Available online at www.sciencedirect.com

SCIENCE DIRECT®

Neurocomputing 69 (2006) 1219–1223

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Storing and restoring visual input with collaborative rank coding and associative memory

Martin Rehn[a,*], Friedrich T. Sommer[b]

[a]KTH, NADA, Royal Institute of Technology, SE-100 44 Stockholm, Sweden
[b]Redwood Center for Theoretical Neuroscience, University of California, Berkeley 132 Barker, Berkeley, CA 94720-3190, USA

Abstract

Associative memory in cortical circuits has been held as a major mechanism for content-addressable memory. Hebbian synapses implement associative memory efficiently when storing sparse binary activity patterns. However, in models of sensory processing, representations are graded and not binary. Thus, it has been an unresolved question how sensory computation could exploit cortical associative memory.

Here we propose a way how sensory processing could benefit from memory in cortical circuitry. We describe a new collaborative method of rank coding for converting graded stimuli, such as natural images, into sequences of synchronous spike volleys. Such sequences of sparse binary patterns can be efficiently processed in associative memory of the Willshaw type. We evaluate storage capacity and noise tolerance of the proposed system and demonstrate its use in cleanup and fill-in for noisy or occluded visual input.

Keywords: Sensory coding; Attractor memory; Rank coding; Sequence memory; Data compression

---

1. Introduction

The microcircuitry of the cerebral cortex shows extensive recurrent connectivity between pyramidal cells in layer II/III. These connections are plastic and have been shown to obey a temporally asymmetric Hebbian learning rule [2]. Associative memories are computational models that describe how cortical circuits may exploit these plastic connections to perform memory operations. Recently, physiological experiments in slice have lent support to these models by showing that activity organizes in repeatable sequences of activity patterns [7]. Under certain conditions, neural associative memories can efficiently store and retrieve large numbers of patterns of neural activity. These conditions are in particular that the patterns are binary and sparse, i.e., that the stored patterns share a low ratio of active cells [17,10,11]. At first glance these conditions seem ill-suited for sensory processing. Visual input, for example, when represented by Gabor-type filters, is graded and nonsparse. However, the recent discovery of discrete so-called UP states in visual cortex provides at least indirect evidence that binary and sparse information processing could be relevant even in early sensory processing [4]. To date, no biologically plausible memory model has been proposed that can store large numbers of chunks of analog raw sensory data, such as images. The aim of this paper is to propose such a model.

It has been shown that principles of efficient coding [14] and also faster but suboptimal techniques of signal representation, such as matching pursuit can sparsify sensory neural representations. These mechanisms can be neuronally implemented by lateral inhibition. Matching pursuit, in particular, has been suggested to convert sensory input into temporal sequences of spikes [13]. For efficient processing of sensory information we propose a combination of a new model of sparse visual coding, extending the model based on matching pursuit, and sparse sequence associative memory [1,18].

---

*Corresponding author. +46 7 90 7784; fax: +46 8 790 0930.
E-mail addresses: rehn@nada.kth.se (M. Rehn), fsommer@berkeley.edu (F.T. Sommer).

0925-2312/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2005.12.080

2. Image coding

2.1. Matching pursuit for the spike coding of visual input

Perrinet et al. [13] have proposed a model of the visual cortex based on a method of signal representation known as matching pursuit. In this model, neurons spike one at a time, each spike being elicited from the neuron that is most strongly excited by the sensory input. Neurons that spike inhibit other cells with similar receptive fields ("explaining away"). The representation of visual input by matching pursuit can be written as

$$\mathbf{x} = \sum_{n=0}^{m-1} \langle \mathbf{r}_n, \mathbf{\Psi}_{\gamma_n} \rangle \mathbf{\Psi}_{\gamma_n} + \mathbf{r}_m, \tag{1}$$

$$\gamma_n = \mathrm{argmax}_{\gamma_n} \langle \mathbf{r}_n, \mathbf{\Psi}_{\gamma_n} \rangle, \tag{2}$$

where $\mathbf{\Psi}$ is a set of basis functions, $\mathbf{x}$ is the visual input and $\mathbf{r}_n$ is the residual after the n:th spike. Eq. (2) indicates that the neuron γ_n to spike next will be the one corresponding to the basis function most similar to the current residual $\mathbf{r}_n$. To determine how many spikes should be used in the sequence, one can optimize the representation based on a cost function. We use the function

$$E(\mathbf{x}, \mathbf{b}) = \frac{1}{2} \sum_{i=1}^{m} \left(x_i - \sum_{j=1}^{n} b_j \Psi_{ji} \right)^2 + f(\mathbf{b}), \tag{3}$$

where the first term quantifies the quality of the representation $\mathbf{b}$, as generated by matching pursuit; $b_j = \sum_{n:\gamma_n = j} \langle \mathbf{r}_n, \mathbf{\Psi}_{\gamma_n} \rangle$. The second term quantifies the metabolic cost of the representation. Assuming that each spike is associated with a fixed metabolic energy cost, we just use the spike count

$$f(\mathbf{b}) = \theta \|\mathbf{b}\|_{L0}. \tag{4}$$

Depending on the choice of the sparseness parameter θ this coding scheme can produce sparse codes. The codes have few nonzero elements but are not yet binary. They contain either zeros or analog expansion coefficients $\mathbf{b}$. It has been demonstrated, however, that the exact analog values are not needed for faithful reconstruction of visual input. They may be replaced by mean values from a rank ordered histogram of coefficients that is averaged over many visual inputs [12]. Thus, an input pattern in a patch of the visual field can be coded by a temporal sequence of spikes, where only the spike order is significant.

2.2. Collaborative rank coding of image fragments

To use spike coding based on matching pursuit in combination with efficient associative memory, the coding strategy explained so far has to be extended. In our model of *collaborative* rank coding, cortical regions processing different patches of visual input collaborate to form spatio-temporal patterns. In our model we assume that an image is tiled by small nonoverlapping patches in the visual space that are processed in parallel by sets of neurons in different regions of primary visual cortex. For each patch, matching pursuit is used to determine the next spiking neuron. But spike timing in different patches is not independent as in the Perrinet model. We include a global, synchronizing influence which could be realized in the cortex by local thresholds that are synchronously oscillating [3]. The effect of this collaborative rank coding is that spikes in different patches organize into synchronized volleys of spikes. The number of spikes per volley can be regulated by the degree of threshold modulation. In our computer model we simply group the k largest coefficients in the first volley, the following k largest in the second and so on until less than k nonzero coefficients remain, at which point the sequence is truncated. For image number μ, we define r^μ as the index vector of the nonzero coefficients in $\mathbf{b}$, ordered according to descending magnitude. We denote by $|r^\mu|$ the cardinality of nonzero coefficients. The input image $\mathbf{x}^\mu$ is then represented by a sequence of patterns $\xi^\mu(t)$ of length T^μ; $t \in [1 \ldots T^\mu]$:

$$\xi_i^\mu(t) = \begin{cases} 1 & r^\mu(i) \in ((t-1)k,\, tk], \\ 0 & \text{otherwise}, \end{cases} \tag{5}$$

$$T^\mu = \lfloor |r^\mu| / k \rfloor. \tag{6}$$

This k-winner-take-all coding strategy implements the collaborative rank coding.

For reconstruction we form a rank code lookup table for the analog coefficients, based on the volley index (see Fig. 1). Comparing the error bars of collaborative and non-collaborative rank coding in Fig. 1 the collaboration seems to increase the accuracy of the lookup. An analysis of the rank code statistics of these coding schemes will be given elsewhere. Most importantly, the collaborative rank coding allows for the reconstruction of a stimulus, given only an joint spike volley sequence and the collaborative rank code lookup table.

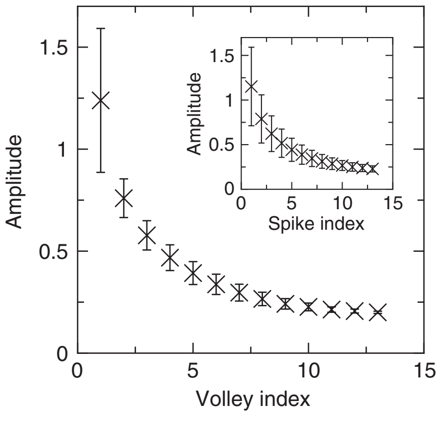

Fig. 1. Coefficient amplitudes as a function of volley index. Error bars show one standard deviation. The inset figure is for non-collaborative rank coding. Note the larger variance for the latter.

3. Memory network model

The collaborative rank coding described in the previous section, converts graded visual stimuli into sequences of coincident spike volleys which correspond to sparse patterns. For storage of the sparse sequences we use a Willshaw memory [17] with dynamic threshold control [15], implementing a k-winner take all scheme. By associating the last pattern in a sequence with the first we turn the image-encoding pattern sequences into closed loops, that can reverberate in the absence of input. When p images are stored, the binary synaptic matrix is given by

$$w_{ij} = \max_{\mu=1}^{p} \max_{t=1}^{T^{\mu}} (\xi_i^{\mu}(t)\xi_j^{\mu}(t-1)); \quad \xi_j^{\mu}(0) := \xi_j^{\mu}(T^{\mu}). \tag{7}$$

To retrieve stored images from noisy or distorted cues, the cues are transformed into spike volley sequences using the collaborative rank coding scheme. Those sequences are then fed to the associative memory. Since the most relevant features will occur in the first few volleys after stimulus onset, we reduce the relative influence of the following volleys in comparison to the retrieved signal produced by the associative memory; we ramp the influence of the input linearly down such that it reaches zero at the end of the input sequence

$$\xi_i^{\text{ret}}(t) = \begin{cases} k\text{WTA}_i(w_{ij}\xi_j^{\text{ret}}(t-1) + \eta(t)\xi_i^{\text{cue}}(t)) & t > 0, \\ 0 & t = 0, \end{cases} \tag{8}$$

$$\eta(t) = \max(0, 1 - t/T^{\text{cue}}). \tag{9}$$

For detecting successful retrieval, we exploit the cyclic nature of the stored sequences. The first pattern in the input sequence is buffered in a short term memory and compared to the output patterns from the associative memory. Whenever both patterns are sufficiently similar, readout of the stored information begins; i.e. when $\langle \xi^{\text{cue}}(1), \xi^{\text{ret}}(t) \rangle \geqslant \Theta$. If such a match does not occur, the input does not elicit memory retrieval. In cases of ambiguous cues the recognition thresholding allows the system to refuse retrieval, rather than to produce random associations. We refer to this as a detected failure; an undetected failure on the other hand is when a random association is produced after all.

4. Results

We evaluate our model by storing image fragments with 24×24 grayscale pixels and then test storage and retrieval under conditions of spatially diffuse noise or spatially defined occlusion. The system is used to store natural images, that have previously been whitened with respect to their spatial frequencies [9]. The images are tiled into nine 8×8 patches. The neuronal representation of the image fragments is three times overcomplete. To reflect properties of real neurons, we restrict all coefficients to positive values. For compensation of this representational restriction we double the number of neurons. Thus, there is a total of $n = 8 \cdot 8 \cdot 3 \cdot 2 = 384$ neurons for each patch. We set sparseness parameter in Eq. 4 to $\theta = 0.0178$, the number of spikes per volley is set to $k = 10$ and the threshold for retrieval detection to $\Theta = 4$.

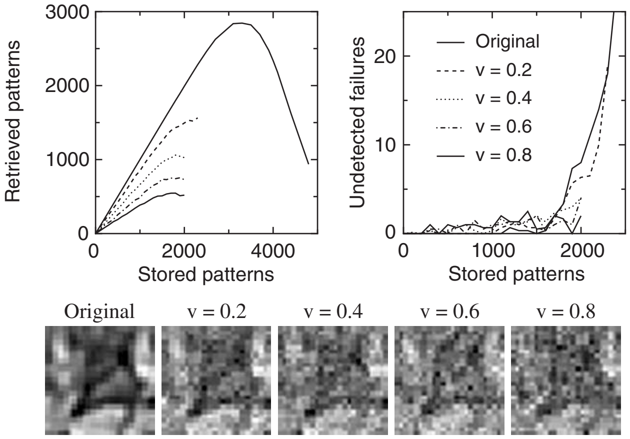

Fig. 2. Storage capacity and noise tolerance. Noisy cues were generated for the stored images, where the noise for each pixel is drawn from an independent Gaussian distribution with zero mean and variance $V_{\text{noise}}/V_{\text{signal}} = v$. The different lines correspond to noise levels $v = [0, 0.2, 0.4, 0.6, 0.8]$. In the left hand panel, storage capacity is shown; except for the noiseless case, the curves are truncated near the peak capacity. In the right hand panel, the number of undetected failures is shown (see text). The bottom panel illustrates the different noise levels in the cues.

The signal-to-noise ratio in reconstructed images is calculated as $S/N = 20\log_{10}(V_{\text{rec}}/V_{\text{err}})$, where V_{rec} and V_{err} are the variances in the reconstructed image and in the reconstruction error, respectively. The signal-to-noise ratio in the image reconstruction with the analog matching pursuit coefficients is about 21. When these are replaced by values from the collaborative rank coding lookup table, the signal-to-noise ratio is still as high as 17. This is also the quality of the images that can be retrieved from memory. As can be seen in Fig. 4, the subjective quality of image retrieval is good.

With the parameters setting described above, the maximum number of images that could be retrieved was found to be around 2800. There are 3456 neurons in the network, so the storage capacity is about 0.8 images per neuron. Each image is represented by a sequence of about 10 patterns. This was achieved at a memory load (ratio of nonzero weights in the memory matrix) of 0.22. The top solid line in Fig. 2 shows how the number of successfully retrieved images depends on the number of stored images under noiseless conditions. We also assessed the signal-to-noise gain when using the system to clean up noisy images; it is illustrated in Figs. 3, 4.

5. Discussion

We have proposed a new model for efficient storage and retrieval of raw images. At the heart of the model is a new scheme to code raw images into sequences of sparse patterns, the method called collaborative rank coding. This

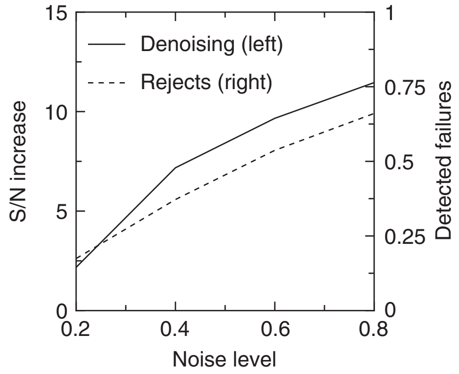

Fig. 3. Denoising by retrieving images from noisy cues. The solid curve (left axis) shows the improvement of signal-to-noise ratio in images retrieved by the network model, as compared to their cues. Detected failures are excluded. The incidence of detected failures is illustrated by the dashed curve (right axis).

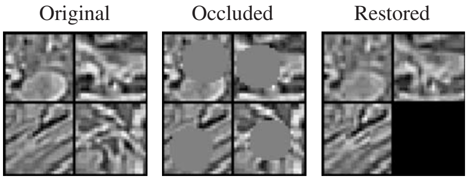

Fig. 4. Retrieval of occluded pictures. Four different original images are shown to the left, occluded cues in the middle and network output on the right. For the bottom right image retrieval failed; a detected failure. The successfully retrieved images also show the quality of the rank coded images.

coding method is combined with an efficient associative memory model [15]. The system uses only binary synapses to store the graded images.

Two common strategies for representing graded information are rate coding and latency coding. The very rapid processing times for visual input severely limit the usefulness of a rate code and also restrict the resolution available to a latency code [5]. In our system, images are represented by sequences of spike volleys, where only synchrony and temporal order is important. The coding is also fast since the most important information is transmitted first. An interesting property of this representation is that the sparse patterns in the sequences all have the same low level of activity k. If an image contains more structure, the sequence will be longer, but the sparsity level can be kept unchanged. Thus, this representational scheme offers a solution to the long standing problem how chunks of raw data with high information content can be represented by sparse patterns with low information content [19]. Our memory system allows massive storage and flexible retrieval of natural images in a neural architecture. Theoretical analyses of associative memories of the Willshaw type suggest that the storage capacity of our system should scale favorably with the size of the network, though this remains to be verified in large-scale simulations.

Whether lower levels in sensory systems of the brain employ content-addressed memory is still an unresolved issue. To resolve this question, predictions from computational models will be important. Here our model has demonstrated how principles of binary sparse sensory coding, sparse associative memory and the conversion from spatial into temporal structure can be used for cleanup, fill-in and memory retrieval of visual representations. The hypothesis of content-addressed memory in earlier stages of vision is supported by various experimental finding, for example, filling-in processes in early visual representations [8] and the recall of low-level visual features in dreams [6,16].

References

[1] S.-I. Amari, Learning patterns and pattern sequences by self-organizations of threshold elements, IEEE Trans. Comput. C-21 (11) (1972) 1197–1206.

[2] G. Bi, M. Poo, Synaptic modification by correlated activity: Hebb's postulate revisited, Annu. Rev. Neurosci. 24 (2001) 139–166.

[3] Braitenberg, Cell assemblies in the cerebral cortex, in: R. Heim, G. Palm (Eds.), Theoretical Approaches to Complex Systems, Springer, Berlin, 1978, pp. 171–188.

[4] R. Cossart, D. Aronov, R. Yuste, Attractor dynamics of network up states in the neocortex, Nature 423 (6937) (2003) 283–288.

[5] M. Fabre-Thorpe, A. Delorme, C. Marlot, S.J. Thorpe, A limit to the speed of processing in ultra-rapid visual categorisation of novel natural scenes, J. Cognitive Neurosci. 13 (2001) 171–180.

[6] D.J. Felleman, D.C. Van Essen, Distributed hierarchical processing in the primate cerebral cortex, Cerebral Cortex 1 (1991) 1–47.

[7] Y. Ikegaya, G. Aaron, R. Cossart, D. Aronov, I. Lampl, D. Ferster, R. Yuste, Synfire chains and cortical songs: temporal modules of cortical activity, Science 304 (5670) (2004) 559–564.

[8] M. Meng, D.A. Remus, F. Tong, Filling-in of visual phantoms in the human brain, Nat. Neurosci. 8 (2005) 1248–1254.

[9] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (1996) 607–608.

[10] G. Palm, On associative memory, Biol. Cybern. 36 (1980) 19–31.

[11] G. Palm, F.T. Sommer, Associative data storage and retrieval in neural networks, in: E. Domany, J.L. van Hemmen, K.S. (Eds.), Models of Neural Networks III, Springer, New York, 1995, pp. 79–118.

[12] L. Perrinet, M. Samuelides, S. Thorpe, Coding static natural images using spiking event times: do neurons cooperate?, IEEE Trans. Neural Network 15 (5) (2004) 1164–1175.

[13] L. Perrinet, M. Samuelides, S. Thorpe, Sparse spike coding in an asynchronous feed-forward multi-layer neural network using matching pursuit, Neurocomputing 57 (2004) 125–134.

[14] M. Rehn, F.T. Sommer, A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields, 2005, submitted for publication.

[15] F. Schwenker, F.T. Sommer, G. Palm, Iterative retrieval of sparsely coded associative memory patterns, Neural Networks 9 (3) (1996) 445–455.

[16] R. Stickgold, A. Malia, D. Maguire, D. Roddenberry, M. O'Connor, Replaying the game: hypnagogic images in normals and amnesics, Science 290 (5490) (2000) 350–353.

[17] D.J. Willshaw, O.P. Buneman, H.C. Longuet-Higgins, Non-holographic associative memory, Nature 222 (1969) 960–962.

[18] G. Willwacher, Storage of a temporal sequence in a network, Biol. Cybern. 43 (1982) 115–126.

[19] C. Zetsche, Sparse coding: the link between low level vision and associative memory, in: R. Eckmiller, G. Hartmann, G. Hauske (Eds.), Parallel Processing in Neural Systems and Computers, Elsevier Science Publishers B.V., North Holland, 1990.

Martin Rehn received his M.Sc. in Engineering Physics from the Royal Institute of Technology. He is currently pursuing a Ph.D. in Computer Science in the Computational Biology and Neurocomputing group at that university. His research interests include sensory representation, attractor memories and modelling of cortical circuits.

Friedrich T. Sommer is with the Redwood Center for Theoretical Neuroscience and the Helen Wills Neuroscience Institute at University of California, Berkeley. 2003–2005 he has been a principal investigator at the Redwood Neuroscience Institute in Menlo Park. In 2002 he received his habilitation in Computer Science from University of Ulm. He holds a Ph.D. and the diploma in Physics from the universities of Düsseldorf and Tübingen.