

# 3G3 Coursework: Coding In Visual Cortex

Tom Xiaoding Lu  
xl402@cam.ac.uk  
Pembroke College

**Abstract**—This coursework investigates two visual image coding mechanisms, the *compact coding* and the *sparse distributed coding* mechanisms. Through image reconstruction using the basis generated by these two mechanisms, the underlying principles involved in sensory information processing used in the visual cortex can be uncovered.

## I. INTRODUCTION & BACKGROUND INFORMATION

The visual system can exploit the structure of natural images' state space, and use efficient coding schemes to achieve image encoding and reconstruction.

Consider the state-space of a receptor array such as the photoreceptors in the retina, modelling the receptive field as an array composed of  $N$  receptors, each of which can represent any value within a range of luminance, each image can then be represented as a single point in such high dimensional array. Although the number of possible images that can be represented is vast, typical (natural) images do not span the entire state space.

By taking adjacent pixels at random locations in natural images and plotting their intensities against each other, figure 1(a) and figure 1(b) show the correlation between neighbouring pixels of random and structured images respectively.

Clearly, for structured images, there is a linear, positive

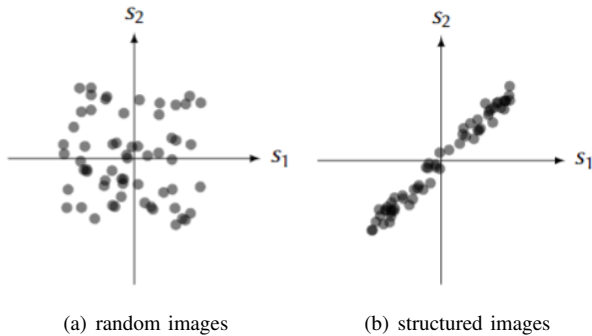


Fig. 1. Correlation between neighbouring pixels

correlation between neighbouring pixels, as the structured (natural) images mainly consist of blocks of colours and colour gradients; edges which result in sharp discrepancies between neighbouring pixels are rare.

As natural images are highly structured, the visual system can exploit such structures through efficient coding mechanisms. Two possibilities, *Compact coding* and *Sparse distributed coding* are investigated and are illustrated in figure 2.

The *Compact coding* scheme proposes using minimal number of neurons to encode the output of the retinal receptors.

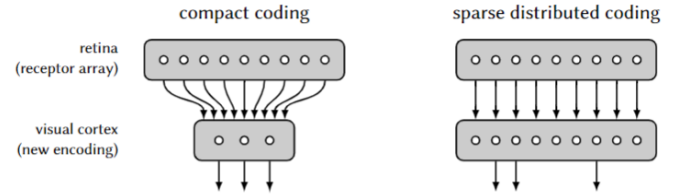


Fig. 2. Two different forms of coding the retinal receptor array to activity in visual cortex

The *Sparse distributed coding* scheme aims to decrease the average activity of neurons in encoding the retinal receptors output.

The aim of the coursework is to compare the retinal receptive field representation through the two coding schemes with the receptive field of a neuron in V1 obtained experimentally. Simple cells in V1 have receptive fields with the following characteristics:

- 1) Localized within a small region of retinal space
- 2) Respond best to a bar of light at a particular position, orientation and width

The responses of simple cells to light can be computed from convolving the retinal image with the two-dimensional filters shown in figure 3, which are the filters the theoretical model aims to produce.

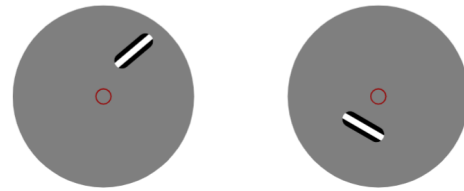


Fig. 3. The receptive fields of two typical simple cells in V1, acting as linear filters

## II. IMAGE RECONSTRUCTION FROM IMAGE BASES

We assume that a given retinal image in visual space can be reconstructed through a **linear combination** of basis functions given by V1 neuron receptive fields. The weights can be seen as the activations of the corresponding V1 neurons.

Given an image  $S(x, y)$ , the neural population output which approximates the image is denoted  $\hat{S}(x, y)$ . This output is constructed through the superposition of a set of neurons, each  $i$ th neuron contributes a basis function  $\phi_i(x, y)$ , with

its activation denoted as  $a_i$ . Equation 1 encapsulates the relationship between individual neuron's response and neural population output (approximated image).

$$\hat{S}(x, y) = \sum_i a_i \phi_i(x, y) \quad (1)$$

### A. Compact Coding

Principal Component Analysis (PCA) is a dimensionality reduction technique, used to find a reduced set of orthogonal vectors that encapsulate the directions of maximum variance of the original retinal state space. These vectors can be used as basis functions used to achieve compact coding. In the

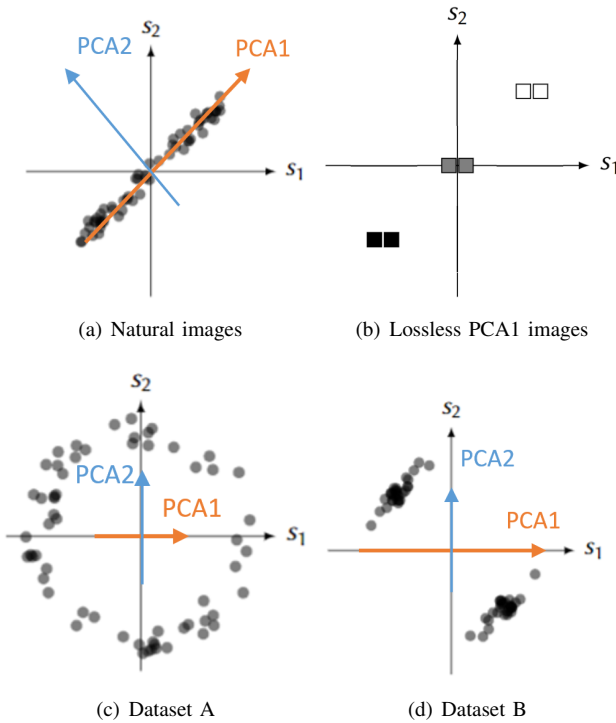


Fig. 4. Principal components of different datasets

simplified example of two-pixel natural images shown in figure 4(a), the axis along the positive diagonal is the first principal component vector as it is the most informative direction. Assuming it also represents the distribution of natural 2-pixel images, figure 4(b) represents examples of 2-pixel images that could be represented without error using only PCA1.

PCA has limitations; in figure 4(c) and 4(d), when a given data set is not linearly distributed but might be arranged along with non-orthogonal axes or well described by a geometric parameter, PCA could fail to represent and recover original data from projected variables. In these cases, feature transformation is required to map the data onto other dimensions.

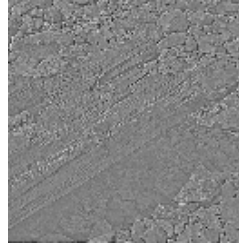
In order to test the different coding mechanisms, different image sets are created as training data for forming basis functions. Figure 5(a) and figure 5(b) are natural images of text and natural scene respectively. Images from I2w set, with an example shown in figure 5(c) are the gradient of the I2 images. Images in I3 are generated white noise, used to further

ugh our senses which provide information that  
uations, termed noise, and may provide ambiguous in  
ie environment. Moreover, when we act on the wov  
mands we send to our muscles are also corrupted by v  
ary and motor variability limits the precision with wh  
d. Here we will review the framework of Bayesian d  
principled approach to handle uncertain information  
how this framework can be used to understand sensory



(a) I1: Text image

(b) I2: Natural scene



(c) I2w: Natural scene gra-  
dient



(d) I3: White noise

Fig. 5. Samples from training images

validate the coding schemes.

By performing PCA on the image sets I1, I2 and I3, their major principal components and the amount of variance that each basis function accounts for on a log-log scale are shown in figure 6. The minimum number of principal components needed to account for over 90% of total variance is 102, 12, 206 respectively for the three image sets. This result should not be surprising as natural scenes in I2 are most structured, text image in I1 is also structured, however, characters have finer defining features than natural images. The images in I3 are generated noise which is unstructured, hence PCA is unable to reduce the dimensions of the image space.

To investigate the image features which the basis functions extract through PCA, the example images are convolved with the top three basis functions accounting for most amount variance. Figure 7 shows the top three basis functions (kernels) for I1. The text image, figure 7(a) shows a horizontal gap extraction kernel, responsible for detecting gaps between lines. Figure 7(b) shows a kernel which responds to the lines of texts themselves. Figure 7(c) shows a kernel which responds to gaps between characters. Intuitively, the basic structure of passages from a corpus consist of mainly these three features, gaps between lines, lines of texts and gaps between characters. Similarly, for I2 natural images set, figure 8 shows that the first basis function is a low pass filter, extracting the smoothed image. The second and third basis functions extract horizontal and vertical edges of the image respectively. The higher order basis functions become more complicated and each account for less variance.

PCA decomposition deconstructs input images into a linear combination of basis functions which can be seen as basic image features. It is seen that through performing PCA on natural images, the most important basis function shown in figure 8(a) matches with the simple cells' receptive fields from the

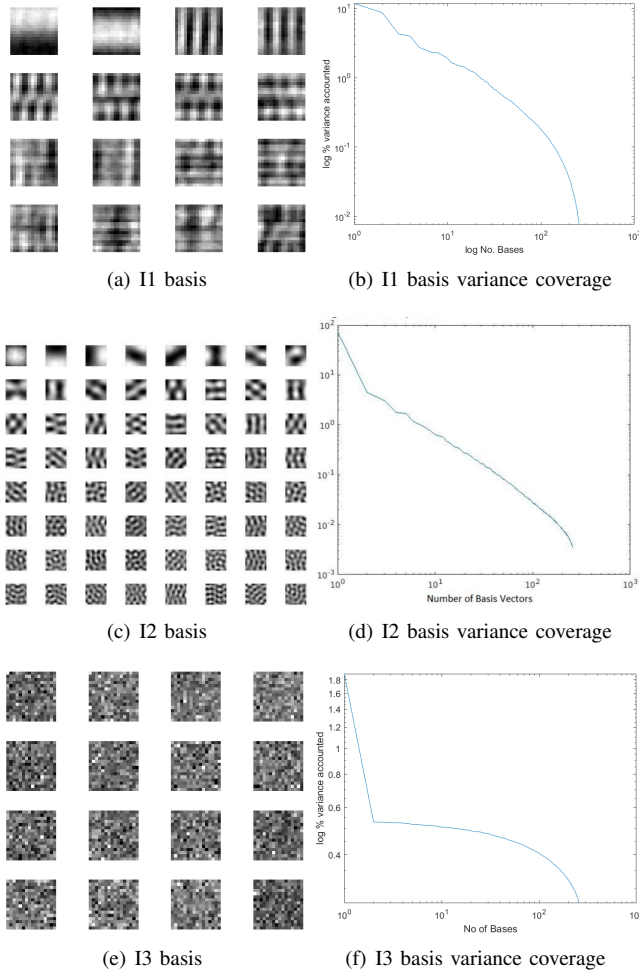


Fig. 6. PCA basis functions and coverage for different image sets

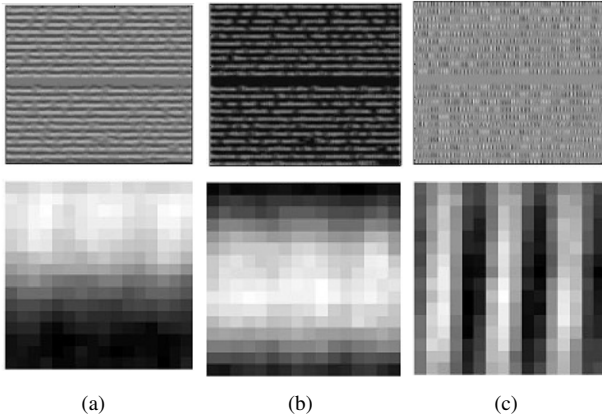


Fig. 7. I1 convolved images and their corresponding features

Lateral Geniculate Nucleus (LGN), which provide on-centre-off surround responses. However, simple cells' responses from the V1-cortex is not simulated by higher-order basis functions, therefore compact coding through PCA is likely not to be a good model for modelling receptive fields from the visual cortex.

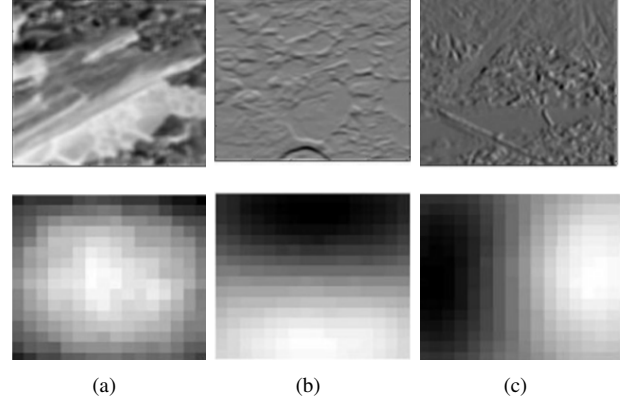


Fig. 8. I2 convolved images and their corresponding features

### B. Sparse Distributed Coding

If PCA used in compact coding is seen as a solution to an optimization problem, its complicated higher order basis functions can be seen as overfitting solutions. In sparse distributed coding, we introduce a regularization term which aims to reduce the total amount of activation (weights) used when reconstruction the retinal space from basis functions. This regularization can also be seen as a prior from the Bayesian perspective. From equation 1, sparse distributed coding's cost function can be written as:

$$\begin{aligned} \mathcal{C}(\{\phi_i\}, \{a_i\}) &= \sum_{x,y} [S(x,y) - \hat{S}(x,y)]^2 + \lambda \sum_i H(a_i) \quad (2) \\ &= (\mathbf{s} - \mathbf{B}\mathbf{a})^T (\mathbf{s} - \mathbf{B}\mathbf{a}) + \lambda \sum_i H(a_i) \quad (3) \end{aligned}$$

Where  $H(a_i) = \log(1 + \frac{a_i^2}{\sigma^2})$  chosen as the regularization term as it is smooth (differentiable) and monotonically increasing. To speed up computation time, equation 2 is vectorized, the matrix  $\mathbf{B}$  has row vectors as basis functions and vector  $\mathbf{a}$  stores all activations for different basis functions.

The cost function is nonlinear because  $H(a_i)$  is nonlinear, standard gradient descent techniques can be used to minimize the cost function with respect to both basis function matrix  $\mathbf{B}$  and activations  $\mathbf{a}$ :

$$\frac{\partial \mathcal{C}}{\partial \mathbf{a}} = \mathbf{s}\mathbf{B}^T\mathbf{B}\mathbf{a} - 2\mathbf{B}^T\mathbf{s} + 2\lambda \frac{\mathbf{a}}{\sigma^2 + \mathbf{a}^2} \quad (4)$$

$$\frac{\partial \mathcal{C}}{\partial \mathbf{B}} = 2\mathbf{B}\mathbf{a}\mathbf{a}^T - 2\mathbf{s}\mathbf{a}^T \quad (5)$$

After calculating the gradients, the gradient descent algorithm can be implemented through:

- 1) Starting with a random set of basis functions (randomized  $\mathbf{B}$  matrix).
- 2) Given this set of fixed basis functions, find the set of activations which minimizes the cost  $\mathcal{C}$  using equation 4.
- 3) Update the basis functions through minimizing  $\mathcal{C}$  using equation 5

The gradient descent method is implemented in MATLAB. Using library function `minimize()`, optimizing the parameters over 500 iterations took 0.756 seconds, which is slow due to

MATLAB's Just-In-Time compilation. The conjugate gradient descent method implemented in C is used instead; optimization over 500 iterations took 2.63 *ms*, which is two magnitudes faster.

### C. Stochastic Gradient Descent and Batch Normalization

Stochastic gradient descent is implemented in order to minimize the cost function in batches. At each time step, subimages are extracted by sampling the training image set at random, and the gradient descent algorithm is used to update  $B$  and  $a$ .

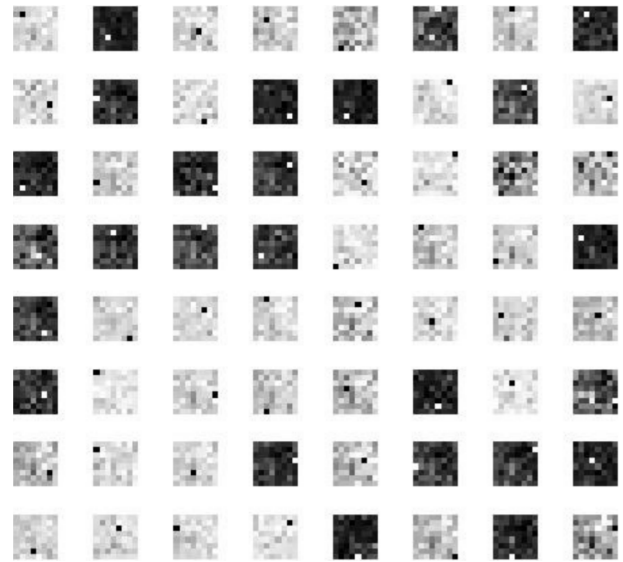
Without batch normalization, the values in the basis functions would tend to grow without bound due to gradient explosion, as an early on large activation causes a large output response from the neuron (basis function), which cascades through each training step. This problem can be solved by re-scaling the magnitude of each basis function after each time step. The re-scaling factor for each basis function is proportional to the amount required to stretch its activation to some arbitrary target activation. This way we constrain the activations after each batch to be centered around some target activation so the problem of exploding gradient is solved. This way higher learning rates can also be afforded as batch normalization also reduces dependence on parameter scale.

By finding the basis functions using the sparse code mechanism, over training datasets I3 and I2w, figure 9 shows the 64 extracted basis functions. For the dataset I3, it is observed that the basis functions extracted are very different from PCA basis functions, they are either only responsive to a single pixel in a subimage or responsive to all but one pixel in a subimage. This suggests that I3 was generated artificially to have uniformly distributed random high intensity pixels on a background of low intensity pixels. The I3 basis functions also validate the performance of the sparse code mechanism, as the way to represent a typical I3 image with minimum number of neurons is to have one neuron responsible for one pixel in the subimage.

Finally, basis functions are extracted from natural images using the 'sphered' version of I2, in which variances in different directions have been equalized (this is seen as an input features normalization). Comparing these basis functions to the simple cell model shown in figure 3, we can see that these basis functions are very representative of the simple cells' receptive fields. They are responsive towards specific gradients at specific locations, which can be characterized as being localized, oriented and bandpass.

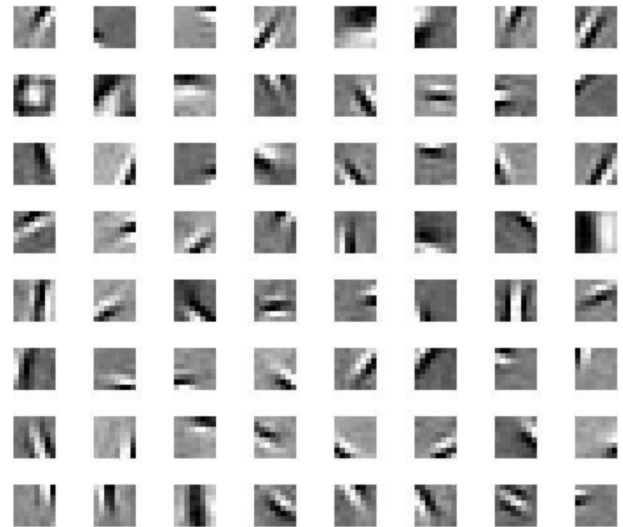
## III. CONCLUSIONS

This coursework aims to predict the receptive field of simple cells through compact coding and sparse coding mechanisms. Through comparing model output with results from neurobiological experiments, we can confirm that compact coding through PCA is a simple way of approximating the receptive field, however, its higher-order basis functions are not representative. Sparse coding is the most likely model to be adapted by the visual cortex as it produces basis functions with similar



(a) I3 noise image set

Fig. 9. Spars code basis functions



(a) I2w natural image set

responses compared with the simple cells, which are localized, oriented and bandpass. The neurobiological implications of sparse coding is not discussed within the coursework. Possible explanations involve better signal to noise ratio, energy efficiency and feature detection which assists recognition in later stages of the visual cortex.

The activation regularization function  $H(a_i)$  in equation 2 is chosen without explanations. Further experiments can focus on the effects of using different regularization functions. Our model also assumes that the activations from individual neurons are reliable and noiseless, whereas in reality neural outputs are noisy. The unreliable activations can be accounted for in future experiments and coding strategies can expand upon sparse coding to achieve more representative coding strategy by the visual cortex.

## IV. APPENDIX: SELECTIONS OF MATLAB CODE

## A. Sparse Code cost and activation gradient

```
function [cost,dcost]=spfunc(a,B,s,sigma,  
    ↪ lambda)  
%s is the subimage vector  
%B is the base matrix  
%a is the activations  
% cost is the cost  
% dcost is the vector of the derivatives  
    ↪ of the cost with respect to the  
% activations  
cost= (s-B*a)' * (s-B*a) + lambda * sum(  
    ↪ log(1 + a.^2/sigma^2));  
dcost= 2*(B'*B)*a - 2*B'*s + 2*lambda * a  
    ↪ ./ (a.^2 + sigma^2);
```

## B. Sparse Code basis function gradients

```
function dcost=basefunc(a,B,s)  
%s is the subimage vector  
%B is the base matrix  
%a is the activity vector  
% dcost is the returned vector of the  
    ↪ partial derivatives  
% of the cost with respect to the bases  
% you should calculate dcost here  
  
dcost= 2* B * (a* a') - 2*s*a';
```