

Self-Organising Learning of Receptive Fields in Multi-Resolution

D. Deng^{†‡}, K.P. Chan[†], Y.L. Yu[†]

[†](The Institute of Radio Engg. & Auto., South China University of Technology, GuangZhou, 510641, China)

[‡](Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong)

ecddeng@scut.edu.cn, kpchan@csd.hku.hk, ecyllyu@scut.edu.cn

Abstract

The statistical likelihood of Gabor filters and primary visual cortex has been of interest for years, yet learning mechanisms proposed did not generate satisfactory Gabor-like receptive fields. In this paper, a new computational model of self-organised Hebbian learning (SOHL) is proposed to work on a multi-resolution image pyramid for the problem of visual receptive field learning. Receptive fields of both orientation and spatial frequency selectivity as observed in our simulation result.

1 Introduction

Simple cells in the primary visual cortex, first detected by Hubel and Wiesel early in 1962, are found to behave approximately like 2-D Gabor filters in spatial domain and spatial frequency domain. With these effective representation of location, orientation and spatial frequency content of the image, it is believed that shape and texture analyses start at these simple cells. And a series of publications by Daugman [2] et al. on mathematic framework eventually draw great attention in the research field of Gabor filtering applications.

The statistical likelihood of Gabor filters and receptive fields (RFs) of primary visual cortex gives rise to the question that how such resemblance could be developed in primary visual systems. In [3] Pattison pointed out the connection between Földiák's Hebbian learning scheme [4] and image reconstruction with Gabor functions by a minimum mean square error (MMSE) criterion. However, experiments carried on Földiák's algorithm have failed to produce Gabor filter-like weights on the feed-forward connections. And the MMSE criterion which Pattison employed is biologically questionable, since it is quite clear that the image is not reconstructed with sheer fidelity in human brain. In [5], Sanger tried to learn receptive fields in a single layered feed-forward network with GHA carried for

K-L Transform. Orthogonality between RFs was resulted and some patterns found no biological explanation.

Barrow [6] has proposed another model of receptive field learning. He applied plain Hebbian learning rule for RFs operating on a band-passed white noise image as well as a natural image. A Winner-Takes-All strategy was used so that for a certain input only the RF which strongestly responds to it would be updated. Compared with Sanger's model, this approach with competitive learning is more of biological plausibility. However, although the model is successfully applied on natural images, considerable limitation exist where

- The RF window size is limited so that the window never covers two edges or blobs at the same time. Such a limitation restrains the possibility to learn information of spatial frequency in the image.
- There is no guarantee on the competitive learning between different units so as to cope with ill initialization and under-learning, and after convergence not all patterns learned are meaningful and useful for image analysis.

2 Computational Model

2.1 Self-organised Hebbian learning: optimization analysis

Barrow's model uses plain competitive learning at the winning unit, and no cooperation between RFs exist. We argue that for the problem of RFs learning with orientation and frequency sensitivity, and learn from it in proportion to their responses. In this case we introduce a neighbourhood for each winning unit and learning is self-organised within the neighbourhood.

Let $p(x)$ denote the probability density function of input \mathbf{x} . If we want to reconstruct the input \mathbf{x}

from the active outputs, a cost or error function E can be defined as

$$E = \int \|\mathbf{x} - \sum_k h_{b,k} \mathbf{w}_k y_k\|^2 p(\mathbf{x}) d\mathbf{x} \quad (1)$$

where b is the index of best-matching unit. Usually $p(\mathbf{x})$ is unknown, and the optimization process is approximated by minimizing the following cost function for each input sample \mathbf{x} :

$$E_1 = \|\mathbf{x} - \sum_k h_{b,k} \mathbf{w}_k y_k\|^2 \quad (2)$$

Given the constraint $\sum_k h_{b,k} = 1$, it goes further to -

$$E_1 = \left\| \sum_k h_{b,k} (\mathbf{x} - y_k \mathbf{w}_k) \right\|^2 \leq \sum_k h_{b,k}^2 \|\mathbf{x} - y_k \mathbf{w}_k\|^2 \quad (3)$$

Define

$$E_2 = \sum_k h_{b,k}^2 \|\mathbf{x} - y_k \mathbf{w}_k\|^2 \quad (4)$$

We can see that, in the case of plain competition among output neurons, weight modification occurs only on the output node with maximum response, and the neighbourhood function $h(\cdot)$ shrinks into the impulse function, i.e.,

$$h_{b,k} = \delta_{b,k} = \begin{cases} 1 & \text{if } b = k \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Then the criterion function defined in Eqn.(2) is equivalent to Eqn.(4). In general, the cost function E_2 is the upper bound of E_1 . To minimize the cost function E_2 , We take the gradient decent of E_2 in Eqn.4. Notice that $y_k = \mathbf{w}_k^T \mathbf{x}$, we have

$$\begin{aligned} \nabla_{\mathbf{w}_i} E_2 &= -2h_{b,i}^2 \nabla_{\mathbf{w}_i} (y_i^2) + h_{b,i}^2 \nabla_{\mathbf{w}_i} (y_i^2 \mathbf{w}_i^T \mathbf{w}_i) \\ &= -2h_{b,i}^2 (2y_i \mathbf{x}) + h_{b,i}^2 (2y_i \mathbf{x} \|\mathbf{w}_i\|^2 + 2y_i^2 \mathbf{w}_i) \end{aligned} \quad (6)$$

If weight vectors are normalized during the learning process, i.e., $\|\mathbf{w}_i\| = 1$ for every i , it yields to

$$\nabla_{\mathbf{w}_i} E_2 = -2h_{b,i}^2 (y_i \mathbf{x} - y_i^2 \mathbf{w}_i) \quad (7)$$

Hence we obtain the following learning rule:

$$\Delta \mathbf{w}_i = \gamma h_{b,i}^2 (y_i \mathbf{x} - y_i^2 \mathbf{w}_i) \propto -\nabla_{\mathbf{w}_i} E_2 \quad (8)$$

where γ is the learning rate. From Eqn.(8) it can be easily seen that the learning algorithm incorporated in the computational model is a combination of the Oja rule and a process of self-organisation similar to Kohonen's approach [9]. We denote the algorithm as 'Self-organising Hebbian Learning' (SOHL).

Because of the locality of the SOHL algorithm, the criterion of minimizing the cost function E_2 is

not aimed at a perfect reconstruction of the original input, but to carry maximum information of the input onto the organised output layer. As shown in [8], the implementation of Hebbian learning under certain condition will gain a maximum information transmission rate on the output layer. And since that only the output cell with maximum response along with its neighbourhood modify their synapse weights at a time, it is unlikely that this competitive implementation will learn all the orthogonal eigenvectors of a input space, as Sanger commented [5]. Rather the input space diverge into several subspace within which the eigenvector for each of them is learned. Hence orthogonality among receptive fields is not necessary, which is biological supported.

3 Self-organised Learning of Receptive Fields

3.1 Multi-resolution image pyramid

A multi-resolution image pyramid is constructed following Burt [7] to provide sufficient spatial frequency information for RF learning. First, the original image I is reduced into a half-sized image I_r by convolving with a gaussian window g :

$$I_r(i, j) = \sum_{m=-2}^2 \sum_{n=-2}^2 g(m, n) I(2i + m, 2j + n) \quad (9)$$

Then the reduced image I_r is expanded into an image which is of the same size as the image I :

$$I_e = I - I_r = 4 \sum_{m=-2}^2 \sum_{n=-2}^2 g(m, n) I((i-m)/2, (j-n)/2) \quad (10)$$

And the Laplacian image I_l is obtained by

$$I_l = I - I_e \quad (11)$$

Fig.1 shows the Laplacian image sequence generated from the original 512x512 image. It can be shown that in Eqn.(11) the Laplacian image is equivalent to the original image convolved with a DOG function. Hence we consider the pyramid construction is equivalent to the pre-cortical processing in retina and LGN units as modelled in [6].

3.2 Frequency selectivity

For the sake of optimal training, input image patches are discriminated by their local frequencies, and patches in different frequency scale are used to train the corresponding RF, resulting in



Figure 1: *Laplacian Pyramid for Lena*

RFs of different orientation and frequency selectivity after competitive learning. A very simple yet fast method is used to estimate the local frequency. If the intensity value sequence within the image patch is treated as a waveform, with D.C. component drawn across it, and if we count the significant zero-crossings on both the x- and y- axes, the estimated frequency f can be calculated as follows:

$$f = \sqrt{(E(n_x)/2)^2 + (E(n_y)/2)^2} \quad (12)$$

where n_x and n_y are zero-crossing rates on x-axis and y-axis respectively, and $E(\cdot)$ means taking average. Zero-crossings within a threshold bound are ignored since they are of little significance in local frequency estimation.

3.3 Computation scheme with SOHL

The computation scheme is shown in Fig.2. At the first stage a Laplacian pyramid of the original image is generated. Meanwhile, RFs corresponding to each spatial frequency stage are randomly initialized. Then a small patch of image is sampled from the Laplacian pyramid, also randomly. The sampled image patch is classified to a certain frequency level after a frequency estimation process, and it is multiplied to the receptive units on the frequency level to generate the responses:

$$y_k = \sum_i^N \sum_j^N I_{ij} R_{k,ij} \quad (13)$$

where \mathbf{R}_k is one of the M RFs in the k -th frequency level, and N is the size of both the image patch \mathbf{I}_k and RFs.

The competitive learning among these M receptive units then picks out the unit with the maximum response, i.e.,

$$y_m^2 = \max_{i=1}^M y_k^2 \quad (14)$$

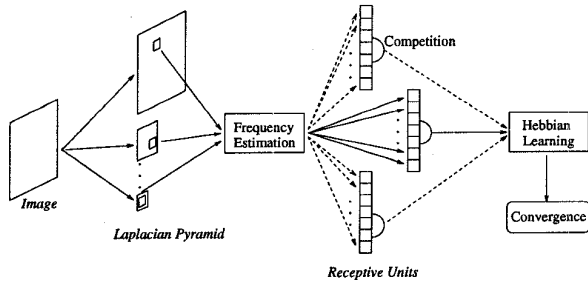


Figure 2: *To learn RFs from multi-resolution*

And the RFs in the neighbourhood of the winning unit m are modified by the SOHL rule:

$$\mathbf{R}'_i = \mathbf{R}_i + \alpha(y_i \mathbf{I}_i - y_i^2 \mathbf{R}_i) h(|i - m|, r) \quad (15)$$

where α is the neighbourhood radius which shrinks gradually with the time, and $h(\cdot)$ is a normalized gaussian neighbourhood function taking the form:

$$h(d, r) = e^{-d^2/2r^2} \quad (16)$$

4 Simulation Result and Discussion



Figure 3: *Grayscale images for training: 'Lena' and a composed brodatz texture image. Both are of 512 × 512 size.*

Due to the nature of the work we choose natural images with detailed frequency information for training. The standard image 'Lena' and a composed Brodatz texture image, both shown in Fig.3, are used in our experiment. Original images are of 512 × 512 size, and 2 cycles of reducing-expanding processes generate an image pyramid of 3 images, each of size 512 × 512, 256 × 256 and 128 × 128 respectively. The receptive fields are of 4 frequency levels, and on each level there are 12 competitive fields aimed for orientation selectivity, randomly initialized at beginning. Image patch of fixed 32 × 32 size are randomly sampled from the image pyramid and fed into the frequency selection

unit. After the frequency level is decided, receptive units on that level will learn from the input image patch through competition. The neighbourhood width begins at 4, and shrinks gradually to 0 during learning. The initial learning rate is $\gamma = 0.0015$. Fig.4 shows the intensity of RF patterns after training, where RFs on the same frequency level are lined up in a row. After convergence, the increasing average selectivity of winning units, defined as $y_{max} / \sum_i y_i$, is found to be saturated, as shown in Fig.5.



Figure 4: Receptive fields learned from multi-resolution

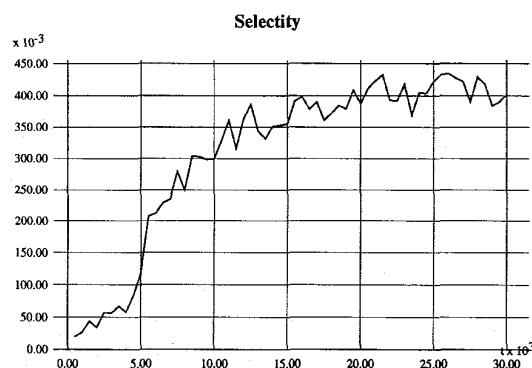


Figure 5: Selectivity of Output Units on Level 2

With the self-organisation implemented, the performance of our computation model is improved in that a much flatter firing rate distribution is gained, which implies that more receptive fields may learn from the input image patch, and therefore, more likely to develop finer orientation selectivity.

It can be seen in Fig.4 that although the receptive fields bear similar frequency and orientation selectivity, they are not strictly Gabor-like. RFs on higher frequency levels suffer from under-learning. Besides, orientations displayed in each frequency level may differ. We argue that these results indicate the dependency to the training image. In real visual systems, Gabor-like cortical synapse weights

can be developed by learning from a large population of visual stimulations, and therefore gain the independency to individual stimulation along with the adaptiveness to almost all of them. With the SOHL carried out on 2-dimensional directions, it is also possible that organised cortical column can be formed. To achieve such a goal, and to test the plausibility of applying SOHL in image analysis tasks, further work needs to be done.

References

- [1] Jones, J.P., Palmer, L.A., "An Evaluation of the two-dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex", *Journal of Neurophysiology*, **58**, pp.1133-1258, 1987.
- [2] Daugman, J.G., "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression", *IEEE Trans. on ASSP*, **36**, No. 7, pp. 1169-1179, 1988.
- [3] Pattison, T.R., "Relaxation Network for Gabor Image Decomposition", *Biological Cybernetics*, **67**, 97-102, 1992.
- [4] Földiák, P., "Adaptive network for optimal linear feature extraction", *Proceedings of IJCNN*, I, 401-406, 1989.
- [5] Sanger, T.D., "Optimal unsupervised learning in a single-layer linear feedforward neural network", *Neural Networks*, **2**, 459-473, 1989.
- [6] Barrow, H. "Learning Receptive Fields", *Proc. IEEE First Annual Conf. on Neural Networks*, 115-121., 1987.
- [7] Burt, P.J., "The Laplacian Pyramid as a Compact Image Code", *IEEE Trans. on Comm.*, **COM-31**, No. 4, 1983.
- [8] Haykin, S., *Neural Networks - A Comprehensive Foundation*, Macmillan, 1994.
- [9] Ritter, H., Martinetz, T., Schulten, K., *Neural Computation and Self-organising Maps*, Addison-Wesley Publishing Company, 1991.