Class Specific Redundancies in Natural Images: A Theory of Extrastriate Visual Processing

Mohsen Malmir, Saeed Shiry

Abstract—Statistical properties of natural signals is an important factor in forming neuronal selectivities of brain sensory system. One such property is the redundancy in visual and auditory inputs to the brain. In this paper, we introduce the concept of class specific redundancies in natural images and propose that selectivity of neurons in extrastriate visual areas are developed to reveal these redundancies. In each extrastriate area, a redundancy reduction mechanism removes these revealed redundancies to provide a more efficient representation of the input image. To test this hypothesis, we implemented a model of area V2 and trained this model with a set of natural images. Experiments on artificial stimulus sets and natural images show the close similarity of model neurons to real V2 neurons and their preference for coding object contours over textures.

I. INTRODUCTION

Studies on statistical properties of natural signals have found that these signals are highly structured and redundant [1]–[3]. A number of studies have used these structures to reduce the effects of noise and represent the signals more efficiently [4]–[7]. Several authors have proposed that in a similar manner, sensory system in brain is adapted to natural signals and properties of sensory neurons are largely due to statistical regularities of sensory input [8]–[11]. Attneave hypothesized that the goal of the visual cortex is to produce an efficient representation of the input image [12]. Barlow in his “efficient coding” hypothesis, suggested that neurons in early sensory areas discard redundant data and supply higher order brain areas with the most informative features in their input [13].

Electrophysiological and neuroimaging studies have provided a wealth of information about the organization and function of neurons in the visual cortex. Neurons in the primary visual cortex are tuned to simple features like edge segments and have small receptive fields (RF) [14]. In the Inferotemporal cortex, the highest pure visual area, neurons have receptive fields that cover much of the retinal image and are activated maximally with complex shapes like faces [15]–[17]. Neurons in areas V2 and V4 have RF sizes between that of V1 and IT neurons and are selective to features of intermediate complexity like angles and contours [18], [19]. These results can be used to evaluate the properties of neurons derived from statistical modeling approaches to the visual cortex.

Primary works used linear transformations to reduce statistical dependency between image pixels. Sparse coding [20] and ICA [21] are classical works that with the aim of reducing redundancies in images, resulted in Gabor like basis functions that resembled V1 classical RF selectivities. However, these models had an inherent limitation in representing images as a linear superposition of basis functions. Analysis on the statistics of linear filters such as Gabor or wavelet revealed that a form of nonlinear dependency exists between their responses. It was observed that variance of response of a filter changes with responses of its neighboring filters [3], [4].

Wainwright and Simoncelli proposed the Gaussian scale mixtures (GSM) model which explained the variance dependency as multiplying a common variance source to independent Gaussian variables [22]. Schwartz and Simoncelli divided the responses of each filter to its variance to achieve independent Gaussian sources and modeled contrast invariant orientation tuning and gain control mechanism observed in V1 neurons [23]. Based on the GSM model, generative models have been developed that learned higher order structures in natural images by capturing the common variance sources among filters responses [24]–[26]. Recent models concern the problem of assigning variance sources to different Gaussian signals in order to produce different images [27], [28].

All the aforementioned models have simulated some properties of simple and complex cells in the primary visual cortex. Models based on local variance dependency cannot be extended to simulate the properties of neurons in extrastriate visual areas. In this paper, we propose the idea of class specific redundancies by looking into the statistics of natural images and the organization of neurons in the visual cortex. We present a model and train it with a set of natural images. Results of experiments demonstrate the close similarity between selectivity of neurons in the proposed model and neurons in the visual cortex.

II. CLASS SPECIFIC REDUNDANCIES IN NATURAL IMAGES

An important property of variance dependency between responses of filters is the size of neighborhood in which filters are dependent. Usually it is assumed that filters, clustered in small regions, are dependent and filters out of these local neighborhoods are independent [23], [29]. Divisive normalization, topographic ICA and other models
use a nonlinear transformation to remove these local redundancies. It is assumed that after local redundancy reduction, responses of filters all over the image are independent and therefore an efficient representation is produced. However, these models do not specify the size of dependency neighborhood. As we discuss below, this depends on the images that evoke filters responses.

Fig. 1 displays the joint histogram of 3 pairs of Gabor filters over 3 different sets of images from the 101 image dataset [47]. Each row displays the histograms for one relative spatial distance between filters with increase in distance from the upper to the lower row. Histograms in 3 columns from left to right are calculated from images of random objects, face images and airplane images respectively. It is obvious that filters in the first row exhibit variance dependency. In the second row, filters display variance dependency for face and airplane images but not for images from different categories. The filters in the third row display dependency only for airplane images. It seems that filters in the second and third row are dependent only for specific images. Previous studies that measured variance dependency did not use images from different categories to calculate the joint histograms. Here we used images from different categories to reveal the dependency between filters at large relative distances.

The relation between dependency neighborhood and specific images become clear from Fig.1. Filters with relatively small distance exhibit dependency over images for all categories. As the distance between filters increases, variance dependency becomes a function of the images that elicit filters responses. This is the case in the second and third rows, where filters do not exhibit variance dependency over images from random categories. For the pair of filters with largest distance, only images of airplanes could reveal that they are dependent. Therefore if we set only a fixed dependency neighborhood for filters then class specific dependencies cannot be detected and removed.

Variance dependency between filters has an intuitive explanation: responses of filters in the dependency neighborhood of a filter provide information about the responses of that filter. Responses of edge selective filters in local neighborhoods are dependent because edges in natural images usually vary smoothly. Therefore based on the response of an edge selective filter, it is possible to make guess about the responses of other filters in its neighborhood. Whenever the images that elicit filters responses are selected from different categories, size of this neighborhood reduces. However, if we limit the category of images, then this dependency neighborhood expands. If the images are from a specific category then one can guess about other filters responses in a larger area. This is the case shown in second and third rows of Fig. 1. For example, if we know that the image is an airplane, then based on responses of a filter we can conjecture about the responses of other filters in a fairly larger area.

Now we take a look to the selectivity of neurons along the ventral visual pathway for further clarification. As we mentioned before, size of RF and complexity of preferred stimuli increase along the ventral visual pathway [30]. Neurons in V1 fire in response to small spatial contrast changes, while neurons in the Fusiform Face Area fire when a face is present in their RF [31]. Neurons in areas V2 and V4 have preferred stimuli of intermediate complexity between V1 and IT neurons. It means that a large number of objects in image can elicit V1 neuronal responses, but IT neurons fire only in response to specific objects.

We propose that neurons in successive areas of the ventral visual pathway reveal redundancies that were hidden in the previous areas. When the complexity of neuronal selectivity increases from V1 to IT, dependency neighborhoods of neurons in images grow larger. It becomes possible then to remove these revealed redundancies through horizontal connections because extrastriate neurons with similar preferred stimuli are clustered together [32]. Therefore the role of hierarchical processing in the visual cortex is to remove redundancies in different spatial extents and provide a more efficient representation of the input image. Redundancy reduction starts in V1 in small vicinities for all images. In successive areas, redundancy reduction becomes specific to classes of images that are defined by the selectivity of neurons. In the final stages, redundancies for images of specific objects are removed and results are provided to the prefrontal cortex for learning and decision making.

One remaining question is how such selectivities develop in extrastriate neurons. It is well known that neurons in the ventral visual cortex have two types of receptive field: Classical RF (CRF) and extra Classical RF (eCRF). The configuration of CRF preferred stimulus is determined mainly by afferent connections from the next area, for
example CRF selectivities of V1 neurons are determined by afferent connections from LGN. There is increasing evidence that eCRF selectivities are formed by interactions between neighboring neurons through horizontal connections [33]. An important point to note is that eCRF selectivity of a neuron develops after its CRF selectivity has developed [34]. Another point is that CRF selectivity in an area develops after the eCRF selectivity of neurons in the next area has been established, e.g., V2 neuronal selectivities develop after eCRF selectivities of V1 neurons are formed [34]. The order of CRF and eCRF development may help describe selectivities of neurons in the ventral visual cortex.

V2 CRF selectivity develops after the horizontal connections in V1 have been established. It means that V2 neurons learn the results of redundancy reduction in area V1. After learning bottom up connections from V1, horizontal connections between V2 neurons emerge which remove revealed redundancies in larger vicinities of the input image. This cycle repeats for areas V4 and IT until their CRF and eCRF selectivities are established. Therefore increasing complexity of preferred stimuli for neurons in the ventral visual cortex is a result of iterative redundancy reduction.

### III. Model Description

We implemented a model of the visual cortex with two layers analogous to visual areas V1 and V2 to test the proposed hypothesis. Neurons in each layer are connected to other neurons in a neighborhood through horizontal connections. Neurons in area V2 receive bottom-up input from V1 neurons. A nonlinear divisive normalization mechanism is used to remove redundancies between responses of neighboring neurons in each layer. Weights for divisive normalization are learned from a set of natural images. Table I summarizes the parameters of the model.

**Table I: Model Parameters**

<table>
<thead>
<tr>
<th>Area</th>
<th>Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1 layer</td>
<td>101×101</td>
</tr>
<tr>
<td>V1 horizontal weights</td>
<td>25×25</td>
</tr>
<tr>
<td>V2 layer</td>
<td>77×77</td>
</tr>
<tr>
<td>V2 horizontal weights</td>
<td>25×25</td>
</tr>
<tr>
<td>V2 bottom up from V1</td>
<td>29×29</td>
</tr>
</tbody>
</table>

Neurons in area V1 of the proposed model are implemented as a set of Gabor filters with different orientations. Gabor function is product of a Gaussian function and a sinusoidal grating [35]:

\[
\begin{align*}
F(x, y) &= \exp\left(-\frac{x_0^2 + y_0^2}{2\sigma^2}\right)\cos((2\pi x_0)/\lambda) \\
&= x \cos(\theta) + y \sin(\theta) \\
&= -x \sin(\theta) + y \cos(\theta)
\end{align*}
\]

Where \(x\) and \(y\) are 2D coordinates of the filter in image, \(\gamma\) is the aspect ratio, \(\theta\) is the orientation, \(\sigma\) is the effective width and \(\lambda\) is the wavelength of the filter. We chose the number of orientations to be 12, which is thought to be a good estimate of the number of orientations in V1 [36].

To remove redundancies between responses of neighboring neurons, we adapted the model proposed by Schwartz and Simoncelli [23]. We used this mechanism because the normalization behavior has been observed in responses of neurons in different areas of the visual cortex [37]-[39]. In their work, Schwartz and Simoncelli assumed that variance of responses of each neuron can be modeled as a weighted sum of rectified responses of its neighboring neurons:

\[
\text{var}(L_x | L_{x}, Y \in C_x) = \sum_{Y \in C_x} w(x, y) L_y^2 + \sigma_x^2
\]

where \(L_x\) and \(L_y\) are responses of neurons \(x\) and \(y\) respectively and \(w(x, y)\) is the weight of connection between them, \(C_x\) is the neighboring region around \(x\) and \(\sigma_x^2\) is the part of variance of neuron \(x\) that is independent of other neurons responses. To remove variance dependency in local neighborhoods, responses of neurons are squared and divided to their variances:

\[
R_x = \frac{L_x^2}{\sum_{Y \in C_x} w(x, y) L_y^2 + \sigma_x^2}
\]

The result of (3) is the independent response \(R_x\) of filter \(x\).

#### A. Learning the Weights of Horizontal Connections in Local Neighborhoods

In their original work, Schwartz and Simoncelli used maximum likelihood estimate to learn the weights of connections between neighboring neurons. Calculation of MLE for a neuron with 25*25 weights over 1000 image patches takes long for hours, let alone the calculation for the entire population of \(\approx 15000\) neurons in the proposed model. Instead, we used a more efficient learning rule which is based on variance estimation for random variables. Given a set of \(N\) samples \(z_i, 1 \leq i \leq N\) of a random variable \(Z\), unbiased estimates for its mean and variance are:

\[
\begin{align*}
\bar{\mu}_Z &= \frac{1}{N} \sum_{i=1}^{N} \hat{z}_i \\
\bar{\sigma}_Z^2 &= \frac{1}{N-1} \sum_{i=1}^{N} (\hat{z}_i - \bar{\mu}_Z)^2
\end{align*}
\]

we used an iterative learning procedure to update the mean and variance estimates for each neuron:

\[
\begin{align*}
\bar{\mu}_x^t &= \left(\bar{\mu}_x^{t-1} \times (t-1) + L_x^t\right)/t \\
\bar{\sigma}_x^2 &= \left(\bar{\sigma}_x^{t-1} \times (t-2) + (L_x^t - \bar{\mu}_x^t)^2\right)/(t-1)
\end{align*}
\]

here \(t\) is the iteration counter and \(L_x^t\) denotes the response of neuron \(x\) to the image in iteration \(t\). Equation (5) is simply an iterative form of (4). Given these estimates, we use a learning rule to minimize MSE between variance estimation
of (2) and (5) over a set of natural images. MSE for neuron \( X \) over a set of \( N \) images is:

\[
E^2 = \frac{1}{N} \sum_{i=1}^{N} \left( Var(L^i_X) - \left( \sigma^2_X \right) \right)^2
\]

(6)

here \( Var(L^i_X) \) is calculated using (2) for image \( i \). We use gradient descent to find the minimum MSE of (6):

\[
\frac{\partial E^2}{\partial w_{xy}} = \frac{2}{N} \sum_{i=1}^{N} \left[ Var(L^i_X) - \left( \sigma^2_X \right) \right] \times \frac{\partial Var(L^i_X)}{\partial w_{xy}} = 0
\]

(7)

\[
w_{xy} = w_{xy}^{i-1} - \eta \times \frac{\partial E^2}{\partial w_{xy}} = w_{xy}^{i-1} - \eta \times \left( Var(L^i_X) - \left( \sigma^2_X \right) \right) \times L^i_X
\]

\[
\frac{\partial E^2}{\partial \sigma^2_x} = \frac{2}{N} \sum_{i=1}^{N} \left( Var(L^i_X) - \left( \sigma^2_X \right) \right) \times \frac{\partial Var(L^i_X)}{\partial \sigma^2_x} = 0
\]

(8)

\[
\sigma^2_x = \sigma^2_x^{i-1} - \eta \times \frac{\partial E^2}{\partial \sigma^2_x} = \sigma^2_x^{i-1} - \eta \times \left( Var(L^i_X) - \left( \sigma^2_X \right) \right) \times \sigma^2_x^{i-1}
\]

where \( \eta \) is the learning rate. We used a bell shaped function for learning rate, because in first iterations of the learning procedure, mean and variance estimates of (5) are not accurate. The learning rate for these estimates is low and increases as the estimate in (5) approaches to the actual mean and variance. To avoid divergence, learning rate again decreases for last iterations. Using rules in (7) and (8) is computationally and memory efficient because only the last mean and variance estimates for each neuron must be kept in memory. Fig.2 displays a set of weights learned for V1 neurons with different orientations.

**B. Learning V2 Bottom-up Weights**

Weights of connections between V1 and V2 are adjusted with a simple Hebbian rule. We used a set of fixed horizontal weights during learning V2 bottom up connections to calculate the inhibitory divisive signals from surround of each V2 neuron. This mechanism is supported by a recent experiment that found the horizontal connections between V2 neurons are not selective before their bottom-up selectivities are established [34]. Learning horizontal connections in area V2 occur after bottom up weights have been learned. Mechanism for learning these horizontal connections is similar to that in V1. Final response of v2 neurons are calculated with divisive normalization acting upon activities arrived from V1 bottom up connections.

**IV. COMPARISON WITH V2 NEURONS**

Visual area V2 is an intermediate stage between V1 and Inferotemporal cortex in the ventral visual pathway. It is proposed that neurons in this area are tuned to visual features like contours and junctions [18]. Despite the importance and relatively large size of V2, there are few studies that have probed the selectivity of its neurons with a carefully selected set of stimulus [18], [40]-[42]. Hedge and Hoyer used a set of gratings and contours to compare the responses of V2 neurons to textures and contours in images [18], [40], [42]. It was observed that population of V2 neurons exhibit a stronger response to contour segments than gratings. In another study, Ito and Komatsu used a set of angles with different orientations and angle sizes to find if angle stimuli could elicit V2 neuronal responses [41]. Results indicated that a large portion of neurons in this area are provoked by angle stimuli. We used these two stimulus sets to compare the neurons in the proposed model with V2 neurons.

Fig. 2. Samples of horizontal weights learned using (7) from a set of natural images.

Fig. 3. Responses of 3 model neurons to stimulus set of [18]. The first row displays the responses of a neuron selective to contour stimuli. The second row is responses of a neuron selective to gratings. Responses in the third row depict a neuron with no specific selectivity to stimuli. These 3 types of neurons were observed in V2 neurons [18].
In the first experiment, we used a set of 128 stimuli consisting of sinusoidal gratings with different frequencies along with sharp and acute curved lines and angles. Responses of three typical model neurons are shown in Fig. 3. The first row in Fig. 3 displays the responses of a neuron selective to contours. This neuron exhibits no response to grating stimuli. The second neuron is provoked largely by gratings but not with contours and angles. The third neuron shows no specific pattern of selectivity and a large number of stimuli elicited response from this neuron. The same three types of neurons were reported in [18].

In the second experiment, we used the set of 66 angles that Ito and Komatsu employed to study the selectivity of V2 neurons. This set is shown in Fig. 4. In the first part of this experiment, the sizes of peak response area for a population of 202 model V2 neurons were measured. This size is an indicator of the specificity of neuronal selectivity to angle stimuli and is equal to the number of angles around the optimal stimulus which elicited neuronal responses larger than 50% of the maximum response. The results are shown in Fig. 5. It can be seen that the distribution of the size of peak response area in the model is very similar to the results obtained from V2 neurons.

Ito and Komatsu found that a large portion of V2 neurons were selective to the sharp angles. This result is consistent with the finding of psychophysical studies that sharp angles are perceptually more salient relative to wide angles [43]. We measured the distribution of optimal angle stimulus over the same population of 202 neurons in area V2 of the model. Fig. 6 displays the results of this experiment and the one reported in [41]. It is obvious that the same pattern of dominance exists in the V2 model neurons. Therefore we can say that the sharp angles are more perceptually salient because they are statistically efficient features in images.

A. Coding Object Contour in Model V2 Neurons

Accumulated evidence from neurophysiological experiments indicates the role of V2 neurons in coding object contours [18], [42]. It is proposed that neurons along the ventral visual pathway code the object contours [44]. Results of different experiments also provide evidence to the fact that V2 neurons prefer contour stimuli over gratings [42]. Therefore we designed an experiment to test the role of model V2 neurons in coding object contours. A set of natural images, previously used to test the contour integration model of V1 neurons [45] were provided as input to the model. For each image, we calculated the final V2 neuronal activities and then provided V1 neurons with a feedback from active V2 neuron. The weights of connections for these feedback signals were selected to be the same as the bottom up weights between V1 and V2. Activities of V1 neurons were modulated whenever they were active and also received a feedback signal from area V2. Results of this experiment are shown in Fig. 7. It is obvious that feedback signals from area V2 to V1 emphasized the response of V1 neurons that were along the edges of objects.

V. CONCLUSION

Selectivity and functional properties of neurons in the extrastriate visual areas have been examined for several years. A number of studies have examined the selectivity of neurons in different areas of the visual cortex and provided useful information about possible processes in the object recognition pathway. In this paper, we proposed that selectivity of neurons in the ventral visual pathway is determined by redundancy reduction in different layers. This hierarchical redundancy reduction results in the most informative features over the entire image.

It is proposed that neurons along the ventral visual...
pathway process object contour information. Experiments have provided evidence that angles are important in object recognition and removing corners from images severely impairs the object recognition ability of human. In accordance with these results, we showed that model V2 neurons preferred to code object contours over textures in a set of natural images.

Future work should address the problem of developing neurons with selectivities similar to V4 and IT neurons. An important issue is that neurons in IT form clusters in which neurons prefer a similar stimulus. To model such selectivities, one must use a set of images from the same object that have been taken from different viewpoints. Another important problem is to examine the pattern recognition ability using features developed in this model. It is proposed that based on the features extracted in IT, a simple linear classifier can perform object recognition with high accuracy [46]. Future works should answer the question that whether the efficient features are the ones that provide such recognition performance.

REFERENCES


