Image Recognition in Hypercolumnar Scale Space by Sparsely Coded Associative Memory* †

H. Janßen, J. Kopecz

Institut für Neuroinformatik, Ruhr–Universität Bochum 44780 Bochum, Germany
e–mail: heja@neuroinformatik.ruhr-uni-bochum.de

We propose a pattern recognition system based on an architecture close to the one found in human visual cortex which is called hypercolumns. We show that this discrete parametric representation can be used to define a short range interaction to make desirable information like edge continuation more explicit. We also show how hypercolumnar representations can be sparsely coded for usage in a very efficient associative memory recognition system. We combine this system with a model for coarse–to–fine search in hypercolumnar scale space thus gaining translational invariance. In principle the application of such a representation appears to be very well suited for data reduction and pattern recognition processes and is part of a *neural instruction set*.

1 Introduction

The structural principles of the nervous system can be viewed at as having emerged in an optimal way from evolutionary adaption to a certain kind of information processing. The most well known examples of such structural as well as processing principles are layered cell structure, Retinotopic Maps, Feature Maps, Associative Memory or Active Vision. In dealing with these principles we are concerned with a *neural instruction set* [5] which will determine alternative ways of information processing compared with classical computers.

A well known example is given by the hypercolumns [2] as found in visual cortex, where orientation selective cells are organized in a way as to form a discrete parametric representation in which the parameter "orientation" is mapped into location. The models presented here will demonstrate that such a structure is extremely well suited for a massive data reduction and offers itself in a natural way for recognition of real world images and scenes. Beyond that, we show with the example of translational invariance, how a search strategy in a hypercolumnar scale space can be used to easily achieve invariance performances for recognition tasks. In our opinion discrete parametric representations are a very flexible and widely applicable structure for task– and knowlegde based image processing and therefore indeed constitute a fundamental element of neural information processing. The practical aptitude of the model is demonstrated by some examplified results and can further be seen in an active vision object recognition system described elsewhere [1].

2 A Dynamic Approach to Hypercolumnar Interactions

The coordinates $i \leq I$, $j \leq J \in \mathbb{N}$ may describe a 2D–intervall in the image space and will later result in the cortical position of a hypercolumn (HC). $u \leq U$ and $v \leq V$ are the coordinates *within* a HC. A certain point on the cortex plane is thus described by a set of four coordinates. A discrete parametric mapping of an image intervall $[i, j]$ onto a HC is now defined by the operation:

*Edited version, originally published in: T. Kohonen et al. (ed.), Artificial Neural Networks, Proc. of ICANN 91, pp. 1203 – 1206, Elsevier Science (1991)

†Supported by the German Federal Department of Research and Technology (BMFT), Grant No. ITR8800K4

$$w_{u,v}^{i,j} = \sum_{\nu,\mu} \left[\frac{f^{i,j}}{\|f^{i,j}\|} * g_{u,v} \right]^{\nu,\mu} \tag{1}$$

$w_{u,v}^{i,j} : \mathbb{R}^2 \rightarrow \mathbb{R}$	cell' s activity at pos. u, v in HC i, j
$\nu, \mu \in \mathbb{N}$	pixelcoordinates of convolved intervall i, j
$f^{i,j}/\|f^{i,j}\|$	normalized greyvalue function of intervall i, j
$g_{u,v} : \mathbb{R}^2 \rightarrow \mathbb{R}$	set of modified gabor functions with $(K = U \times V)$ orientations

Because of its physiological structure we assume that a prominent part of early visual information processing in the brain is done by local interactions. One can then easily formulate constraints and characteristics of such dynamical system by means of differential equations having attractors at "places of desired behaviour". In our approach, three assumptions enter the formal description of the system:

Firstly, we introduce nearest neighbour crossinhibition that is, cells being responsive preferently to a stimulus of a certain orientation tend to inhibit adjacent cells' activities that code approximately perpendicular orientations. This is described as a weighted sum over the adjacent Hypercolumns, being asymmetric in two aspects: HCs perpendicular to a cell's orientation have stronger inhibitory weightings than others and the cells' couple asymmetric since the nearest neighbours enter as a positive definite function with a higher power than the actual activity itself. In the following, we will denote a cell's activity by w_{ijk}, where k codes the orientation. A damping term

$$U_c = [\sum_{nm} \sum_{l<K} w_{i^n j^m l}^2 w_{ijk} \times \mathcal{T}_{kl}^{nm}]$$

ensures that higher activity is less inhibited. The term $\mathcal{T}$ codes the spatial anisotropic and crossinhibitory interactions.

Secondly, a cell can build up activity if the activation distribution in its vicinity is of proper shape although it has not been stimulated by its related receptive field. This relates to e.g. optical illusions that tend to close and smooth interrupted or faulty lines. Such an excitatory contribution is written as $U_{ex} = +v(t)$ where $v(t)$ is coupled to $w_{ijk}(t)$ by a monotone decreasing function that prevents the system from unconstrained growth.

At third, the system receives visual input that enters as a force: $U_{st} = -[w_{ijk}(t) - w_{ijk}(0)]$.

So the we end up with a set of coupled nonlinear differential equations of the following form:

$$\frac{dw_{ijk}(t)}{dt} = \alpha U_{st} + \beta U_c + \gamma U_{ex} \tag{2}$$

By construction this system always converges. The solution of eq.(2) together with a projection back into image space can be regarded as a regularization. Fig.1 shows the effect on a greyvalue image.

3 Recognition by Sparsely Coded Associative Memory

The high data reduction rate, achieved by hypercolumnar representations and the potential of regularisation brings into question, how much and which relevant information of the original image is preserved. A pattern recognition system on the base of a hypercolumnar mapping gives the possibility to examine this question further. Because of the data reduction it is possible to implement a system with high speed and low memory consumption. The goals for our pattern recognition system were robustness of algorithm and a reliable error measure. Some invariance to translation and scaling is intended too, however the invariance problem is defered to the scale space extension described later.

Let $\mathbf{x}$ with $\mathbf{x} \in \mathbb{R}^m$, $m = IJK$ be a discrete parametric representation. We use the transformation

$$\mathbf{x} = (x)_M, \;\; x_M = w_{i,j,k} \;\text{ with }\; M = k + iK + jIK, \tag{3}$$

i.e. the 2–dimensional discrete distribution of excitation is transformed to a 1–dimensional vector, destroying the topographical neighbourhood and preserving only neighbourhood of orientations.

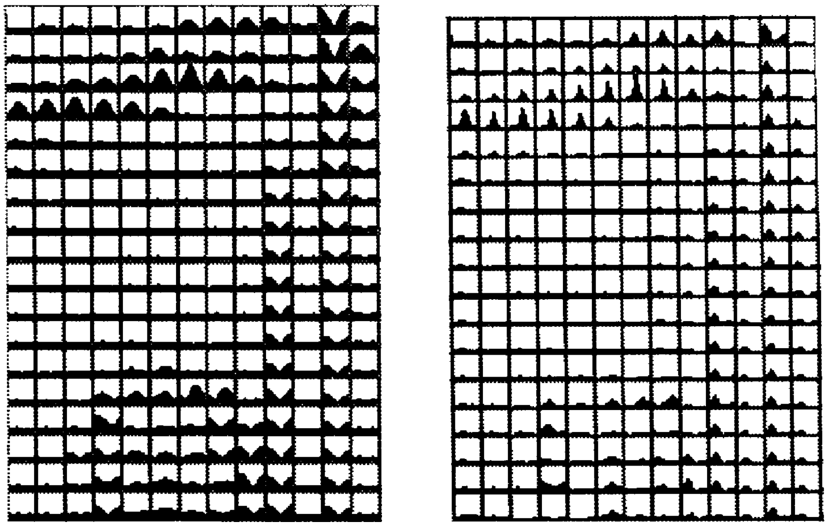

Figure 1: *Effect of local interaction as described by eq.(2) on an ensemble of hypercolumnar cells. On the x axis of each little histogramm are the orientations, the y axis represents the excitation of a given cell. On the left is the original representation of an example image, on the right the enhanced representation with sharpened peaks and suppressed noise.*

This vector $\mathbf{x}$ now is coded to a binary vector

$$\mathbf{p} = \hat{\mathbf{C}}\mathbf{x} \ \text{ with } \ \mathbf{p} \in \mathbb{B}^m. \tag{4}$$

The coding operator $\hat{\mathbf{C}}$ will be discussed in more detail later. Actually it is only required, that

$$\sum_{i=1}^{m} p_i = l \ \text{ with } \ 0 < l < m \quad \text{for all } \mathbf{p}. \tag{5}$$

holds. The transformed and coded pattern vectors $\mathbf{p}$ therefore are binary vectors with a constant number of 1–elements. These patterns are stored in an associative memory. The model used here is a classifying binary associative memory similar to the one proposed by Willshaw et al. [8] and mathematically analyzed by Palm [6].

The memory itself is a matrix $\mathbf{S} = (s)_{i,j} \in \mathbb{B}^{m \times n}$ within which the patterns are stored as column vectors. In case of an inquiry, an arbitrary input pattern $\mathbf{p}^k$ from a pattern set $\tilde{K}$ is multiplied by the memory matrix:

$$\mathbf{q}^{k^T} = \mathbf{S}\mathbf{p}^k \quad \text{with } \mathbf{q}^k \in \mathbb{N}^n, \, \mathbf{p}^k \in \{0,1\}^m, \, k \in \mathbb{N}. \tag{6}$$

One element q_i^k with $i \in \{1, \ldots, n\}$ contains the correlation of the input vector $\mathbf{p}_k$ with the stored vector $\mathbf{p}_i$, and is therefore a measure for similarity of both binary patterns.

A significant increase of speed can be achieved by a sparse coding of the pattern vectors, i.e. $u \ll m$. In this case, it is not necessary to search the whole matrix for multiplication by the input vector, but only the subset of line vectors, which corresponds to 1–elements in the input vector.

Now we have to discuss in more detail, which coding is suitable to reflect the meaning of real world similarity to the correlation values of the binary patterns. A coding which is in this sense *meaningfull* can only be obtained heuristically, because of the necessarily inexact definition of the term *similarity*. The excitation distribution in a hypercolumn usually is highly redundant: The main information of such a histogram is which orientation is dominating (position of the maximum) and the strength of the edge (amplitude of excitation distribution). Only these two parameters shall be coarsely coded now, thus gaining a high correlation of patterns which are similar in this way.

The various methods of coding are described in previous papers [4, 3] and will not be presented here. The main idea is to code the maximum position in a hypercolumnar histogram plus its nearest neighbours by a 1–element leaving the other positions zero. Improvements use random coding and adaption of l to the "mean edge contents" of an image series to overcome the problem of the fixed value of l for every pattern.

4 Scale Space Searching for Translational Invariance

As an extension to the simple pattern recognition system mentioned above, we propose a model to gain translational invariance using a sequential coarse–to–fine search in a hypercolumnar scale space. The basic approach of this model can be applied in a very similar way to the problem of scale invariance.

The model uses the small translational invariance of the simple recognition system, which results from the local averaging property of the used gabor filtering representing image edges with a mean positional invariance of 1/2 of the segment width. Therefore the invariance can be increased by increasing the segment size, thus reducing the recognition abilities. A combination of large invariance and good recognition can be gained by the scale space matching model.

The image is not only coded in the initial level of resolution but in different lower resolutions in octave distance [7]. By reducing the resolution and size of the images, leaving the filtering alike, we get a pattern pyramid with large translational invariance at the top, and recognition precision at the bottom.

This pattern pyramid is stored in the memory as one block thus magnifying the storage consumption to a factor of $\sum_1^\infty 1/n^2$ at maximum. In the recognition phase, the image is coded and associated to the memory resulting in one correlation histogram for every step of resolution. These histograms minus an inhibitory histogram are summed up giving a sumhistogram which is used as an overall similarity measure. The inhibitory histogram is initialized with the sum of the individual histograms for those patterns whith correlation lower than a minimum threshold. At this point the following algorithm starts:

- If the maximum correlation value of the sumhistogram is satisfactory to some appropriate condition, the process stops giving the maximum position as result.
- If there are positive values in the sumhistogram, the maximum is chosen and a shift process is started, searching from the coarsest to the finest resolution. If there is no positive value, the recognition is stopped with a "pattern not found" result.
- If the value in the sumhistogram can be raised enough corresponding to step one, the whole process is stopped, otherwise the pattern position in the inhibitory histogram is set to the sum of the individual histograms.

Thus the inhibitory histogram is used to overcome the problem of doubly searching for a pattern by storing the falsificated pattern.

The algorithm outlined here can be understood as a search in a classifying tree with erroneous data. In contrast to the only error free solution (full search), the algorithm reduces search time significantly, tolerating some errors in the result. A systematic statistical analysis of the algorithm is left to further research, but testing with some hundered images showed promising results.

5 Implementation and Results

The simple associative memory was tested extensively with different forms of coding with and without regularisation. Some results are given in previous papers [3, 4]. The scale space approach with sequential search, mentioned here in the following benefits very much from the speed of the simple recognition system[1]

The real implementation of the scale search process uses a central filtered representation of the input image thus needing only one (time consuming) filtering operation per association. A result in case of a relatively difficult recognition task is given in Fig. 2.

[1] With a pattern vector length of 3136 bit and a memory size of 3136 patterns (1.2 MBytes) the recognition time on a SUN–4 is about 1/10 second.

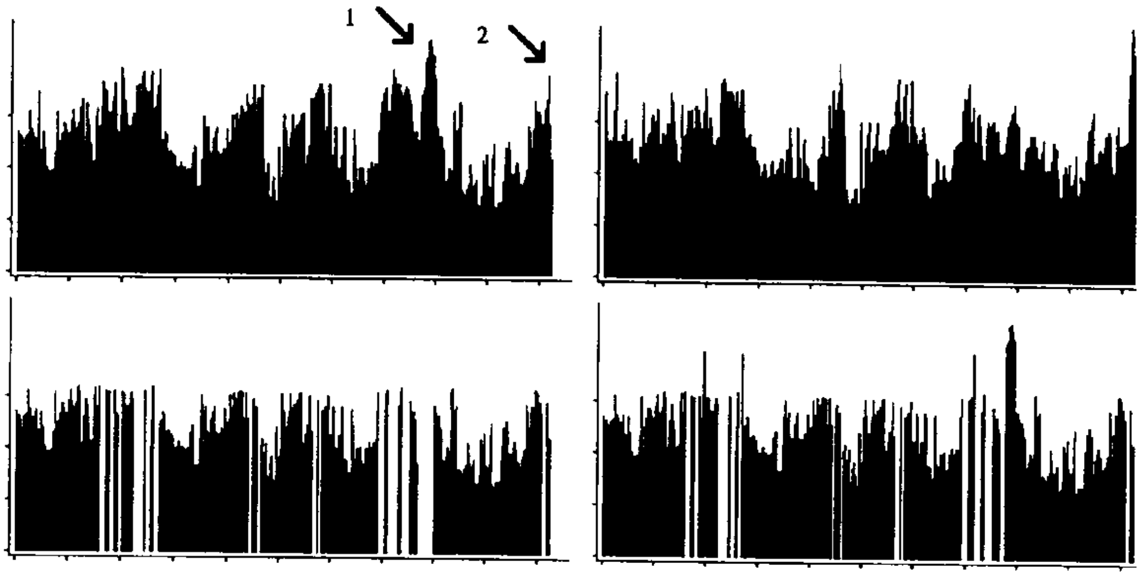

Figure 2: *A typical result of the search process in scale space: the left pictures show the sumhistogram and the inhibitory histogram at the start of the search process, the right ones at its successfull end. All histograms display correlation against pattern number. The memory contained 512 patterns of 128 square pixel grey valued camera images with laboratory scenes. The input image (corresponding to stored images at arrow 2) was shifted by 15 pixels down and 13 left and associated. Note that the unshifted image was nearly satisfactory matched to some "false" patterns (see arrow 1), but the matching procedure inhibits these solutions in the search process and took the right pattern.*

References

[1] G.-J. Giefing, H. Janßen, and H. A. Mallot. A saccadic camera movement system for object recognition. this volume, 1991.

[2] D. Hubel and Wiesel. Functional architecture of macaque visual cortex. *Proc. R. Soc. London B*, 198:1 – 59, 1977.

[3] H. Janßen and J. Kopecz. Image representation in hypercolumnar structure my means of associative memory. In *Close–Range Photogrammetry meets Machine Vision*, pages 1050 – 1057. ISPRS, SPIE, 1990.

[4] H. Janßen, J. Kopecz, and H. A. Mallot. Erkennung natürlicher Bilder mit Hilfe diskreter parametrischer Bildrepräsentationen und Assoziativspeichern. In *Mustererkennung 1990, Informatik Fachberichte 254*, pages 232 – 233. DAGM, Springer-Verlag, 1990.

[5] H. A. Mallot, W. von Seelen, and F. Giannakopoulos. Neural mapping and space-variant image processing. *Neural Networks*, 3:245 – 263, 1990.

[6] G. Palm. On associative memory. *Biol. Cybern. 36, 19-31*, 1987.

[7] M. Porat and Y. Y. Zeevi. The generalized gabor scheme of image representation in biological and machine vision. *IEEE PAMI*, 10:452 – 468, 1988.

[8] D. J. Willshaw, O. P. Buneman, and H. C. Longuet-Higgins. Non-holographic associative memory. *Nature*, Vol. 222:960–962, 1969.