (2010) What and where: A Bayesian inference theory of attention. *Vision Research*, **50**, 2233–2247.

9 Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007) Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29**, 411–426.

10 Mutch.J and Lowe.D.G (2008) Object class recognition and localization using sparse features with limited receptive fields. *Int. J. Comput. Vision*, **80**, 45–57.

11 Theriault, C., Thome, N., and Cord, M. (2013) Extended Coding and Pooling in the HMAX Model. *IEEE Transactions on Image Processing*, **22** (2), 764–777.

12 Hoffman, W. (1989) The visual cortex is a contact bundle. *Applied Mathematics and Computation*, **32**, 137–167.

13 Koenderink, J. (1988) Operational Significance of Receptive Field Assemblies. *Biological Cybernetic*, **58**, 163–171.

14 Petitot, J. (2003) The neurogeometry of pinwheels as a sub-riemannian contact structure. *J.Physiology, Paris*, **97**, 265–309.

15 Citti, G. and Sarti, A. (2006) A Cortical Based Model of Perceptual Completion in the Roto-Translation Space. *Journal of Mathematical Imaging and Vision*, **24** (3), 307–326.

16 Ben-Yosef, G. and Ben-Shahar, O. (2012) A Tangent Bundle Theory for Visual Curve Completion. *IEEE Trans. Pattern Anal. Mach. Intell.*, **34** (7), 1263–1280.

17 Felleman, D.J. and Essen, D.C.V. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, **1** (1), 1–47.

18 Priebe, N.J., Lisberger, S.G., and Movshon, J.A. (2006) Tuning for Spatiotemporal Frequency and Speed in Directionally Selective Neurons of Macaque Striate Cortex. *The Journal of Neuroscience*, **26** (11), 2941–2950.

19 Lennie, P. and Movshon, J. (2005) Coding of color and form in the geniculostriate visual pathway (invited review). *J Opt Soc Am A Opt Image Sci Vis.*, **22**, 2013–2033.

20 Born, R.T. and Bradley, D.C. (2005) Structure and function of visual area MT. *Ann. Rev. of Neurosci.*, **28** (1), 157–189.

21 Christopher C, P., Richard T, B., and Margaret S, L. (2003) Two-Dimensional Substructure of Stereo and Motion Interactions in Macaque Visual Cortex. *Neuron*, **37**, 525–535.

22 C C, P. and R T, B. (2001) Temporal Dynamics of a Neural Solution to the Aperture Problem In Visual Area MT of Macaque Brain. *Nature*, p. 1040.

23 Rolls, E. and Deco, G. (2002) *Computaitonal neuroscience of vision*, Press:Oxford, 1st edn..

24 Cloutman, L.L. (2013) Interaction between dorsal and ventral processing streams: Where, when and how? *Brain and Language*, **127** (2), 251 – 263.

25 Zanon, M., Busan, P., Monti, F., Pizzolato, G., and Battaglini, P. (2010) Cortical Connections Between Dorsal and Ventral Visual Streams in Humans: Evidence by TMS/EEG Co-Registration. *Brain Topography*, **22** (4), 307–317.

26 Schenk, T. and McIntosh, R. (2010) Do we have independent visual streams for perception and action? *Cognitve Neuroscience*, pp. 52–62.

27 McIntosh, R. and Schenk, T. (2009) Two visual streams for perception and action: current trends. *Neuropsychologia*, pp. 1391–1396.

28 Thorpe, S., Fize, D., and Marlot, C. (1996) Speed of processing in the human visual system. *Nature*, **381**, 520.

29 Johnson, J.S. and Olshausen, B.A. (2003) Timecourse of neural signatures of object recognition. *J Vis*, **3** (7), 499–512.

30 Hubel, D. (1962) The visual cortex of the brain. *Scientific American.*, **209** (5), 54 – 62.

31 Adams, D. and Horton, J. (2003) A precise retinotopic map of primate striate cortex generated from the representation of angioscotomas. *J Neurosci.*, **23** (9), 3371 – 3389.

32 Roe, A.W., Chelazzi, L., Connor, C.E., Conway, B.R., Fujita, I., Gallant, J.L., Lu, H., and Vanduffel, W. (2012) Toward a Unified Theory of Visual Area V4. *Neuron*, **14** (1), 12–29.

33 David, S.V., Hayden, B.Y., and Gallant, J.L. (2006) Spectral Receptive Field Properties Explain Shape Selectivity in Area V4. *Journal of Neurophysiology*, **96** (6), 3492–3505.

34 Rolls, E.T., Aggelopoulos, N.C., and Zheng, F. (2003) The Receptive Fields of Inferior Temporal Cortex Neurons in Natural Scenes. *Journal of Neuroscience*, **23**, 339–348.

35 Hubel.D and Wiesel.T (1959) Receptive fields of single neurones in the cat's striate cortex. *Journal of physiology*, pp. 574–591.

36 De Valois, R.L. and De Valois, K.K. *Spatial vision*, Oxford psychology series, Oxford University Press, 1988, (1990 [printing]), New York, Oxford.

37 L. Maffei, B. G. Hertz, A.F. (1973) The visual cortex as a spatial frequency analyser. *Vision Research*, **13**, 1255–1267.

38 Campbell, F.W., Cooper, G.F., and Cugell, E.C. (1969) The Spatial Selectivity of the Visual Cells of the Cat. *Journal of Physiology (London)*, **203**, 223–235.

39 Bosking, W., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997) Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex. *Journal of Neuroscience*, **17**, 2112–2127.

40 Sanguinetti, G. (2011) *Invariant models of vision between phenomenology, image statistics and neurosciences.*, Ph.D. thesis, Universidad de la República, Facultad de Ingeniería.

41 Koenderink, J. and van Doorn, A. (1987) Representation of Local Geometry in the Visual System. *Biological Cybernetic*, **55**, 367–375.

42 Young, R.A. (1987) The Gaussian derivative model for spatial vision: I. Retinal mechanisms. *J. of Physiology*, **2**, 273–293.

43 Gabor, D. (1946) Theory of Communication. *J. Inst. Elect. Eng.*, **93**, 429–457.

44 Daugman, J.G. (1980) Two-dimensional Spectral Analysis of Cortical Receptive Field Profiles. *Vision Research*, **20** (10), 847–856.

45 Daugman, J.G. (1985) Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, **2** (7), 1160–1169.

46 Simoncelli, E. and Heeger, D. (1998) A model of neural responses in visual area MT. *Vision Research*, **38**, 743–761.

47 T.Lindeberg (1998) Feature detection with automatic scale selection. *Int. J. of Computer Vision*, **30**, 77–116.

48 Masquelier, T. and Thorpe, S.J. (2007) Unsupervised Learning of Visual Features through Spike Timing Dependent Plasticity. *PLoS Comput Biol*, **3** (2), e31.

49 Lampl, I., Ferster, D., Poggio, T., Riesenhuber, M., Ferster, D., and Poggio, T. (2004) Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *J. Neurophys*, **92**, 2704–2713.

50 Fukushima, K. and Miyake, S. (1982) Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition*, **15** (6), 455–469.

51 Fukushima, K. (2003) Neocognitron for handwritten digit recognition. *Neurocomputing*, **51**, 161–180.

52 Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998) Gradient-based learning applied to document recognition, in *Proceedings of the IEEE*, pp. 2278–2324.

53 Wiskott, L. and Sejnowski, T. (2002) Slow feature analysis: Unsupervised learning of invariances. *Neural Comput.*, **14**, 715–770.

54 Rolls, E.T. and Milward, T.T. (2000) A Model of Invariant Object Recognition in the Visual System: Learning Rules, Activation Functions, Lateral Inhibition, and Information-Based Performance Measures. *Neural Comput.*, **12**, 2547–2572.

55 Foldiak, P. (1991) Learning invariance from transformation sequences. *Neural Comput.*, **3** (2), 194–200.

56 Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P.A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *Journal of Machine Learning Research*, **11**, 3371–3408.

57 Goh, H., Thome, N., Cord, M., and Lim, J.H. (2012) Unsupervised and Supervised Visual Codes with Restricted Boltzmann Machines, in *European Conference on Computer Vision (ECCV 2012)*.

58 Fabio, A., Joel Z, L., Lorenzo, R., Jim, M., Andrea, T., and Poggio, T. (2014) Unsupervised learning of invariant representations with low sample complexity: the magic of sensory cortex or a new framework for machine learning?, *Technical Memo CBMM Memo No. 001*, Center for Brains, Mind and Machines, MIT, Cambridge, USA.

59 Riesenhuber.M and Poggio.T (1999) Hierarchical models of object recognition in cortex. *Nature Neuroscience*, **2**, 1019–1025.

60 Kohonen, T. (2001) *Self-organizing maps*, Springer series in information sciences, 30, Springer, Berlin, 3rd edn..

61 Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y. (2009) What is the best multi-stage architecture for object recognition?, in *ICCV*, IEEE, pp. 2146–2153.

62 Hu, X., Zhang, J., Li, J., and Zhang, B. (2014) Sparsity-Regularized HMAX for Visual Recognition. *PLoS ONE*, **9** (1), e81 813.

63 Pinto, N., Cox, D.D., and DiCarlo, J.J. (2008) Why is real-world visual object recognition hard? *PLoS Computational Biology*, **4** (1), e27.

64 Lyu, S. and Simoncelli, E.P. (2008) Nonlinear image representation using divisive normalization., in *CVPR*, IEEE Computer Society.

65 Field, D.J., Hayes, A., and Hess, R.F. (1993) Contour Integration by the Human Visual System: Evidence for a Local Association Field. *Vision Research*, **33**, 173–193.

66 Thériault, C. (2006) *Mémoire visuelle et géométrie différentielle : forces invariantes des champs de vecteurs élémentaires*, Ph.D. thesis, Université du Québec à Montréal.

67 Hoffman, W. (1966) The Lie Algebra of Visual Perception. *Mathematical Psychology*, **3**, 65–98.

68 Hoffman, W. (1970) Higher visual perception as prolongation of the basic Lie transformation group. *Mathematical Biosciences*, **6**, 437–471.

69 Ben-Yosef, G. and Ben-Shahar, O. (2012) Tangent Bundle Curve Completion with Locally Connected Parallel Networks. *Neural Computation*, **24** (12), 3277–3316.

70 Petitot, J. and Tondut, Y. (1999) Vers une neurogéométrie. fibrations corticales, structures de contact et contours subjectifs modaux. *Mathématiques et sciences humaines*, **145**, 5–101.

71 Ullman, S. (1976) Filling-in the gaps: The shape of subjective contours and a model for their generation. *Biological Cybernetics*, **25**, 1–6.

72 Horn, B.K.P. (1983) The curve of least energy. *ACM Trans. Math. Softw.*, **9** (4), 441–460.

73 Mumford, D. (1994) Elastica and computer vision, in *Algebraic Geometry and its Applications* (ed. C.L. Bajaj), Springer-Verlag, New York.

74 Gilbert, C.D., Das, A., and Westheimer, G. (1996) Spatial Integration and Cortical Dynamics. *Proceedings of the National Academy of Sciences USA*, **93**, 615.

75 Bosking, W., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997) Orientation Selectivity and the Arrangement of Horizontal Connections in Tree Shrew Striate Cortex. *J. of Neuroscience*, **17**, 2112–2127.

76 Somers, D.C., Dale, A.M., and Tootell, R.B.H. (1999) Functional MRI Reveals Spatially Specific Attentional Modulation In Human Primary Visual Cortex. *Proceedings of the National Academy of Sciences USA*, **96**, 1663.

77 Coppola, D. and Purves, D. (1996) The extraordinarily rapid disappearance of entopic images. *Proceedings of the National Academy of Sciences USA.*, **93**, 8001–8004.

78 Dodwell, P. (1983) The Lie transformation group of visual perception. *Perception & Psychophysics*, **34**, 1–16.

79 Grunewald, A., Bradley, D.C., and Andersen, R.A. (2002) Neural Correlates of Structure-from-Motion Perception in Macaque V1 and MT. *Journal of Neuroscience*, **22**, 6195–6207.

80 Mitchison, G. (1991) Removing Time Variation with the Anti-hebbian Differential Synapse. *Neural Comput.*, **3** (3), 312–320.

81 Isik, L., Leibo, J., and Poggio, T. (2012) Learning and disrupting invariance in visual recognition with a temporal association rule. *Frontiers in computational neuroscience.*, **6** (7).

82 Rolls, E.T. and Stringer, S.M. (2001) Invariant object recognition in the visual system with error correction and temporal difference learning. *Network*, **12**, 111–130.

83 Sutton, R.S. (1988) Learning to predict by the methods of temporal differences, in *Machine Learning*, Kluwer Academic Publishers, pp. 9–44.

84 Theriault, C., Thome, N., Cord, M., and Perez, P. (2014) Perceptual Principles for Video Classification With Slow Feature Analysis. *Journal of Selected Topics in Signal Processing*, **8** (3), 428–437.

85 Theriault, C., Thome, N., and Cord, M. (2013) Dynamic Scene Classification: Learning Motion Descriptors with Slow Features Analysis, in *IEEE CVPR*.

86 Azevedo, F.A.C., Carvalho, L.R.B., Grinberg, L.T., Farfel, J.M., Ferretti, R.E.L., Leite, R.E.P., Filho, W.J., Lent, R., and Herculano-Houzel, S. (2009) Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *The Journal of Comparative Neurology*, **513** (5), 532–541.

87 Leuba, G. and Kraftsik, R. (1994) Changes in volume, surface estimate, three-dimensional shape and total number of neurons of the human primary visual cortex from midgestation until old age. *Anatomy and Embryology*, **190** (4), 351–366.

88 Curcio, C.A. and Allen, K.A. (1990) Topography of ganglion cells in human retina. *J Comp Neurol*, **300** (1), 5–25.