Image Representation in Hypercolumnar Structure by means of Associative Memory[*†]

H. Janßen, J. Kopecz, H. Mallot

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

Abstract

We propose a pattern recognition system based on an architecture close to the one found in human visual cortex which is called Hypercolumns. We use this discrete parametric representation in connection with a sparsley coding associative memory. 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* [2].

1 Introduction

Besides using properties of single cells and simple neural nets the nervous system exploits structural principles for visual information processing. These principles 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* [2] which will determine alternative ways of information processing compared with classical computers.

A special example is given by the hypercolumns [1], 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. Thus, on a coarse grain topographic neighbourhoods of the retinal picture are conserved, whereby different orientations are represented on a finer grain for each area in the picture respectively.

The paper presented here will demonstrate with the example of pattern recognition at hand, 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. 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.

2 Theoretical principles of discrete mapping and associative image recognition

2.1 Building a discrete parametric representation

In the following paragraphs a brief and formal description of a discrete parametric representation will be given, together with a short discussion of possible practical implications.

The coordinates $i \leq I$, $j \leq J \in \mathbb{N}$ may describe a 2 D intervall in the image space and will later result in the cortical position of a hypercolumn (HC).

*Edited version, originally published in: A. Gruen, E. Baltsavias (ed.), Close–Range Photogrammetry meets Machine Vision, Proc. of ISPRS 90, pp. 1050 – 1057, SPIE Bellingham (1990)

†Supported by the Bundesministerium für Forschung und Technologie, Grant No. ITR8800K4

$u \leq U$ 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

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

where

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

so the hypercolumnar mapping can be seen as the application of an ensemble of $(I + U) \times (J + V)$ filters to a certain image. Fig.2 visualizes such a mapping with a natural picture taken by a video camera. The amount of data reduction, as referred in Fig. 1 as 20, is thus dependent on the following factors:

1. the size of each picture intervall, the larger the intervalls the larger the reduction factor (loosing more information at the same time)
2. the number of orientationally selective cells
3. the amount of overlap between the different intervalls

One advantage of this kind of discrete representation e.g. compared to a Hough Transform, is the twofold information coded in the stimulus position which means position plus orientation of a feature. The position in the HC codes orientation and position in a way that similar features in neighbouring retinal positions correspond to similar stimuli in neighbouring Hypercolumns – a fact that is of high importance for an efficient coding in an Associative Memory. As already shown by [3] a total steadiness cannot be achieved since a 3 dimensional structure is mapped onto a 2 dimensional stucture.

By means of this formal description of discrete mapping it is now possible to define processing operations on a HC structure which may e.g. correspond to a regularization of a vector field.

2.2 Defining processing operations on a hypercolumnar structure

By means of local evaluation of neighbouring stimuli and simple geometric manipulations coherent features like lines or edges can be enhanced thus filling blanks between lines of the same orientation, an effect well known in the human visual system.

Regularization in this context means assigning a certain set of (orientation–) parameters to each original picture segment respectively dependent firstly on local relationships between adjacent Hypercolumns and secondly on external requirements (frequently referred to as "knowledge").

Let us consider $\mathbf{h}^{i,j}$ as a vector describing all cells in a HC i, j. Then a given regularizing step can be written as:

$$\tilde{\mathbf{h}}^{i,j} = \mathbf{h}^{i,j} + \hat{W}_U \mathbf{h}^{U(i,j)} \tag{2}$$

where $U(i, j)$ means adjacent Hypercolumns, e.g. the nearest neighbours. $\hat{W}$ can be seen as a weighting matrix depending on local relations whose elements are a function of each regularization step and which has two important features:

1. For each cell firing after presentation of an edge $\hat{W}$ gives a connection stength to all cells of the adjacent Hypercolumns lying in the direction the firing cell "points" to. Thus geometry and strength of the connections depend upon the structure of the picture presented to the HC array.
2. besides this, a phase difference can be assigned to cells in adjacent HCs. By defining a constraint function this results in inhibition of orthogonal cells and enhancement of similar oriented cells, as shown in Fig. 1.a.

This constraint function can be modelled by the Associative Memory and therefore induces the element of context knowledge to the recognition process as will be described on page 3.

This type of interaction can be modelled mathematically as a dynamic system which has certain patterns as attractors, however we will not go into this in this paper.

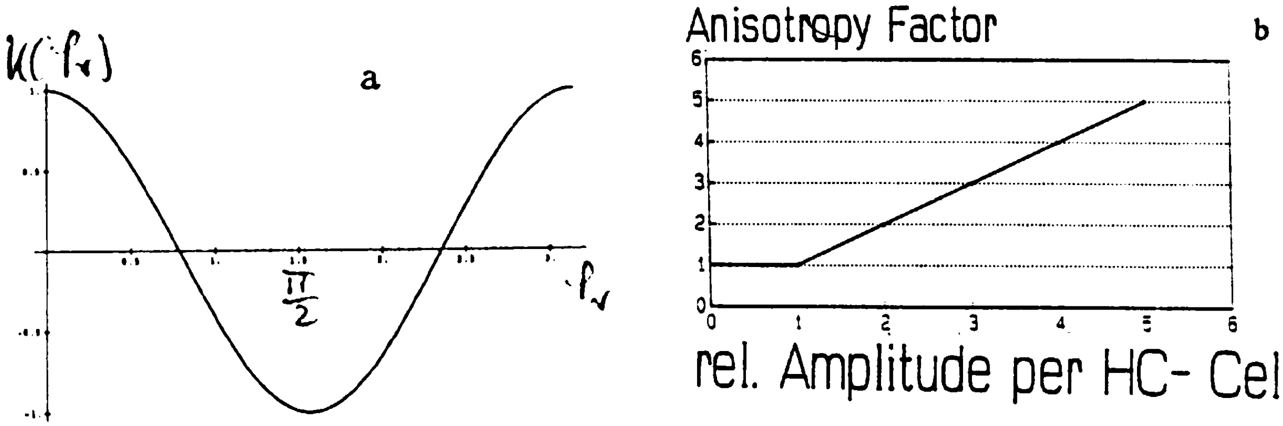

Figure 1: *a. Constraint function $k(\varphi_r)$, which models the relation between weighting factors and relative phases of HC cells. Here, the most steady continuatio nof lines is modelled, signals having a relative phase of $\frac{\pi}{2}$ are inhibited. b.: Anisotropy factor in the connections as a function of cell activity.*

3 System implementation, application to natural scene recognition

3.1 Discrete mapping and regularization

We have realized edge detection by convolution of picture intervalls with a set of Gabor functions as well as by performing a scalar multiplication with a set of orientation selective filters. In any case, the result is a set of numbers each representing the quality of an edge in a given picture intervall, which may as well be interpreted as neural activity of a cortical cell.

Thus amplitude and position code a specific edge at a given position in the original picture. For each Hypercolumn the result of Eq. (1) can be visualized in a histogramm of the amplitudes, as shown in Fig. (3a). Projecting back from HC space to image space yields pictures similar to Fig. 3b in which obviously by far not the whole information of the original is contained. However, it carries with it the most characteristic information necessary to perform a recognition process of high quality.

3.2 Image representation in sparsely coded associative memory

The high data reduction rate, achieved by a hypercolumnar representation brings into question, how much and which (behavioral) relevant information of the original image is preserved, and how this information can be used in a more complex system. A pattern recognition system on the base of a hypercolumnar mapping – like the one proposed here – gives the possibility to examine this question further. Because of the amount of data reduction, it is possible to implement such a task with practicable speed and memory consumption. Other goals for the pattern recognition system briefly sketched here are robustness of algorithm and a reliable error measure. Some invariance to translation and scaling is intended too, however the task of invariance is not pursued here, since it is mainly adressed to other parts of the system.

Let $\mathbf{x}$ with $\mathbf{x} \in \mathbb{R}^m$, $m = I * J * K$ be a discrete parametric representation. In case of the hypercolumnar

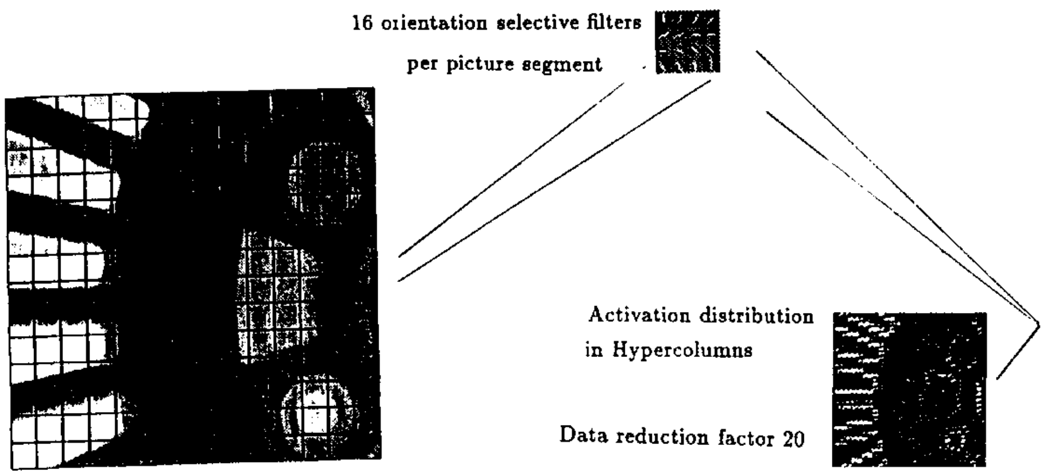

Figure 2: *Application of mapping process to a video image representing a section of a wheel. The picture is segmented into intervalls of a certain overlap and then analysed by means of orientational selective filters. Afterwards, each pixel in the hypercolumn represents orientation and position of an egde. Thus a considerable amount of data reduction can be achieved.*

representation as it is dicussed above, we get this by the transformation[1]

$$\mathbf{x} = (x)_M, \;\; x_M = h_k^{i,j} \;\; \text{with} \;\; M = k + i * K + j * I * K, \tag{3}$$

i.e. the three dimensional discrete distribution of excitation is transformed to a onedimensional vector, destroying the topographical neighbourhood and preserving only neighbourhood of orientations. This vector $\mathbf{x}$ now is coded to a binary vector[2]

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

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

$$\sum_{i=1}^{m} p_i = w \;\; \text{mit} \;\; 0 \leq w \leq m \quad \text{for all } \mathbf{p}. \tag{5}$$

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 sparsely coded associative memory similar to the ones proposed by Palm [4, 5] and Willshaw et al. [8].

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. Therefore maximal n patterns can be stored. In case of an inquiry, an arbitrary input pattern $\mathbf{p}^k$, from a pattern set K is multiplied with 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 therefore a measure for similarity of both binary patterns.

Besides simplicity this type of associative memory used here offers the advantage of effective implementations on conventional computer architectures and an interesting mathematical property: Because every pattern vector

[1] The indices u, v are combined to one index k for the orientation in one hypercolumn.

[2] In the following, we speak of binary vectors as of vectors, which are elements of vector spaces over $\mathbb{R}$, and whose elements are in a subset of $\mathbb{R}$, namely in $\mathbb{B} = \{0, 1\}$. The same holds for matrices.

b

a

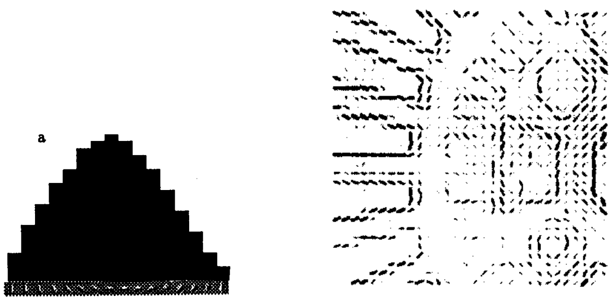

Figure 3: *a: Histogramm of activity distribution in a hypercolumn, here, the response of all HC cells to a horizontal egde is shown. The broad tuning of the cells is also found in biological systems and is improved by the regularization process.*
b: Reconstruction of an original picture by projecting back into image space.

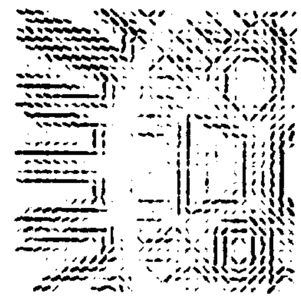

Figure 4: *Reconstruction similar to Fig. 3b after some regularization steps. Straight lines are enhanced, smaller structures supressed.*

p w contains the same number of 1-elements, the matrix as a whole contains $W = wn$ 1-elements. The number of 1-elements in every line vector is j

$$z_j = \sum_{i=1}^{n} s_{ij} \tag{7}$$

which is a measure for the quality of a feature. If

$$z_j \simeq W/m \quad \text{for all } j \leq m, \tag{8}$$

holds, the same amount of information is contributed to the pattern classification by all features, while large variations in the line sums z_j point to a collection of features, which is inadequate in respect to the used pattern set.

In a sequential implementation, 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 with 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 give 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*. As can be seen in Fig. 5a, 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 theexcitation distribution). Only these two parameters shall be coarsely coded now, thus gaining a high correlation of patterns which are similar in this way.

The first possibility is to code the position of the excitation maximum in every hypercolumn in a bit pattern with a 1 for the maximum of orientation and 0 otherwise. To come up to some invariance to rotation of edges, the neighbours of the maximum can too be coded by 1. In the same way we can attain a coarse coding of the strength of the edge: A high excitation amplitude is coded by more 1-elements than a less sharp peak in a histogram.

One problem concerning this first attempt is the differing number of ones used to code the hypercolumns. Because of the estimation of feature qualities (8) and the advantage to get the absolute correlation values as a measure for similarity without normalization, (5) has to hold. This can be achieved by a random distibution of the *surplus* of 1-elements, or expressed as a condition of constraint:

$$\forall_{0 \leq r \leq \frac{m}{h-1}} \sum_{s=1}^{K} p_{s+r*K} = w/K, \tag{9}$$

with K as the number of orientations in a hypercolumn. Thus this is a very strong restriction to condition (5). This procedure will be referred to as *local coding* further on.

But local coding too fails to solve the problem of large differences in the excitation distribution of real world images. In principal, this can only be obtained by a nonlocal coding, i.e. by a coding, that does not only depend on single hypercolumns and holds the number of 1-elements constant here, but does the same thing globally, i.e. for the whole pattern vector. This can be achieved under the assumption of a constant number of *relevant* features in every image and requiring a simple relation to calculate the *relevance value*. A reasonable value again is the amplitude of the excitation histogram of a hypercolumn.

Formally we can write this as:

$$C(h_k^{ij}) = G(M(h_k)^{ij}, \mathbf{r}) \qquad \text{with } M : \mathbb{R}^K \mapsto \{1,\dots,K\} \times \mathbb{R} \quad \text{and } G : \{1,\dots,K\}^{IJ} \times \mathbb{R}^{IJ} \mapsto \mathbb{B}^{IJK}. \tag{10}$$

By application of $\hat{\mathbf{M}}(u, v, \mathbf{t})$ we get the maximum orientation position and the excitation amplitude for every hypercolumn.

Application of $\hat{\mathbf{G}}(i, j, \mathbf{r})$ results in the code vector from these positions and amplitudes.

Therein $\mathbf{r}$ is the parameter set, by which $\hat{\mathbf{G}}$ can be modified in wide borders: fuzzyness of coding, number of HC's to be coded by ones and eventually various levels of coding strength. An example for such a *global coding* will be given in the discussion of the implementation.

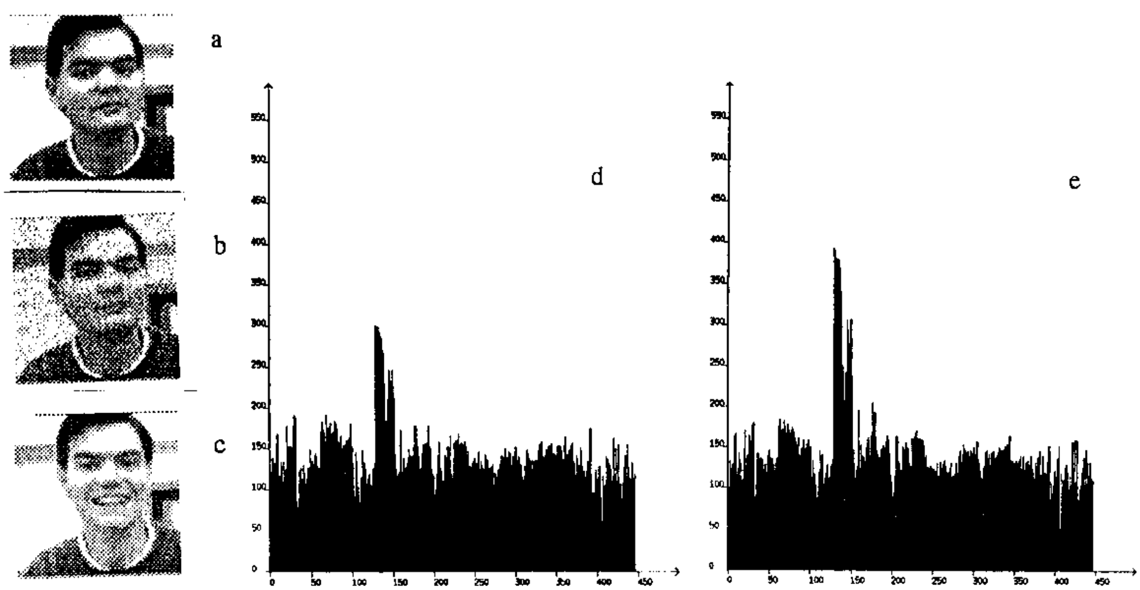

Figure 5: *Some results: correlation of different natural video images to the slightly noised image 128. The images 128 to 150 are a sequence of very similar images. a) image 128 a) image 128 slightly disturbed (gaussian noise with $\sigma = 10\%$) c) image 150 d) histogram in case of local coding; the maximum correlation value is 588, theoretical number of different codings is $16^{196} \approx 1 \cdot 10^{236}$ e) histogram in case of global coding; 2/3 of the image segments are coded, theoretical number of different codings is $16^{130} \cdot \binom{196}{66} \approx 5 \cdot 10^{209}$*

3.3 Implementation of the pattern recognition system

The associative memory was implemented[3] and tested with a variety of codings. The two most important types of coding, namely the simple local coding and the global sparse coding were tested by storing about 450 natural

[3] In all examples mentioned here, the length of the pattern vectors is 3136 bit, with a memory size of 3136 patterns (1.2 MBytes) the recognition time on a sun-4 is about 1/10 second.

video images (mostly scenes in laboratory and portraits) and 2000 random images (white noise)[4]. In selecting scenes, we paid carefull attention to get both sequences of very similar and very different images, to check the generalization capabilities of the system.

The results show that even the simple local coding yields very good classification properties in recognition of nondistorted images. With this form of coding, random images are coded as maximal different, because the noise in the greyvalues generates weak edges in random directions and therefore the binary patterns contain random distributed ones (under constraint of condition 9). That this is a disadvantage is shown by checking the system with a set of patterns got from a natural image by random disturbance of different intensity and by cutting off parts of the image plane (setting the grey value to zero). While cutting off parts of the image (especially in case of more or less uniform background parts) does not affect the recognition very much, *noising* of uniform image parts prevents acceptable results.

In contrast to this the global coding improves the recognition process by coding images more sparsely (Fig.5). But because of the strong dependance on the image contents, it is reasonable to combine this technique with the regularization process as a preprocessing procedure step in future (Fig.6).

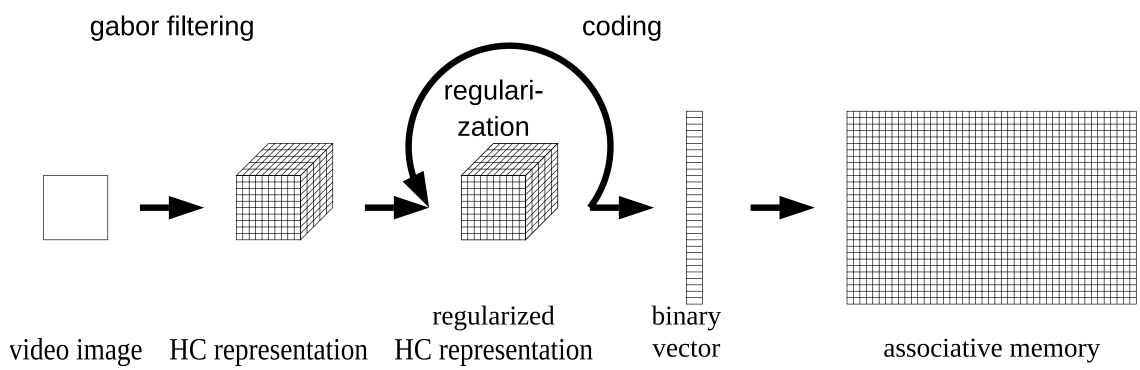

Figure 6: *The whole process of pattern recognition (The step of regularization is not included in the present results.)*

4 Perspective

In the present paper only a part of the potential of pattern recognition systems on the base of discrete parametric representations is evaluated. Actually our main interest is – besides testing other regularization and coding methods – to install a feedback path from the associative memory to the process of regularization and coding. Our main idea is that the discrete representation of an image depends on a small set of parameters, which controls the regularization process. Depending on these parameters (for instance one for relative strength of interaction and one for the strength of anisotropy) one can get different information from one hypercolumnar representation by different regularizations. Examples are the coarse shape of an object or texture gradients. In this way *optimal* representations for every object can be stored together with the parameter values and used to apply a second step to the recognition process, in which the first estimation for image similarity is checked by means of stored parameters. This can be looked at as a verification or falsification of a recognition guess, which can be extended to a step by step hierarchy of recognition levels.

In addition to that we are presently working on a process, which constructs a whole scale space of hypercolumnar representations out of a single image and achieves invariance of translation and scaling by a hierarchical

[4]All images were aquired with a resolution of 128^2 pixels and 256 levels of grey, and were filtered with a set of 16 different oriented gabors. To code an edge, any coding sets 3 out of 16 bits for maximum orientation. Altogether this means a data reduction of about 42:1.

search process in this data structure.

A third extension is the inclusion of dynamic patterns, which a especially well suited for edge oriented representations.

As a part of our group's active vision concept with neural architecture, the system presented here will be integrated in a complex attention driven and knowledge based system, which reacts to new stimuli in a changing environment, and as well recognizes known objects. Therefore the system combines top down and bottom up approaches.

References

[1] Hubel and Wiesel (1977) *Functional Architecture of macaque visual cortex* Proc. R. Soc. London B 198:1-59

[2] H. Mallot, W. v. Seelen; Naturwissenschaften, in preparation

[3] R. Durbin and G. Mitchinson *A dimension reduction framework for understanding cortical maps* Letters to Nature Vol 343 Feb.1990.

[4] G. Palm: *On Associative Memory.* Biol. Cybern. 36, 19-31 (1980)

[5] G. Palm: *On the storage capacity of an associative memory with randomly distributed storage Elements.* Biol. Cybern. 39, 125-127 (1981)

[6] A. Fuchs, H. Haken: *Pattern Recognition and Associative Memory as Dynamical Processes in a Synergetic System (I).* Biol. Cybern. 60, 17-22 (1988)

[7] A. Fuchs, H. Haken: *Pattern Recognition and Associative Memory as Dynamical Processes in a Synergetic System (II).* Biol. Cybern. 60, 107-109 (1988)

[8] D.J. Willshaw, O.P. Buneman, H.C. Longuet-Higgins: *Non-Holographic Associative Memory* Nature 222, 960-962 (1969)