Available online at www.sciencedirect.com
ScienceDirect
Journal of Physiology - Paris 100 (2006) 125–132
Journal of Physiology Paris
www.elsevier.com/locate/jphysparis

Learning receptive fields using predictive feedback

Janneke F.M. Jehee [a,*], Constantin Rothkopf [b], Jeffrey M. Beck [b], Dana H. Ballard [c]

[a] Center for Visual Science, Department of Computer Science, University of Rochester, 242 Meliora Hall, Rochester, NY 14627-0270, USA
[b] Center for Visual Science, Brain and Cognitive Sciences, University of Rochester, Meliora Hall, Rochester, NY 14627-0270, USA
[c] Center for Visual Science, Department of Computer Science, University of Rochester, Computer Science Building, Rochester, NY 14627-0226, USA

Received 6 June 2006; revised 8 September 2006; accepted 20 September 2006

Abstract

Previously, it was suggested that feedback connections from higher- to lower-level areas carry predictions of lower-level neural activities, whereas feedforward connections carry the residual error between the predictions and the actual lower-level activities [Rao, R.P.N., Ballard, D.H., 1999. Nature Neuroscience 2, 79–87.]. A computational model implementing the hypothesis learned simple cell receptive fields when exposed to natural images. Here, we use predictive feedback to explain tuning properties in medial superior temporal area (MST). We implement the hypothesis using a new, biologically plausible, algorithm based on matching pursuit, which retains all the features of the previous implementation, including its ability to efficiently encode input. When presented with natural images, the model developed receptive field properties as found in primary visual cortex. In addition, when exposed to visual motion input resulting from movements through space, the model learned receptive field properties resembling those in MST. These results corroborate the idea that predictive feedback is a general principle used by the visual system to efficiently encode natural input.
Published by Elsevier Ltd.

Keywords: Efficient coding; Predictive feedback; Visual system; Minimum description length; Matching pursuit; Computational model

1. Introduction

Neurons in primary visual cortex (V1) respond to bars of different orientation (Hubel and Wiesel, 1968; Jones and Palmer, 1987), neurons in extrastriate cortex respond to angles and contours (Pasupathy and Connor, 1999; Hedgé and Van Essen, 2000), and neurons in area MST respond to optic flow (Orban et al., 1992; Duffy, 1998). How do these selectivities come about? A longstanding approach to understanding selectivity has been to consider it in terms of efficient coding of natural images (e.g., Barlow, 1961; Atick, 1992; Olshausen and Field, 1996). Natural stimuli are typically very redundant, and a direct representation of these signals by the array of sensory cells would be very inefficient. It has therefore been argued that the role of early sensory processing is to remove redundancy in the input, resulting in a more sparse and statistically independent output.

Building on these early ideas, it was posited that early-level response properties might result from predictive feedback (Rao and Ballard, 1999). Specifically, given that the visual system is hierarchically organized and that its connections are almost always reciprocal (Felleman and Van Essen, 1991), Rao and Ballard proposed that higher-level areas predict lower-level input through feedback connections, while lower-level areas signal the difference between actual neural activity and the higher-level predictions. This removes redundancy in the input by removing the predictable. To test the hypothesis, Rao and Ballard trained a computational neural network model on image patches taken from natural scenes. After training, tuning properties of the model neurons resembled tuning properties found for neurons in area V1 and V2, corroborating the predictive feedback hypothesis.

* Corresponding author. Tel.: +1 585 275 2203; fax: +1 585 271 3043.
E-mail addresses: jjehee@cvs.rochester.edu (J.F.M. Jehee), crothkopf@bcs.rochester.edu (C. Rothkopf), jbeck@bcs.rochester.edu (J.M. Beck), dana@cs.rochester.edu (D.H. Ballard).

0928-4257/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.jphysparis.2006.09.011

From anatomical studies it is clear that a hierarchical feedforward–feedback design is characteristic of many sensory areas (Felleman and Van Essen, 1991). This suggests that predictive feedback might be a general mechanism by which neuronal tuning properties are formed. In this paper, we use the predictive coding framework to explain receptive field properties as found in MST. Neurons in area MST respond to planar, radial, and circular motion, which are components of optic flow—optic flow refers to perceived motion of the visual field resulting from an individual's own movements through space. The area receives connections primarily from middle temporal area (MT), where neurons code for magnitude and direction of motion in small regions of the visual field (Maunsell and Van Essen, 1983; Allbright, 1984). We will show that, using the hierarchical predictive coding model, neurons tuned to optic flow naturally emerge when presented with local motion fields.

The Rao and Ballard implementation exhibited sparse coding (i.e., encoding of input with a small set of active neurons, see e.g., Olshausen and Field, 1996), which was inspired by minimum description length theory (MDL) (Rissanen, 1978; see also Grunwald et al., 2005). MDL chooses as the best model for a given set of data the one that leads to the largest compression of the data. Although this theory has an attractive information–theoretic formulation, robust algorithms that realize its promise have exhibited delicate convergence behavior. Here, we show that a new algorithm based on matching pursuit (Mallat and Zhang, 1993) has fast convergence properties and good behavior with respect to the MDL metric. Moreover, matching pursuit has a straightforward hierarchical implementation, and, as an emergent property of the algorithm, results in a sparse neural code. We illustrate the algorithm by modeling the connections between Lateral Geniculate Nucleus (LGN) and primary visual cortex. We show that our new mathematical implementation of the predictive feedback hypothesis not only reproduces orientation tuning as found for simple cells in cortical area V1, but also captures tuning to optic flow as found in MST cells.

2. Model

2.1. General architecture

Higher-level units try to predict the responses of units in the next lower level via feedback connections. Lower-level units signal the difference between the higher-level predictions and the actual activity through feedforward connections. Difference signals are then used to correct higher-level predictions. Thus, each module consists of two kinds of cells: coding units ('predictive estimators') and difference-detecting units (Fig. 1). If a lower-level module has information for the receptive fields in a more abstract higher-level module, then its coding units connect to the difference-detecting units of that module. Higher-level units have larger receptive fields. Cortical hierarchies can be built by combining modules. Here, we consider only one hierarchical level for each of two circuits. One simulates the LGN–V1 feedforward–feedback circuit and the other simulates the MT–MST feedforward–feedback circuit. The model can easily be extended to more levels.

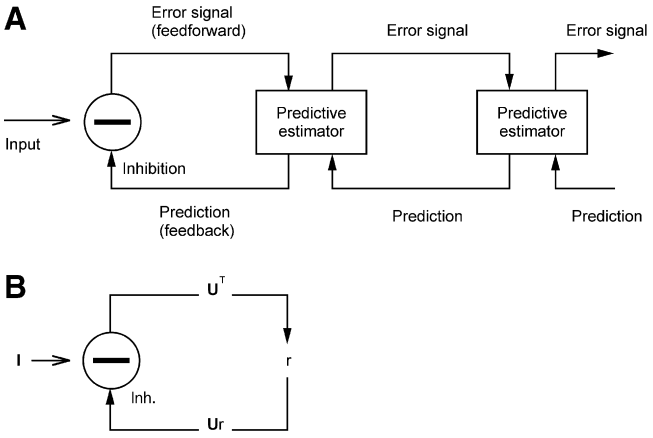

Fig. 1. Hierarchical model for predictive coding. (A) General architecture. Higher-level units attempt to predict the responses of units in the next lower level via feedback connections. Lower-level units signal the difference between the higher-level predictions and the actual activity through feedforward connections. Difference signals are then used to correct higher-level predictions. (B) Components of a module. Feedforward connections encode the synaptic weights $\mathbf{U}^{\mathrm{T}}$, coding units maintain the current estimate of the input signal and convey the top-down prediction $\mathbf{U}r$ to the lower level via feedback connections. Difference units compute the difference $(\mathbf{I} - \mathbf{U}r)$ between current activity $\mathbf{I}$ and its top-down prediction $\mathbf{U}r$.

2.1.1. LGN and V1

Scenes are 768 by 768 black-and-white images of natural surroundings (Fig. 2), preprocessed by filtering with a zero-phase whitening/lowpass filter (Atick, 1992; Olshausen and Field, 1996)

$$R(f) = f\mathrm{e}^{-(f/f_0)^4}$$

and subtracting the mean, where f stands for spatial frequency and $f_0 = 300$ cycles/image. The pixel values obtained in this way are taken as the initial activation values of neurons in the first layer, which would correspond to the LGN. The second layer, which would correspond to a small part of cortical area V1, is represented by 128 units. We limit the LGN input into model V1 to 8 by 8 pixels (or 64 LGN cells). Such 8-by-8 image 'patches' are randomly selected from the filtered input image, represented as a single vector, and fed into the V1 model neurons using feedforward connections. In the language of matching pursuit, we say that these feedforward connection weights from LGN to each of the V1 neurons constitute a basis vector. Basis vectors are initialized with random values and zero mean and then constrained to have unit length.

2.1.2. MT and MST

Image sequences consist of ten frames each and have a resolution of 480 by 480 pixels. They come from two differ-

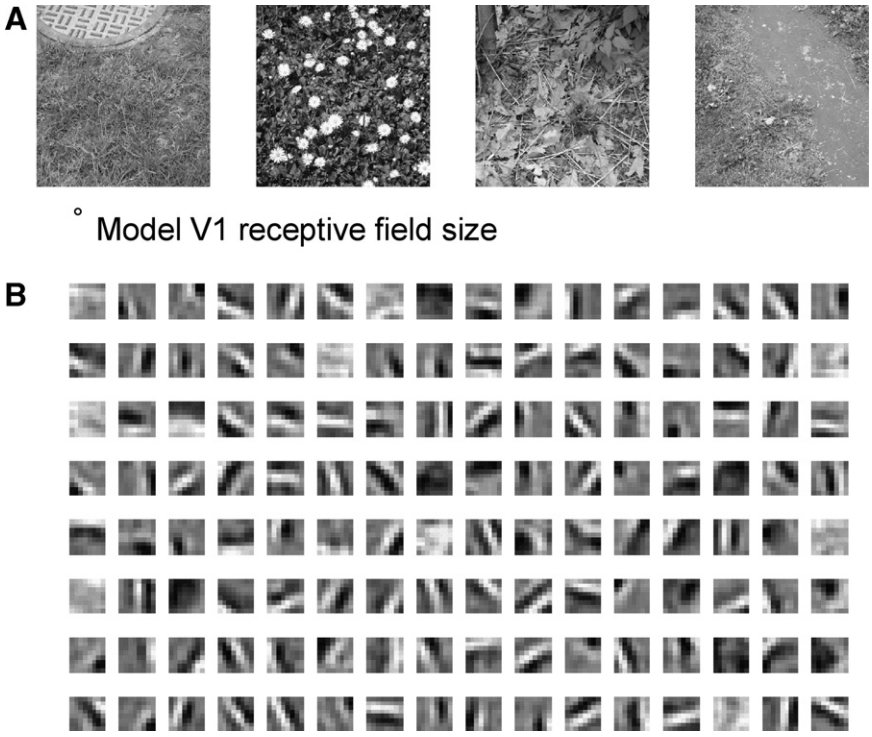

Fig. 2. Receptive fields of model V1 units after training on natural images. (A) Subset of natural images used for training. The circle denotes model V1 receptive field size. (B) Learned V1 receptive fields. Plots are scaled in magnitude so that each fills the gray scale, but with zero always represented by the same gray level. Black is negative, white is positive.

ent sources: half of the sequences are synthetic images obtained from models constructed in OpenSceneGraph (toolkit available online: http://www.openscenegraph.org/) and moving the virtual camera in the scenes according to translations and rotations. Combinations of translations or rotations along two axes are used in each sequence. The other half of the image sequences are obtained from real-world image recordings of human subjects walking in wooded areas and a cityscape. A digital video camera with an equivalent focal length of 40 mm was mounted on a helmet subjects wore while locomoting (Pelz and Rothkopf, in press). From these image sequences dense optical motion fields (represented as vectors pointing in the direction of motion) are computed using a multiscale implementation of the Lucas–Kanade tracker (Lucas and Kanade, 1981). The resulting motion fields are averaged over regions of size 30 by 30 leading to a motion field representation of 256 motion vectors arranged in a 16 by 16 array. Motion vectors are centered and whitened by subtracting their means and the singular value decomposition of their covariance, respectively (see Jabri et al., 2000; for similar preprocessing steps). The obtained values are used as initial activation values of the 256 units in the first layer, which would correspond to cortical area MT. The second layer, which would correspond to a small part of area MST, is represented with 128 neurons. The feedforward connectivity structure ('basis vector') into each MST unit is 16 by 16 in size, and initialized with random values and zero mean. At each new matching pursuit cycle (see below), one of the motion fields is randomly selected, represented as a single vector of unit length, and fed into the algorithm.

2.2. Network dynamics and learning

As previously indicated, the synaptic weights which drive a (let us say) V1 neuron form a basis vector which represents the preferred stimulus of that neuron. The action of many such V1 neurons is to predict the (say) LGN input, $\mathbf{I}$, as a linear combination of N basis vectors, where the weighting coefficient of each basis vector $\mathbf{u}_i$ is given by the response r_i of its corresponding V1 neuron. In this way, the original input can be predicted using the activity in V1 by simply applying the linear transformation

$$\mathbf{I} \approx \sum_{i=1}^{N} r_i \mathbf{u}_i$$

The typical matching-pursuit algorithm seeks to provide an accurate prediction of the input using the least number of basis vectors or equivalently the least number of active V1 neurons (Mallat and Zhang, 1993; see also Appendix). This is accomplished iteratively by first determining which of the 128 basis vectors best predicts the input, or equivalently, by finding the V1 neuron whose receptive field structure best matches the LGN input. This is done by determining which basis vector maximizes the inner product with the LGN input. The best-matching basis vector weighted by the neuronal response r (given by the maximal

inner product) is then subtracted from the input, and the process is repeated on the residual input. After k iterations the residual input is given by

$$\Delta \mathbf{I}_k = \Delta \mathbf{I}_{k-1} - r_{i_k}\mathbf{u}_{i_k} = \mathbf{I} - \sum_{l=1}^{k} r_{i_l}\mathbf{u}_{i_l}$$

and the next neuron is chosen by again determining which of the remaining V1 basis vectors best predicts this residual input. In a neural network, the subtractive process is carried out using feedback connections, so that at each iteration of the algorithm the residual input is described by the activity of the LGN and the projection of the residual input onto the remaining basis vectors is implemented using feedforward connections from the LGN to V1.

To better capture the input statistics in the future and enhance the sparseness of the neural code, basis vectors are incrementally updated in each feedforward–feedback cycle. This is done by minimizing the description length of the joint distribution of inputs and neural responses (Appendix). In earlier work (Rao and Ballard, 1999), sparseness was achieved by the *a priori* assumption that the distribution of neural responses is of a particular form that does not necessarily correspond to the correct neural response distribution. Here, we do not specify a (possibly incorrect) sparse prior distribution, but rather create a sparse code via the action of the matching pursuit algorithm, which generates a prediction and then learns receptive fields/basis vectors to better make predictions in the future. This is a slightly different notion of both sparsity and optimality than previously used as it is based on an analysis of the algorithm rather than on an assumption that the inputs were generated by a specific probabilistic model (Appendix). However, this new learning algorithm causes the model to converge quickly to a set of basis vectors that optimally capture the input statistics and allow for making predictions thereof. In the appendix, we show that the optimal learning rule is well approximated by a local learning rule, which takes the form of the traditional Hebbian rule:

$$\Delta \mathbf{u}_{i_k} = \gamma \langle r_{i_k} \Delta \mathbf{I}_{k-1} \rangle$$

where γ is given by $0.3/(1+\beta)$, and β is initially equal to 1 and increases by 1 every 1000 image patches. Parameter values are kept constant throughout all simulations. The model is allowed to process each image patch using four feedforward–feedback cycles. Basis vectors are updated and then normalized each time a neuron is chosen in one of the feedforward–feedback cycles. The V1 basis vectors are trained on 10 000 image patches extracted from 16 natural images. MST basis vectors are trained on motion fields extracted from 640 natural image sequences.

3. Results

To test the predictive coding hypothesis, we trained our matching pursuit model on image patches extracted from natural scenes, the motivation being that receptive field properties might be largely determined by the statistics of their natural input (see also, Field, 1987; Atick, 1992; Dan et al., 1996; Rao and Ballard, 1999). After exposure to several thousand natural image patches, the basis vectors learned by the model show orientation tuning as found for simple cells in V1 (Fig. 2). The basis vectors determine the feedforward responses, and can be considered as classical receptive fields of the higher-level units.

To explicitly test the matching-pursuit model on sparseness, we calculated the number of feedforward–feedback loops needed to accurately predict the input by computing the amount of overlap between a presented image patch and the linear combination of basis vectors chosen by the algorithm in each successive feedforward–feedback loop, weighted by their responses. Fig. 3 shows the amount of overlap before training (dashed line) and after training (solid line). After training, the prediction converges with the visual input after only few feedforward–feedback iterations, which corresponds to choosing the same number of basis vectors in model area V1. Since there are on the order of 128 V1 units, the code in this case is extremely sparse. Similar results are obtained using different image patches.

To test the generality of the predictive feedback approach, we also trained the model on visual motion input resulting from movements through space, which would resemble area MST's natural MT input. After exposure to visual motion input, basis vectors in the model exhibit tuning to translation and expansion (Fig. 4), which are components of optic flow. Thus, the model presented here not only captures V1 receptive field properties, but also MST receptive field properties in terms of cortico-cortical feedback used by the visual system to efficiently encode natural input.

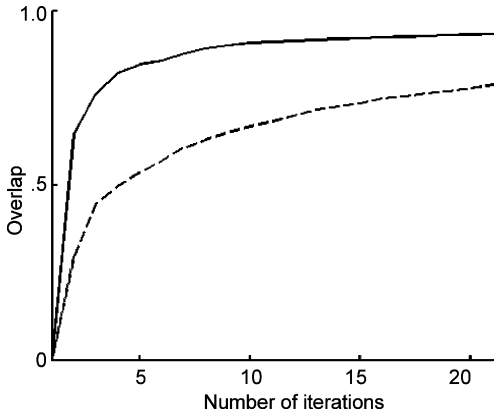

Fig. 3. Convergence of the matching pursuit algorithm for one representative image patch after training. Amount of overlap was computed by taking the dot product between a presented image patch and the linear combination of basis vectors chosen by the matching pursuit algorithm, weighted by their responses. Fewer steps are needed after training (solid line) than before training (dashed line). The small number of steps is a consequence of our algorithm and corresponds to a correspondingly small number of V1 neurons used to represent the stimulus.

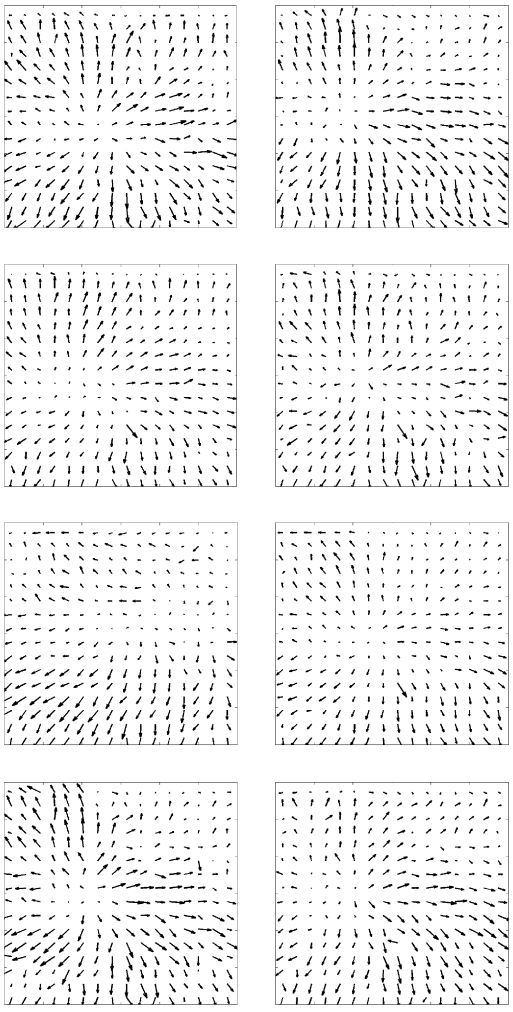

Fig. 4. Feedforward receptive fields of model MST after training (representative subset). Basis vectors in the model exhibit tuning to translation and expansion, which are components of optic flow.

4. Discussion

We have shown that efficiently encoding an image using predictive feedforward–feedback cycles captures neuronal receptive field properties as found in primary visual cortex, as well as receptive field properties of neurons in area MST. Moreover, the model exhibits sparse coding, which is the encoding of any particular stimulus with a small set of neurons. These results corroborate the idea that predictive feedback is a general principle used by the visual system to efficiently encode natural input. Although we have only demonstrated the utility of the algorithm for encoding one hierarchical level of the cortex, the extension to multiple levels is straightforward. Each module contains two kinds of neurons: coding neurons and difference-detecting neurons. If coding neurons have an output that can be combined with other outputs to be further coded, they can be combined at a new module's difference-detecting neurons. If a more abstract module wants to effect the code of a lower-level module, its error output is fed to the input of the lower-level coding neurons.

Previously, the predictive-feedback framework has been used to explain endstopping and other extra-classical receptive field effects in V1 (Hubel and Wiesel, 1968; Bolz and Gilbert, 1986), which were interpreted as responses from difference detectors that signal the error between a higher-level prediction and actual activity (Rao and Ballard, 1999). That is, areas along the visual hierarchy represent the image with increasing receptive field size, and higher-level predictions can be made on increasing spatial context. As a consequence, when a center stimulus matches the surround, less activity is elicited in lower-level difference-detecting neurons because the center can be predicted from the surround through feedback from higher-level areas. Such a prediction fails, however, when the stimulus is presented in isolation, generating a much larger response in the error-detecting neuron (Rao and Ballard, 1999). Suppressed responses are not phenomena restricted to area V1: similar extra-classical receptive field effects have been found in other visual areas (Hubel and Wiesel, 1968; Allman et al., 1985; Bolz and Gilbert, 1986; Desimone and Schein, 1987; Hubel and Livingstone, 1987). For example, some neurons in MT are suppressed when the direction of motion in the surround matches that in the classical receptive field (Allman et al., 1985), and it was conjectured that these might also be accounted for by predictive feedback (Rao and Ballard, 1999). We consider these applications as a topic of future research, and predict that our matching pursuit model will capture extra-classical receptive field effects in MT as well.

Other modeling studies have suggested mechanisms for learning simple cell receptive fields based on competition between input from LGN on-cells and off-cells (Soodak, 1987, 1991; Tanaka, 1992; Miller, 1994). However, extending such a mechanism in a straightforward way to other areas is difficult. In this paper, we presented a general mechanism for learning receptive fields, and showed that V1 and MST receptive field properties naturally emerge when using an efficient hierarchical and predictive strategy for encoding natural input.

Olshausen and Field (1996) posited that neurons in cortical area V1 could be understood as a sparse code for the visual stimulus. The idea was that area V1 was trying to code the stimulus with as few active neurons as possible. Here, we showed that sparse coding can be thought of as a case of a general coding principal termed minimum description length (Appendix, see also Rao and Ballard, 1999), implemented in the model's feedforward–feedback loop through the matching pursuit algorithm (Mallat and Zhang, 1993). Coding an image using such serial feedforward–feedback computations costs in time, but as the number of separate projections needed is typically very small, the total delay is modest. Furthermore, there is a shortcut when the circuit is part of a hierarchy. The largest projection is computed first and sent on to higher areas so that they do not have to wait for the entire computation to settle to get started on their own projection computation. Subsequent iterations refine the predictions from previous

cycles in both lower- and higher-level areas. Consistent with this idea, neurophysiological studies have found dynamic changes in the tuning properties of neurons in lower- (Lamme, 1995; Ringach et al., 1997; Hupé et al., 1998) and higher-level visual areas (Sugase et al., 1999) that have been shown to result from feedforward–feedback interactions (Hupé et al., 1998; Lamme et al., 1998).

Recent theories propose that feedforward processing involves rapid and automatic processes that enable basic object categorizations, however incorporating a limited amount of spatial detail (Hochstein and Ahissar, 2002; Lee et al., 1998; Roelfsema et al., 2000). For a detailed and complete representation, higher areas would need to reach back to the lower-level areas via feedback connections. Lower-level areas contain neurons with smaller receptive fields than neurons in higher areas, and are in that respect better suited for signaling of spatial detail. Our model is compatible with these theories in the sense that feedback mechanisms are used to highlight information that was not captured in the first feedforward sweep, but we put less emphasis on feedback interactions being necessary for processing of spatial detail.

In conclusion, edge selectivity and tuning to optic flow may result from feedforward–feedback interactions used by the visual system to efficiently encode natural images. Given the hierarchical layout of sensory areas (Felleman and Van Essen, 1991), it is likely that predictive feedback is a general mechanism used by the sensory system to shape receptive fields, of which the ones presented here are but examples.

Acknowledgement

We thank Wei Ji Ma for helpful comments on an earlier draft of this article.

Appendix

Maximum-likelihood formulation of sparseness

We seek a model that not only accurately predicts its inputs but also minimizes the number of neurons needed to predict any given input; in other words, we seek a code that gives a sparse representation of the input. In the standard approach (e.g., Rao and Ballard, 1999), sparseness is enforced via an assumption on the shape of the prior distribution on the neural responses. That is, it is supposed that the 8-by-8 input patch is predicted by a noisy linear combination of some basis vectors, i.e.,

$$\mathbf{I} = \mathbf{U}\mathbf{r} + \boldsymbol{\eta}_{\mathbf{I}}$$

where the basis vectors (or receptive fields) are the columns of the matrix $\mathbf{U}$, $\mathbf{r}$ is a vector with neuronal responses and $\boldsymbol{\eta}_{\mathbf{I}}$ is some additive Gaussian noise which is assumed to be independent with variance σ^2. This constitutes a likelihood function of the form $\Pr(\mathbf{I}|\mathbf{r}) = \text{Normal}(\mathbf{U}\mathbf{r}, \sigma^2)$. With the addition of a sparse prior $\Pr(\mathbf{r})$ on the responses one then determines ('learns') the maximum-likelihood values for the basis vectors which make up the matrix of basis vectors $\mathbf{U}$ by maximizing the expected log likelihood, L, of the joint distribution of $\mathbf{I}$ and $\mathbf{r}$, which is given by

$$\begin{aligned} L &= \langle \log \Pr(\mathbf{I}|\mathbf{r}) + \log \Pr(\mathbf{r}) \rangle \\ &= \left\langle -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|\mathbf{I} - \mathbf{U}\mathbf{r}\|^2 + \log \Pr(\mathbf{r}) \right\rangle \end{aligned}$$

for any (input) data set composed of vectors $\mathbf{I}$ of length N. Here $\|\cdot\|$ indicates the L2 norm. For a given input, one builds the vector of neuronal responses $\mathbf{r}$ by determining which neural responses maximize the likelihood of a given input being observed; in other words, one determines the most likely responses given this particular input.

Problems with the common formulation of minimum description length

In MDL, the goal is to minimize the total number of bits needed to encode both the higher-level neural responses (the r_i's) and the residual error of the predictions (the errors represented by the lower-level area) (Rissanen, 1978; see also Grunwald et al., 2005). There is a link between maximum-likelihood modeling as shown above and the minimum description length approach. We define the Kullback–Leiber (KL) divergence between the empirically observed joint distribution of $\mathbf{I}$ and $\mathbf{r}$ and the distribution given by the generative model L (which is specified to be sparse):

$$D_{\mathrm{KL}}(p(\mathbf{I},\mathbf{r}), q(\mathbf{I},\mathbf{r})) = -H(\mathbf{I},\mathbf{r}) - \langle \log q(\mathbf{I},\mathbf{r}) \rangle_{p(\mathbf{I},\mathbf{r})}$$

where $H(\mathbf{I},\mathbf{r})$ gives the entropy or description length of the true joint distribution $p(\mathbf{I},\mathbf{r})$, the $\langle\,\rangle$ indicate a expected value with respect to the true distribution $p(\mathbf{I},\mathbf{r})$, and $q(\mathbf{I},\mathbf{r})$ is the distribution of a proposed generative model (i.e., L above). If the true model corresponds to the generative model, the KL divergence would be zero. In that case, maximizing the expected log likelihood L would be equivalent to minimizing the description length. However, since the KL divergence is positive definite and it is unlikely that the assumed form of the generative model is correct, we can conclude that the entropy $H(\mathbf{I},\mathbf{r})$ or description length of the true distribution p is bounded above by the negative of the expected log likelihood, i.e.,

$$H(\mathbf{I},\mathbf{r}) \leqslant \langle -\log q(\mathbf{I},\mathbf{r}) \rangle_{p(\mathbf{I},\mathbf{r})} = -L$$

This result implies that maximizing the expected log likelihood of a generative model will yield parameter values which minimize an upper bound on the description length, regardless of whether or not the chosen prior was indeed a sparse prior. Thus, though often linked, the information–theoretic notion of the minimum description length principle is only indirectly related to the notion of sparseness of neural activity, as the description length is minimized by a sparse generative model only when the generative distribution corresponds to the true distribution. Choosing an

incorrect prior on the neural responses (regardless of whether it is too sparse or not sparse enough) has the effect of leading to an overall increase in the description length. This result also implies that a sparse prior, like most priors used in a generative model framework, is an additional assumption which is imposed upon the data regardless of whether or not it is appropriate.

Optimizing matching pursuit: a new algorithm for building a sparse representation

Here, we suggest a more algorithmic approach to building a sparse neural representation which is based upon matching pursuit (Mallat and Zhang, 1993). This algorithm naturally generates a sparse representation and may also be formulated probabilistically allowing us to utilize the description length (as opposed to an upper bound) as the objective function for learning the basis vectors. That is, rather than specifying an approximate (reads incorrect) generative model $\Pr(\mathbf{I}|\mathbf{r})$ and prior $\Pr(\mathbf{r})$, we suggest analyzing the properties of the algorithm which will be used to generate the neural responses, $\Pr(\mathbf{r}|\mathbf{I})$, directly. This approach has the advantage of making no incorrect assumptions regarding the form of $\Pr(\mathbf{I})$ and places only very natural constraints on the shape of the prior $\Pr(\mathbf{r})$. Moreover, rather than minimizing an upper bound on the description length this formulation allows for the description length itself to be minimized directly. Specifically, note that the description length or joint entropy $H(\mathbf{I},\mathbf{r})$ can be written in terms of the conditional entropy of the neural responses:

$$H(\mathbf{I},\mathbf{r}) = H(\mathbf{r}|\mathbf{I}) + H(\mathbf{I})$$

Since the entropy of the inputs, $H(\mathbf{I})$, is independent of any parameters which are used to give the neural responses, we can conclude that minimizing the description length is equivalent to minimizing the conditional entropy of the neural responses, $H(\mathbf{r}|\mathbf{I})$, which means maximizing the expected log likelihood of the neural responses, $\langle \log \Pr(\mathbf{r}|\mathbf{I})\rangle$.

With this in mind, we now consider a biologically plausible, probabilistic implementation of matching pursuit and the associated learning rule. In this context, biologically plausible is assumed to imply two additional constraints: (1) responses in the model are constrained to be positive, and (2) the learning rule must be local much like the Hebbian learning rule shown above (Section 2). Since typical matching pursuit generates its neural responses iteratively we will also choose the kth neuron by sampling from the distribution

$$\Pr(i_k = j|\Delta\mathbf{I}_{k-1}, \{i_1 \ldots i_{k-1}\}) = \frac{\Theta(\mathbf{u}_j \cdot \Delta\mathbf{I}_{k-1}) \exp(\alpha \mathbf{u}_j \cdot \Delta\mathbf{I}_{k-1})}{\sum_j \Theta(\mathbf{u}_j \cdot \Delta\mathbf{I}_{k-1}) \exp(\alpha \mathbf{u}_j \cdot \Delta\mathbf{I}_{k-1})}$$

where i_k is the index of the kth neuron, $\mathbf{u}_j$ is the basis vector associated with the jth neuron, $\Theta(x)$ is the Heaviside function and $\alpha^{-1} = 1/10$ is a temperature parameter. Once a neuron has been selected, the actual response of that neuron is drawn from a normal distribution with mean given by $\mathbf{u}_{i_k} \cdot \Delta\mathbf{I}_{k-1}$ and small variance σ^2. Evaluation of the gradient of the conditional entropy generated from this distribution leads to terms that are either non-local or approximately proportional to $r_{i_k}\Delta\mathbf{I}_{k-1}$. From this, we conclude that the optimal local learning rule takes the form of the traditional Hebbian rule:

$$\Delta\mathbf{u}_{i_k} = \gamma\langle r_{i_k}\Delta\mathbf{I}_{k-1}\rangle$$

where γ is a learning rate (see also Section 2).

Note that this learning rule may also be obtained from the gradient of the error function for the kth iteration, i.e.,

$$\frac{\partial E_k}{\partial \mathbf{u}_{i_k}} = \frac{\partial}{\partial \mathbf{u}_{i_k}} \|\Delta\mathbf{I}_{k-1} - r_{i_k}\mathbf{u}_{i_k}\|^2 = 2r_{i_k}\Delta\mathbf{I}_{k-1}$$

References

Allbright, T.D., 1984. Direction and orientation selectivity of neurons in visual area MT of the macaque. J. Neurophys. 52, 1106–1130.

Allman, J., Miezin, F., McGuinness, E., 1985. Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local–global comparisons in visual neurons. Annu. Rev. Neurosci. 8, 407–429.

Atick, J.J., 1992. Could information theory provide an ecological theory of sensory processing? Network 3, 213–251.

Barlow, H.B., 1961. Possible principles underlying the transformation of sensory messages. In: Rosenblith, W.A. (Ed.), Sensory Communication. MIT Press, Cambridge, MA, pp. 217–234.

Bolz, J., Gilbert, C.D., 1986. Generation of end-inhibition in the visual cortex via interlaminar connections. Nature 320, 362–365.

Dan, Y., Atick, J.J., Reid, R.C., 1996. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. J. Neurosci. 16, 3351–3362.

Desimone, R., Schein, S.J., 1987. Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. J. Neurophysiol. 57, 835–868.

Duffy, C.J., 1998. MST neurons respond to optic flow and translational movement. J. Neurophysiol. 80, 1816–1827.

Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1, 1–47.

Field, D.J., 1987. Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A 4, 2379–2394.

Grunwald, P., Pitt, M.A., Myung, I.J. (Eds.), 2005. Advances in Minimum Description Length: Theory and Applications. MIT Press, Cambridge, MA.

Hedgé, J., Van Essen, D.C., 2000. Selectivity for complex shapes in primate visual area V2. J. Neurosci. 20, RC61.

Hochstein, S., Ahissar, M., 2002. View from the top: hierarchies and reverse hierarchies in the visual system. Neuron 36, 791–804.

Hubel, D.H., Livingstone, M.S., 1987. Segregation of form, color, and stereopsis in primate area 18. J. Neurosci. 7, 3378–3415.

Hubel, D.H., Wiesel, T.N., 1968. Receptive fields and functional architecture of monkey striate cortex. J. Physiol. (London) 195, 215–243.

Hupé, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P., Bullier, J., 1998. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature 394, 784–787.

Jabri, M.A., Park, K.-Y., Lee, S.-Y., Sejnowksi, T.J., 2000. Properties of independent components of self-motion optical flow. Paper Presented at the 30th IEEE International Symposium on Multiple-Valued Logic, Portland, OR.

Jones, J.P., Palmer, L.A., 1987. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J. Neurophysiol. 58, 1233–1258.

Lamme, V.A.F., 1995. The neurophysiology of figure-ground segregation in primary visual cortex. J. Neurosci. 15, 1605–1615.

Lamme, V.A.F., Supèr, H., Spekreijse, H., 1998. Feedforward, horizontal, and feedback processing in the visual cortex. Curr. Opin. Neurobiol. 8, 529–535.

Lee, T.S., Mumford, D., Romero, R., Lamme, V.A.F., 1998. The role of the primary visual cortex in higher level vision. Vision Res. 38, 2429–2545.

Lucas, B., Kanade, T., 1981. An iterative image registration technique with an application in stereo-vision. In: Proceedings of the 7th IJCAI, pp. 674–679.

Mallat, S., Zhang, Z., 1993. Matching pursuit with time-frequency dictionaries. IEEE T. Signal Process. 41, 3397–3415.

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

Miller, K.D., 1994. A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between on- and off-center inputs. J. Neurosci. 14, 409–441.

Olshausen, B.A., Field, D.J., 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381, 607–609.

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

Pasupathy, A., Connor, C.E., 1999. Responses to contour features in macaque area V4. J. Neurophysiol. 82, 2490–2502.

Pelz, J.B., Rothkopf, C., in press. Oculomotor behavior while navigating natural and man-made environments. In: van Gompel, R.P.G., Fischer, M.H., Murray, W.S., Hill, R.L. (Eds.), Eye Movements: A Window on Mind and Brain, Elsevier, Oxford.

Rao, R.P.N., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive field effects. Nature Neurosci. 1, 79–87.

Ringach, D.L., Hawken, M.J., Shapley, R., 1997. Dynamics of orientation tuning in macaque primary visual cortex. Nature 387, 281–284.

Rissanen, J., 1978. Modeling by the shortest data description. Automatica 14, 465–471.

Roelfsema, P.R., Lamme, V.A.F., Spekreijse, H., 2000. The implementation of visual routines. Vision Res. 40, 1385–1411.

Soodak, R.E., 1987. The retinal ganglion cell mosaic defines orientation columns in sriate cortex. Proc. Natl. Acad. Sci. USA 84, 3936–3940.

Soodak, R.E., 1991. Reverse-Hebb plasticity leads to optimization and association in a simulated visual cortex. Visual Neurosci. 6, 507–518.

Sugase, Y., Yamane, S., Ueno, S., Kawano, K., 1999. Global and fine information coded by single neurons in the temporal visual cortex. Nature 400, 869–873.

Tanaka, S., 1992. A mathematical model for the self-organization of orientation columns in visual cortex. Neuroreport 3, 69–72.