Network 3 (1992) 61–70. Printed in the UK

The principal components of natural images

Peter J B Hancock, Roland J Baddeley and Leslie S Smith

Centre for Cognitive and Computational Neuroscience, Departments of Psychology and Computing Science, University of Stirling, Stirling FK9 4LA, UK

Received 15 November 1991

Abstract. A neural net was used to analyse samples of natural images and text. For the natural images, components resemble derivatives of Gaussian operators, similar to those found in visual cortex and inferred from psychophysics. While the results from natural images do not depend on scale, those from text images are highly scale dependent. Convolution of one of the text components with an original image shows that it is sensitive to inter-word gaps.

1. Introduction

We live in, and are required to make sense of, a complex visual world. One key to interpreting images is to know something of their statistics. The simplest kind of statistics, based on pixel grey levels, are first order: means, variances and probability distributions of brightness values. Such statistics are very useful, for instance, for setting thresholds. We can also ask more complicated questions, such as: how does the value of one pixel depend on that of its neighbours? In images of the real world, nearby pixels will often have common causes and thus be statistically related. A common method in statistics for analysing inter-relations between variables is factor analysis and the most basic form of this is principal component analysis (PCA).

The idea of PCA is to convey the most information about a set of data given a limited number of linear descriptors. High-dimensional data is projected onto a smaller number of dimensions which are chosen so as to maximize the variance on the new axes. The axes produced are mutually orthogonal. Given a two-dimensional probability distribution in the shape of an ellipse, PCA would return the two axes of the ellipse, the longer one first. The first principal component (PC) is the linear descriptor (normalized weighted sum), which gives the most information about the data, assuming a normal distribution and no knowledge of higher-order statistics. The second principal component will give the most additional information, given the value of the first. Thus one application of PCA is in data reduction: it may be possible to represent a data point sufficiently accurately by giving the values of only the first few PCs. In some cases exact representation may be possible. For example, data lying on a plane can be fully specified using only two variables. The two axes may be defined by simple linear combinations of the axes of the raw data.

In this study we are extracting principal components of 64×64 pixel pieces of natural images. The exact method of finding PCs is to find the eigenvectors of the correlation matrix of the input data. We have 4096 variables, which would produce

0954-898X/92/010061+10$04.50 © 1992 IOP Publishing Ltd

a matrix with 2^{24} entries: beyond reasonable computation. We are using a neural network technique developed by Sanger [1] that is able to find a good approximation to the solutions within a few hours on a 1 MFlop workstation.

2. Related work

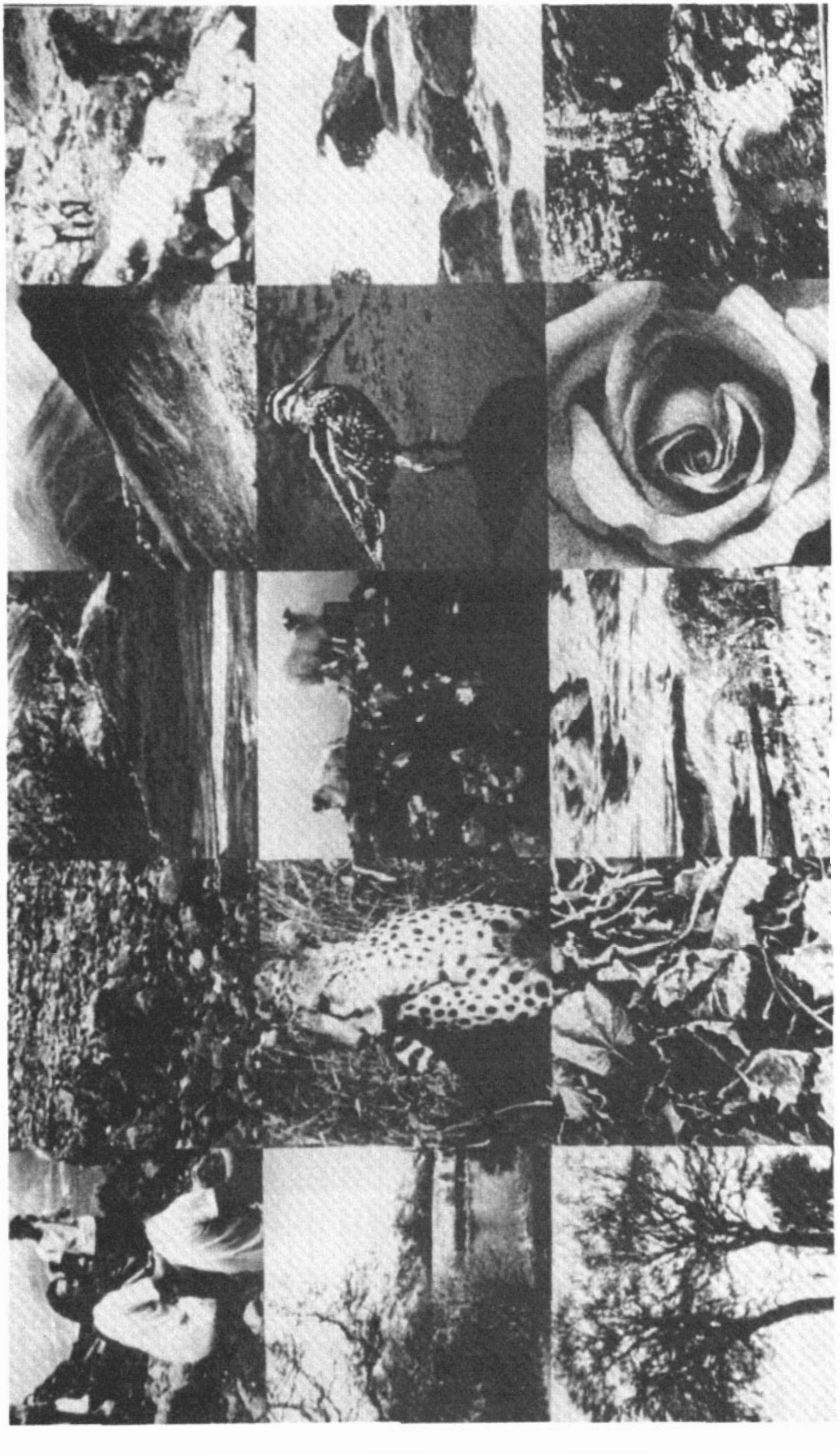
Sanger has applied his algorithm to natural images [1], with the aim of compressing the data. He used 8 × 8 pixel receptive fields and obtained PCs that resemble oriented first- and second-derivative operators. We remove the effects of the square edge of the receptive field by using a Gaussian window. One problem associated with square windows is that the edges may distort the solutions that are found. We are also averaging over a number of different images.

Barrow [2] looked at the possibility of learning the receptive fields of primary visual cortex cells. He modelled retinal and lateral geniculate processing with difference of Gaussian operators and applied portions of a natural image that were pre-processed in this way to a competitive learning network of cells. The cell which responded most strongly to the given input adjusted its connection strengths. The receptive fields that developed also look like oriented first- and second-derivative operators. While this method may resemble what is computed in visual cortex, the competitive algorithm is not guaranteed to give principal components.

Other workers have considered the development of primary visual cortex. Linsker [3,4] looked at the development of a multi-layer system given only random input. He discovered that orientation-selective units developed, despite the lack of any features in the input. The work suggests a mechanism by which appropriate sensitivities may be developed before birth, requiring only tuning up on exposure to the real world. Linsker's system has been shown [5,6] to converge to the PCs of the input under some parameter regimes. Rubner and Schulten [7] use an alternative network model that extracts PCs from correlated noise input by using an anti-Hebbian learning rule to push correlated units apart. They obtained similar operators aligned with the edges of their square input array.

3. Method

Fifteen images were obtained by scanning photographs at a resolution of 300 dpi and 256 grey levels. The pictures (shown in figure 1) were chosen to avoid man-made structures, which tend to have large vertical and horizontal components, and also to avoid obviously straight horizons. The original photographs came from a variety of different cameras and lenses, so no attempt was made to correct for any optical irregularities. Each image was 256 pixels square. Square samples of 64 × 64 pixels were obtained by choosing an image and an area within it at random. The mean grey level (estimated over 20 000 samples) was subtracted from each pixel value. The sample was then masked by (*not* convolved with) a Gaussian with a standard deviation of 10 pixels. This means that the borders of the square sample are more than three standard deviations from the centre, which is far enough to avoid edge effects. The sample vector was then normalized to unit length.



Figu e 1. The 15 natural images used in the experiment.

Sanger's net is a generalization of a single-unit rule proposed by Oja [8]. For a linear unit this is:

$$y = \sum_{i=1}^{N} x_i w_i \qquad \Delta w_i = \eta y(x_i - y w_i).$$

Here, $\boldsymbol{x} = (x_1, \ldots, x_N)$ is a real-valued input, y is the output signal, w_i is the strength of the connection (or weight) from input unit i to the single output unit, and η is the learning rate. This rule can be shown [8,9] to produce a weight vector corresponding to the eigenvector of the correlation matrix of all the inputs which has a maximal eigenvalue; that is, it extracts the principal component of the input data. The weight vector also tends to unit length. Sanger [1] extended the algorithm to multiple output units in a way that extracts the principal components in sequence,

$$\Delta w_{ij} = \eta y_j (x_i - \sum_{k=1}^{j} y_k w_{ik}).$$

The net thus consists of a small number of output units each receiving inputs from, in our case, all 4096 input units. There are no connections between the output units. The net is initialized by setting all the weights to small random values such that the sum of the squares of each unit is approximately unity. We used a rectangular scatter in the range ±0.03. Training consists of applying randomly selected inputs and updating the weights. Convergence is assisted by gradually reducing the learning rate η: we usually started it at 1.0 and then halved the value every 20 000 presentations, for a total of 120 000 presentations.

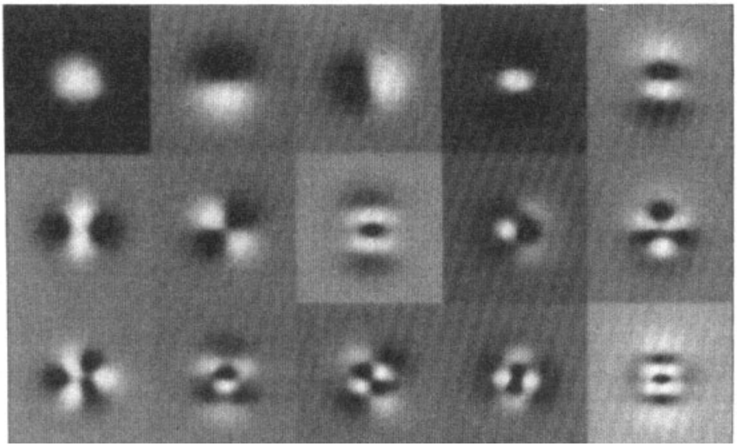


Figure 2. The first 15 principal components of our images, numbered from left to right and top to bottom.

4. Results

The first 15 principal components extracted from our images are shown in figure 2. Note that the sign of each operator shown has no significance: the net may converge such that a given unit has either positive or negative output. The first component is approximately Gaussian. The size is not determined simply by the Gaussian window of our pre-processing, but reflects the correlation scale of the images, as is demonstrated by the use of text images below. The second and third components resemble respectively the horizontal and vertical first-derivative operators, modulated by the Gaussian window. We can show that the orientation of these operators is not caused by residual edge effects by the simple expedient of rotating the photographs in the scanner by 45 degrees. The results of this, shown in figure 4, indicate that the orientation specificity does come from the images. An unexpected result is that the orientation tuning curves of the two 'bar-detector' components (4 and 6) give a good match to a model derived from psychophysical research by Foster and Ward [10]. This finding is explored in more detail in a companion paper [11].

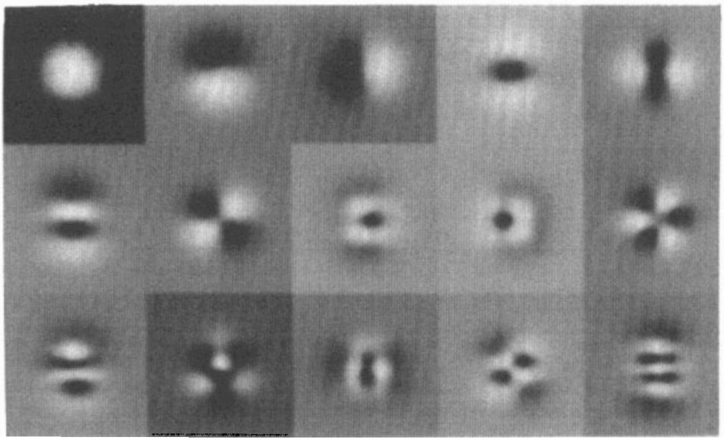


Figure 3. The first 15 principal components of an extended set of 40 images, numbered from left to right and top to bottom (with the permission of the Royal Society).

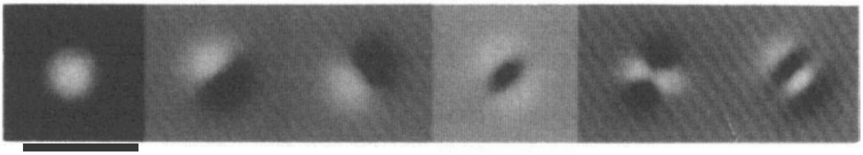


Figure 4. The first 6 principal components of 15 images, when rotated by 45 degrees on the scanner.

We measured the output variance of each component by applying a test set of randomly chosen inputs. The confidence limits on the variances are quite large, so we used 10 000 inputs to generate the results shown in figure 5. The variances decrease with increasing component number. The log against log plot of figure 5 shows an approximately straight line. However, there are some deviations caused by groups of two or three components having quite similar variances. Thus components 5, 6 and 7 are similar. We assessed the consistency of these results by doing another run with 40 images, adding another 10 natural scenes and 15 of man-made structures. Despite the distinctly different input, the first 15 PCs, shown in figure 3, are very similar. Some pairs, such as 5 and 6, have reversed their order, while others, such as 8 and 9, have accounted for similar input variance in a slightly different way. The variance accounted for by these pairs (figure 5) is also similar. The operators that have changed order or become mixed tend to be those which deviate from the generally straight line and have unusually similar variance.

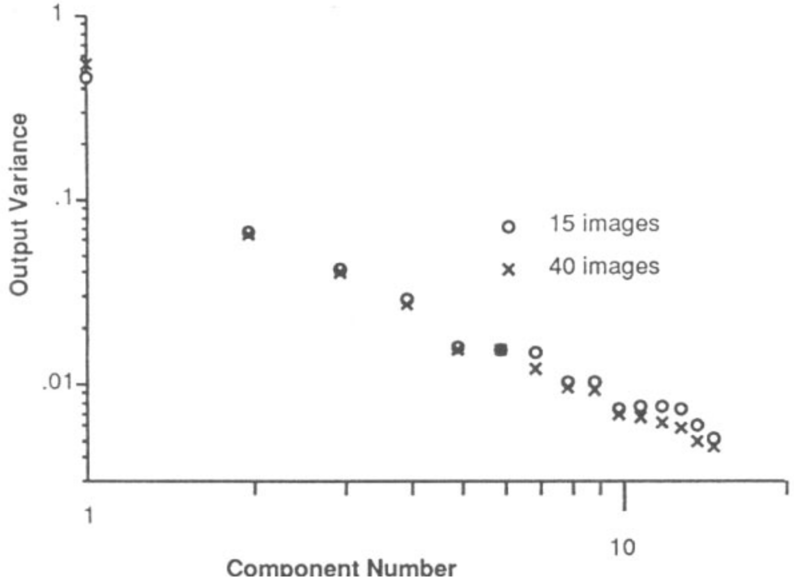


Figure 5. Output variances of the first 15 PCs from the 15 image set and from the 40 image set, over 10 000 inputs.

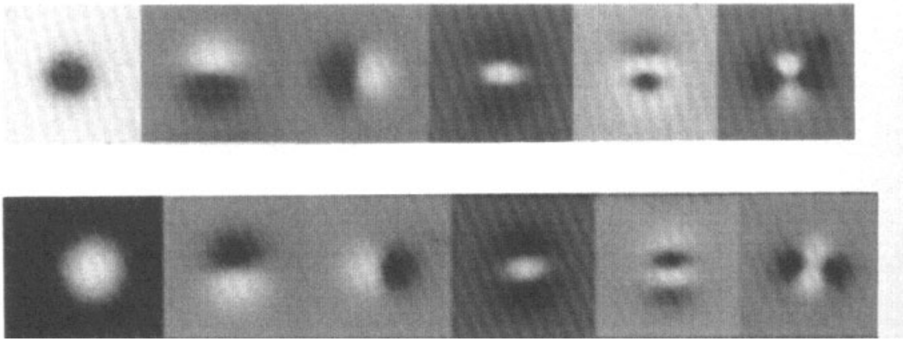


Figure 6. The first six principal components of the images, at double (top) and half (bottom) the original scale.

Because of the variety of our images, there is no reason to suppose that our results would depend on the particular scale that we have used for the analysis. To confirm this, we re-ran the experiment at double and half the original scales. For instance, we used 128 × 128 pixel samples, with a Gaussian window of size 20. The results shown in figure 6 (top) and (bottom) are indeed very similar. By way of contrast, we used the same technique to extract principal components of text at various scales.

We used four different samples of the text at each scale for input. Rather than change the size of our windows, we simply changed the magnification of the text on the scanner. The results (figure 7) now show a very marked scale dependence (note that the sign differences are again not significant). The first and second components are matched in spatial frequency to the inter-line spacing. Subsequent components match the pitch of the letter strokes. Note that, particularly for the finest scale text, the size of the horizontal filters (components 3 and 4) is substantially bigger than the vertical ones (5 and 6). This reflects the use of a proportional font: the lines of text are evenly spaced while horizontal letter positioning varies. The horizontal correlation distance is therefore smaller than the vertical correlation distance. The last two components at the finest scale are tuned to word boundaries. This may be demonstrated by treating the component as a filter and convolving it with one of the original text images. The results of this are shown in figure 8: there is a blob in every word gap, with the biggest blobs making sentence ends. At the coarsest scale, the components are approximately the same size as individual letters.

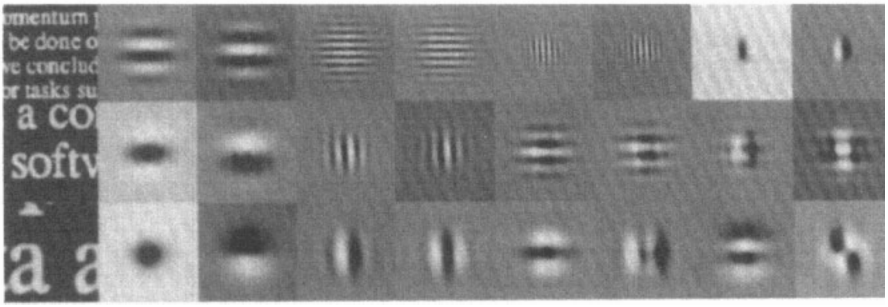


Figure 7. Some samples of Times font text, at three different scales, and their first eight principal components.

5. Discussion

The correlations that PCA detects in natural images arise from the likelihood that adjacent pixels have a similar cause (e.g. the same object). It is apparent that horizontal correlations are on average longer than those in any other direction. It is known [6] that the principal components of symmetrically Gaussian blurred random noise include both oriented and rotationally symmetric operators like those reported here. We have investigated the PCs that result from anisotropic synthetic images: Brownian fractals and Gaussian filtered random noise [11]. The horizontal dominance produces PCs that resemble those from natural images, particularly for the Brownian fractals where the spatial frequency distribution matches that of natural images better.

Our analysis indicates which second-order features are best able to discriminate between our selection of images†. A natural question is whether early visual cortex is performing such a principal components analysis. The general form of the receptive fields of the earlier components certainly resembles those of simple cells, for instance in cats [12]. We are not, however, aware of V1 cells which have some of the more

† Given no information about higher-order statistics.

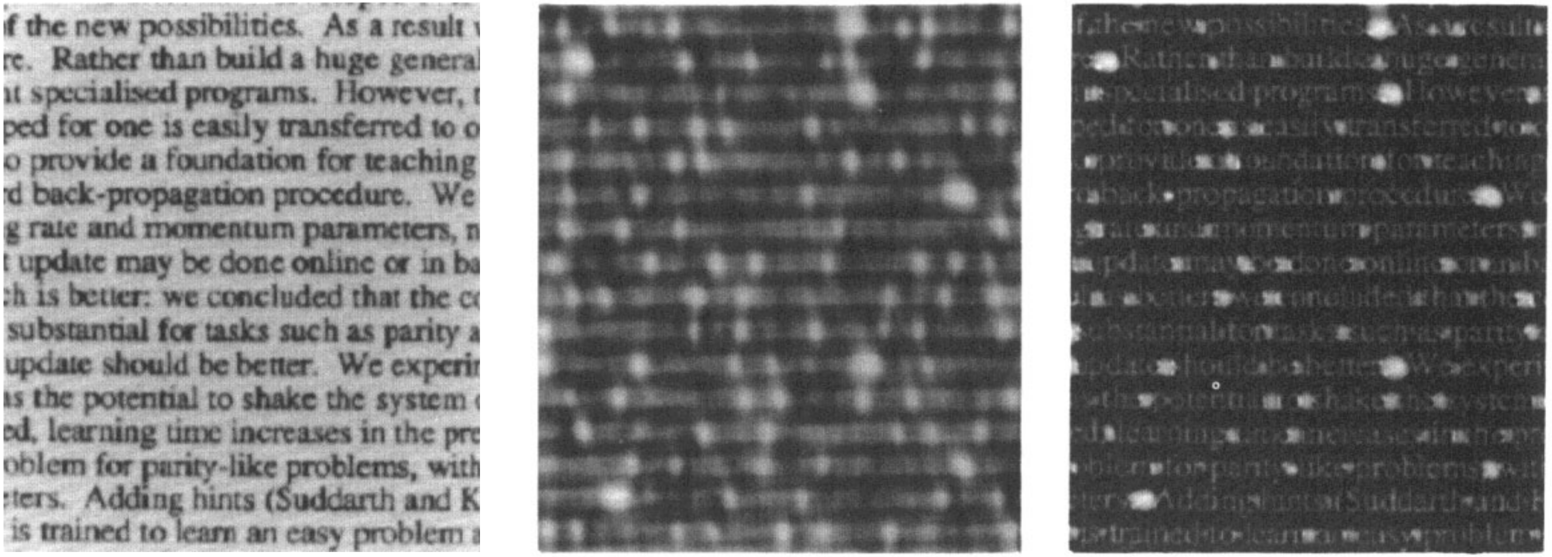


Figure 8. Left: one of the original text images, at fine t scale. Centre: same image, convolved with seve th PC from figure 7. Right: central image thresholded and superimposed on the left-hand image.

complex responses demonstrated by our lower components. While this could be because they have not been tested for (such testing would be difficult), it seems more likely that the different constraints operating in V1 lead it to do something other than simple PCA.

One major difference between the constraints in our system and those operating in cortex is the (relative) absence of noise in our model. Principal components will convey the most information provided the signals are accurate. Noise may have two effects. One is that an alternative coding, that spans the same space as the PCs, becomes preferable. The outputs can then be adjusted to have similar variances, so that each cell carries a similar amount of information. The second effect is that, in the presence of noise, minimizing mutual information no longer implies mutual orthgonality. Passing on the maximum information requires only that knowledge of one cell output gives no information about the output of another one. In the short processing times required for vision, apparently overlapping receptive fields may be effectively independent.

Another difference in constraints is that the brain is able to process the whole image at once. We are looking at second-order statistics, whereas the brain can use correlations between the outputs of second-order filters that look at adjacent bits of the image. Thus output from two appropriately oriented 'bar-detectors' may indicate a line in the image. It is not obvious how useful some of our lower- order components would be in predicting the activity of their neighbours. While economy suggests that mutual information between cells looking at the same part of the image should be minimized, representations useful for further image processing may be produced by maximizing mutual information between cells looking at adjacent parts of the image. The development of interesting unit properties such as depth sensitivity by maximizing mutual information has been explored by Becker and Hinton [13].

Acknowledgments

PH is funded by grant no GR/F97393 from the UK Science and Engineering Research Council Image Interpretation Initiative; RB is a SERC student. The work was performed on equipment funded by SERC and by the EC BRAIN initiative, grant no ST2J-0415-C. We have benefitted from discussions with Roger Watt.

References

[1] Sanger T D 1989 Optimal unsupervised learning in a single-layer linear feedforward neural network *Neural Networks* **2** 459–73

[2] Barrow H G 1987 Learning receptive fields *IEEE Int. Conf. on Neural Networks (San Diego 1987)* (New York: IEEE) p 115–21

[3] Linsker R 1986 From basic network principles to neural architecture *Proc. Natl Acad. Sci. USA* **83** 7508–12, 8390–4, 8779–83

[4] Linsker R 1988 Self-organization in a perceptual network *Computer* 105–17

[5] MacKay D J C and Miller K D 1990 Analysis of Linsker's simulation of Hebbian rules *Neural Computation* **2** 173–87

[6] MacKay D J C and Miller K D 1990 Analysis of Linsker's simulations of Hebbian rules to linear networks *Network* **1** 257–97

[7] Rubner J and Schulten K 1990 Development of feature detectors by self-organization *Biol. Cybernetics* **62** 193–9

[8] Oja E 1982 A simplified neuron model as a principal component analyzer *J. Math. Biol.* **15** 267–73

[9] Hertz J, Krogh A and Palmer R G 1990 *Introduction to the Theory of Neural Computation (Santa Fe Institute Studies in the Sciences of Complexity)* (Reading, MA: Addison-Wesley)

[10] Foster D H and Ward P A 1991 Asymmetries in oriented-line detection indicate two orthogonal filters in early vision *Proc. R. Soc.* B **243** 75–81

[11] Baddeley R J and Hancock P J B 1991 A statistical analysis of natural images predicts psychophysically derived orientation tuning curves *Proc. R. Soc.* B submitted

[12] Orban G A 1984 *Neuronal Operations in the Visual Cortex* (Berlin: Springer)

[13] Becker S and Hinton G E 1989 Spatial coherence as an internal teacher for a neural network *Technical Report* TR CRG-TR-89-7, University of Toronto, Department of Computer Science