ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

Learning of cognitive maps from sequences of views

Hanspeter A. Mallot and Bernhard Schölkopf

Max–Planck–Institut für biologische Kybernetik
Spemannstr. 38, 72076 Tübingen, Germany

Abstract. This paper presents a scheme for learning a cognitive map of a maze from a sequence of views and movement decisions. The scheme is based on an intermediate representation called the *view graph*, whose nodes correspond to the views while the labelled edges represent the movements leading from one view to another. By means of a graph theoretical reconstruction method, the view graph is shown to carry complete information on the topological and directional structure of the maze. Path planning can be carried out directly in the view graph without actually performing this reconstruction. A neural network is presented that learns the view graph during a random exploration of the maze. It is based on an unsupervised competitive learning rule translating temporal sequence (rather than similarity) of views into connectedness in the network. The network uses its knowledge of the topological and directional structure of the maze to generate expectations about which views are likely to be encountered next, improving the view recognition performance.

1. Introduction

1.1. Representation of 3D shape and space

Information on spatial relations in the environment is crucial for the generation and control of most kinds of behaviours both in living beings and in robots. One source of such information is the sequence of retinal images. Extensive research efforts have been directed towards extracting the 3D information included in these images in an *explicit* form, i.e. to construct various representations of depth from them. Examples of such representations include generalized cylinders in the context of object recognition (Marr and Nishihara, 1978) or cognitive maps for representing the environment (see Gallistel 1990, O'Keefe 1991).

One way to think about representations is as a piece of information that has been made explicit in the brain. If the 3D structure of an object is explicitly known, it should in principle be possible to predict novel views of this object. Recent psychophysical work shows that this is not the case: even if ample 3D information is provided (e.g., by stereo or the kinetic depth effect), recognition

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

Table 1: Map behaviour, possible sources of information, and hypothetical representations employed

Behavioural Competences	**Sources of Information**	**Possible Representations**
• Repeat a previously travelled path • Find known target from new starting point • Find shortcuts • By–pass newly blocked connections • Communicate about paths	• Pointers (guidances) and general rules • Global compasses (sun, chemical gradients, etc.) • View sequences and Landmarks • Path integration (dead reckoning)	• Associations between views and motor commands • Relational information on places, connections, and views (e.g., graph structures) • Topographic maps representing metric relations

is much harder for novel views than for familiar views included in the training set (Bülthoff and Edelman 1992). One possible interpretation of this result is that 2D views rather than some explicit representation of 3D shape are stored in the brain.

Instead of making all available image information explicit and storing it in a representation, one could think of picking just the right pieces of information required for a given behavioural task. In this view, perception is part of a perception–action cycle that uses image information with the least possible amount of intermediate computation. A particularly interesting way to state the underlying question is this: how complicated can spatial behaviour get without using explicit representations of space? In this paper, this question is explored in the field of navigation in space, i.e., maze exploration.

1.2. Cognitive maps and the perception–action cycle for map behaviour

A *cognitive map* is a neural mechanism which enables its user to solve navigation and orientation tasks as if using a real map of the environment (see Table 1). Examples of map behaviour include: repetition of previously travelled routes, finding detours around obstacles, approaching a goal from a novel starting point, finding a path between two arbitrary points of the environment (topological use of the map), finding the shortest path (topological and metrical use of the map), etc. These competences are based on a number of information

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

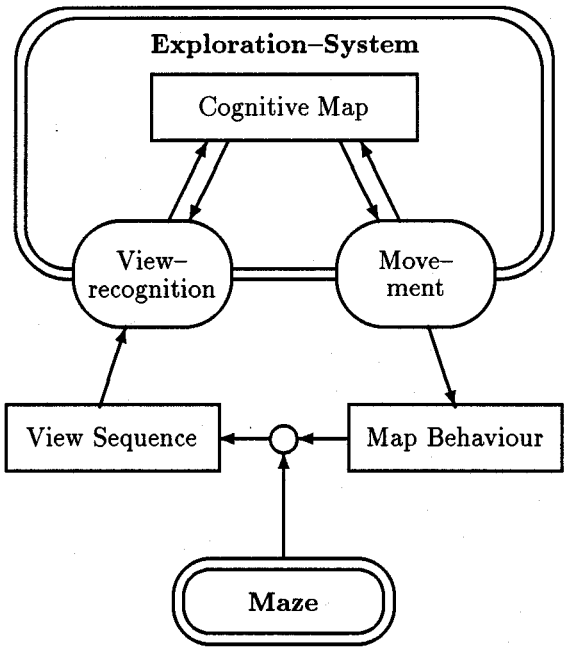

Figure 1: Cognitive maps and the perception–action cycle. The sensory input available to the system is a sequence of views reflecting the structure of the maze and the observers movement decisions. The cognitive map stores information on the interrelation of the view sequence and the movements taken (upward arrows). It also generates "expectations" in the sense that (i) similar views can be distinguished by their position in the maze and that (ii) movement decisions are guided by the cognitive map (e.g., for exploration or goal–finding). For further explanations see text.

sources some of which are listed in Table 1. In this paper, we restrict ourself to sequences of views as the sole source of information.

A cognitive map contains two types of information, concerning (a) the recognition of places and (b) the connections between them. Previous approaches have often focused on the first problem and tried to learn paths as an association between recognized views or places and motor commands (for review, see O'Keefe, 1991). In this paper, we give an explicit account of the connectivity structure of the explored environment in terms of the view–graph.

The notion of a cognitive map pursued in this paper is illustrated in Fig. 1 as part of an action–perception cycle. Since we restrict ourselves to cognitive maps of mazes, sensory input can be described as a sequence of views reflecting the connectivity structure of the maze together with the movement decisions taken by the system exploring the maze. Another advantage of mazes is the fact that there are only a small number of possible (egocentric) movements, such as "go left", "go right", "go back", "turn left", etc. The cognitive map does

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

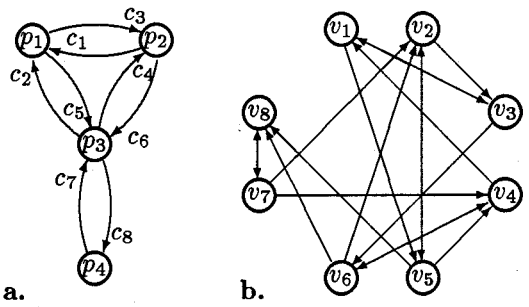

Figure 2: **a.** Simple maze shown as a directed graph with places p_i and corridors c_j. **b.** Associated view–graph where each node v_i corresponds to one view, i.e. one directed connection in the place graph. Only the edges with go–labels are shown. In graph theory, b. is called the *interchange graph* of a. Simpler plots of b. are possible but not required for our argument. **c.** Adjacency matrix of the view–graph with labels indicating the movement leading from one view to another. Go–labels (involving a locomotion from one place to another): g_l (go left), g_r (go right), g_b (go backward). Turn–labels (e.g., probing of corridors): t_l (turn left), t_r (turn right), t_o (stay).

not represent the complete information of the maze but only those aspects that are relevant to (i) the interpretation of view–sequences or (ii) the generation of map behavior. The left double–arrow in Fig. 1 indicates the interaction of view recognition and the cognitive map: if a view is recognized, the current position in the map can be updated. Vice versa, knowledge on the current position in the map helps distinguish similar views occuring at different locations. The right double–arrow indicates the fact that movement decisions must be associated with the resulting view changes in order to make the cognitive map predictive; also, map information will be used to make movement decisions with respect to a given plan (such as exploration, approach to a target, etc.).

In Sect. 2., we present the view–graph as a data–structure that is built on concepts such as views and movements and is able to represent all relevant information on mazes. The argument will be based mostly on mathematical graph theory. Sect. 3. deals with an artificial neural network reconstructing a cognitive map (i.e. a view–graph) from a sequence of views and motion decisions.

2. The view–graph as a sufficient representation of mazes

2.1. Places, views, and movements

Consider a simple maze composed of *places* $p_1, ..., p_n$ and *corridors* $c_1, ..., c_m$ (Fig. 2a). One way to think of this maze is a graph where the places are the nodes and the corridors are the edges. We consider all corridors to be

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

directional but allow for the existence of two corridors with opposite directions between any two nodes. Throughout the paper, we will assume that there are no isolated places in the maze, i.e., each place can be reached from any other place by at least one sequence of corridors.

When exploring a maze, the observer generates a sequence of movement decisions defining a path through the maze. In doing so, he encounters a sequence of *views* from which he wants to recover the place–graph. In order to study the relation of views, places and movements, we make the following simplifying assumptions.

Views: There is a one–to–one correspondence between directed corridors and views. All views are distinguishable and there are no "inner views" in a place that do not correspond to a corridor.

Movements: At each time step, one movement from a finite (usually small) set is selected. Thus, the observer knows if he simply probed a corridor (by turning towards it without moving) or if he actually walked it. Clearly, the same view would be encountered in both cases.

With these assumptions, we can construct the *view–graph* that an observer will experience when exploring a maze (Fig. 2b,c). Its elements are:

- The **nodes** of the view–graphs are the views v_p, which, from the above assumption, are simply identical to the corridors in the place–graph. We denote the start and target place of a view v by $P_{out}(v)$ and $P_{in}(v)$, respectively. Of course, the functions P_{in} and P_{out} are not known when exploring the maze.

- The **edges** of the view–graph indicate temporal coherence: two views are connected, if they can be experienced in immediate temporal sequence. The edges are labelled with the movements resulting in the corresponding view sequence. It is convenient to distinguish two types of labels:
  - "Go–labels" specify movements leading from one place to another. In Fig. 2c, three types of go–labels are shown, indicating leftward, rightward and backward movements.
  - "Turn–labels" specify movements within one place, i.e. probing movements where a view is looked at but not walked to.

The resulting adjacency matrix with movement labels is depicted in Fig. 2c. Note that all edges starting from the same node will have different movement labels.

For the graph–theoretic argument in the rest of this section, we identify all go–labels and neglect the turn–labels. Thus, two views v_p and v_q are connected iff $P_{in}(q) = P_{out}(p)$. We denote by $A = (a_{pq})$ the simplified *adjacency matrix* of the view–graph:

$$a_{pq} = \begin{cases} 1 & P_{in}(q) = P_{out}(p) \\ 0 & \text{otherwise} \end{cases} \qquad (1)$$

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

It can be obtained from the adjacency matrix of Fig. 2c by replacing all go–labels by ones and the turn–labels and bars by zeros. Taken together, the (now unlabelled) view–graph defined here is the *interchange graph* (e.g., Wagner 1970) of the place–graph.

2.2. Which information does the view–graph represent?

In this Section, we assume that the view–graph has been learnt, e.g. by the neural network presented in Sect. 3. The question then is: does the view–graph contain enough information to guide map behaviour? To answer this question, we will now mathematically reconstruct the underlying maze–graph from the view–graph. It turns out that this can be done already with the simplified (unlabelled) view–graph introduced in the previous section.

It should be clear that the place–graph can be recovered from the view–graph only up to permutations of the place– and view–numbers. Strictly speaking, we will construct a third graph whose nodes are sets of views leading to one place. This graph will be isomorphic to the original place–graph. We start by defining the *successor* $\mathcal{S}(v_p)$ of a view v_p as the set of views that can follow v_p:

$$\mathcal{S}(v_p) = \{v_r | a_{rp} = 1\} \tag{2}$$

The reconstruction method rests on the idea that two views lead to the same place, if they have the same successor:

$$P_{in}(v_p) = P_{in}(v_q) \Longleftrightarrow \mathcal{S}(v_p) = \mathcal{S}(v_q). \tag{3}$$

To prove this equivalence, we substitute from Eqs. 2, 1 and obtain:

$$\begin{aligned} \mathcal{S}(v_p) = \mathcal{S}(v_q) &\iff \{v_r | a_{pr} = 1\} = \{v_s | a_{qs} = 1\} \\ &\iff \{v_r | P_{out}(r) = P_{in}(p)\} = \{v_s | P_{out}(s) = P_{in}(q)\} \end{aligned}$$

The last two sets cannot be empty, because the maze does not contain isolated places. Since the indices r and s refer to the same arbitrary numbering of the set of views, it is easy to see that the final equality implies $P_{in}(p) = P_{in}(q)$, which proves our proposition. Note that in addition to Eq. 3, we have proved $\mathcal{S}(v_p) \cap \mathcal{S}(v_q) \neq \emptyset \Longleftrightarrow \mathcal{S}(v_p) = \mathcal{S}(v_q)$: Successors are either disjoint or identical.

While we cannot recover the functions P_{in} and P_{out} explicitly, Eq. 3 shows that the equality of their successors partitions the set of all views into n subsets correponding to the places $\{p_i | i \leq 1 \leq n\}$. These are the nodes of the recovered graph. The corridors can be found by inspection of the view–graph.

The above results were obtained for the unlabelled version of the view–graph. Since this can always be derived from the labelled view–graph, the results hold for the richer graph *a forteriori*. It should be clear that the reconstruction described here need not be carried out in the brain; it simply illustrates the sufficiency of the representation.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

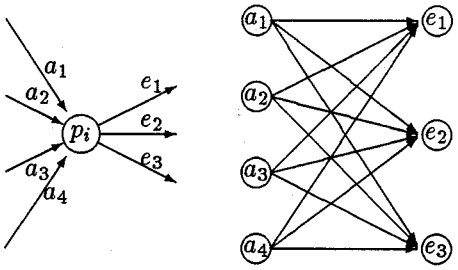

Figure 3: Subgraphs of maze– and view–graph corresponding to one place. The subgraph of views is a complete bipartite graph since every entry view, a_i, is connected to every exit view, e_j (i.e., the successors of all entry views are the same).

2.3. Redundancy of the view–graph and the completion matrix

The fact that all views leading into the same place have the same successor can be restated by saying that each place in the maze corresponds to a complete bipartite subgraph in the view–graph (see Fig. 3). A complete bipartite graph consists of two subsets of nodes such that each node from one subset is connected to each node of the second subset while there are no connections within subsets. Here, the two subsets are the views leading into one place and the views leading away from this place. This second set is the successor of each member of the first set.

This structure of the view–graph entails an interesting type of redundancy which might be useful in exploration. Suppose that view a_1 is known to have the views e_1, e_2 in its successor and view a_2 is known to have e_1 in its successor. It can than be predicted that the sequence $a_2 \rightarrow e_2$ should also exist. If it doesn't, the views e_1 as seen in the sequences $a_1 \rightarrow e_1$ and $a_2 \rightarrow e_1$ are probably confused and must be distinguished by further examination.

The bipartite structure of the subgraphs has an interesting counterpart in the adjacency matrix. The successor of view v_p (Eq. 2) corresponds to the pth column of the adjacency matrix A; its size is the number of ones in that column. The reconstruction result (Eq. 3) therefore implies that A can have only up to n (the number of places) different columns. Two columns of A are either identical or orthogonal (no coinciding ones). Therefore, for a suitable permutation of the views, the symmetrical matrix $A^\top A$ will have block structure with n blocks each corresponding to one bipartite subgraph of views or to one place in the maze. If A is interpreted as a transition matrix, $A^\top A$ describes a step forward followed by a step backward, not necessarily the reverse of the step forward. Starting with some view, this movement will reach all views leading to the same place. We call $A^\top A$ the *completion matrix* of the view–graph. The size of the blocks equals the frequency of a particular column (successor) in a matrix, i.e., the in–degree of the corresponding place. The values taken by the coefficients inside each block are constant and correspond to the out–degrees of the corresponding places. Outside the blocks, the matrix takes the value zero (cf. Fig. 6).

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

2.4. Is the view–graph planar?

In the previous section, we have shown that the view–graph contains all information required to reconstruct the place–graph. If we want to represent view–graphs in a neural network, it is important to find a network topology in which they can actually be embedded. For example, two–dimensional grids are best suited for planar graphs, i.e., graphs that can be drawn on a sheet of paper without intersecting edges. Since the problem of planarity is completely solved for bipartite graphs (e.g., Wagner, 1970), we can give a simple sufficient criterion for non–planarity of the view–graph: Complete bipartite graphs are non–planar if both subsets contain at least three nodes. Therefore, the view–graph of a maze containing a place where three or more corridors meet, cannot be planar. This is the case for all interesting mazes.

2.5. Paths and movement sequences

So far, we have shown that the (unlabelled) view–graph contains all information contained in the place–graph. However, for path planning some additional information is required. This can be seen from the fact that simple transformations such as mirroring applied to the maze would affect neither its graph structure nor the structure of the view–graph. It would, however, strongly affect the paths taken through the maze since all left/right decisions would have to be reversed.

In the original maze, directional information can be included by allocentric or world centered direction labels assigned to the corridors (e.g., in Fig. 2a, corridor c_1 would be labelled "west", corridor c_4 "north–east", etc.). In the view–graph, we can now use the egocentric labelling system introduced in Sect. 2.1. The reconstruction method given above was based solely on the go–labels and ignored the connections having turn–labels. Of course, the turn–labels will be useful for reconstruction of the places. It is an empirical problem to find out the relative importance of these labels in biological maze learning.

For the generation of movement sequences, the turn–labels are irrelevant. If a path is given connecting two views in the view–graph by edges carrying go–labels, a corresponding movement sequence can easily be generated by simply listing the labels along the path.[1]

3. Learning mazes from view sequences

3.1. Self–organizing sequence map

We construct a neural network consisting of three sets of units: the input, the movement, and the map layer. In order to support orientation behaviour in a maze, the network must represent information concerning two problems:

[1] If two views are connected by a turn–label, they will also be connected by a two–step path using go–labels.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

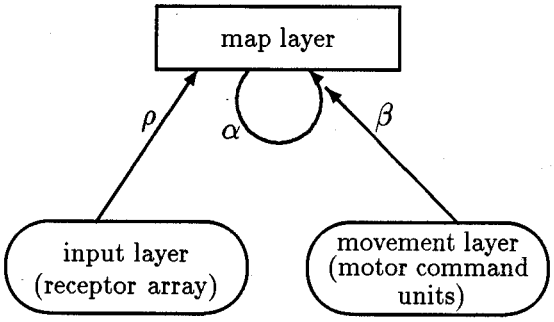

Figure 4: Structure of the neural network. ρ: input weights, α: map layer weights, β: presynaptic, modulatory weights from the movement units. Note the similarity to the intra–organismic part of the action–perception cycle depicted in Fig. 1.

P1 Each view must be identified and associated with a particular location in the map. In our network, this is achieved by a set of input weights ρ_{ij}. After learning, each view will be represented by activity in the associated map unit, i.e. the unit whose input weights are most closely tuned to the presented view. Neurons representing particular places have been recorded from rats (see O'Keefe, 1991). For results on directional place cells, which more closely correspond to our views, see McNaughton et al. (1983).

P2 The topology of the maze is represented by weights α_{ik} connecting units within the map layer. Weights between units will be assigned according to the temporal sequence of the views represented by these units. Movement labels are associated to the connections in the map layer by presynaptic facilitating connections from the corresponding movement unit.

While the proposed network is reminiscent of a standard Kohonen map (Kohonen 1982), there are two important differences: First, nearness in the map corresponds to temporal adjacency, not to featural similarity. In fact, similar views can occur at great distances in the maze and must not be confused. Second, distance in the map is measured as the minimal number of synapses that must be passed between two units (the "combinatorial distance" in graph theory). Therefore the topological structure of the resulting "map" is not limited (see Martinetz and Schulten 1994). This is desirable since view–graphs need not be planar.

3.2. Description of the model

Network structure. The network consists of an *input layer* (J units) with activity variables f_j^t, a *map layer* (I units) with activity variables e_i^t and thresholds θ_i^t, and a *movement layer* (K units) with activity variables m_k^t; superscripts denote time steps. The input layer is fully connected to the map layer via the input weights ρ_{ij}^t. For each map unit i, the weight vector $\mathbf{r}_i := (\rho_{i1}, ..., \rho_{iJ})$ is called its "receptive field". As a consequence of the Cauchy–Schwarz inequality, the map unit will be activated most strongly by an input vector identical

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

to its receptive field. Within the map layer, synaptic weights α_{in}^t are initialized to zero but may evolve during learning. The movement layer contains one unit for each possible movement (label of the view–graph connections). Each unit will establish facilitating presynaptic connections to all appropriate map layer connections. The map layer weight α_{in}^t may be facilitated by a presynaptic weight $\beta_{in,k}^t$, if m_k is the movement type associated with view connection α_{in}.

Input sequence. A sequence of movement signals is given as external input to the movement units. At the same time, a sequence of input vectors, $\mathbf{f}^t = (f_1^t, ..., f_J^t)^\top, t \in \mathbf{N}$, is simulated from the corresponding movements in some underlying maze–graph and presented to the input layer. In the first series of simulations, the movement decisions were chosen at random with equal probability, resulting in a random walk through the maze. For each view (directed corridor in the maze) a fixed view vector is chosen which is fed to the input layer of the network each time the random walk passes by. The view vectors can be either canonical base vectors (only one component different from zero) or random. In the first case, problem P1 above is assumed to be solved by some ideal preprocessing; in the second case, P1 and P2 are approached simultaneously.

Activation dynamics. The acitivity of the map layer units is described by

$$e_n^t = g\left(-\theta_n^{t-1} + \sum_{j=1}^{J} \rho_{nj}^t f_j^t + \sum_{i=1}^{I} \tilde{\alpha}_{ni}^t e_i^{t-1}\right), \tag{4}$$

where $g : \mathbf{R} \to [0, 1]$ denotes the logistic function. Let $w(t)$ denote the index of the most active cell (the winner cell), i.e.: $e_{w(t)}^t = \max_i\{e_i^t\}$. It represents the view currently perceived in the maze. The effects of incoming information are biased by an intrinsic term (the second sum in Eq. 4) such that the current winner is likely to be a unit connected to the last winner via a strong weight. By $\tilde{\alpha}$, we denote the joint effect of map layer weight and the presynaptic facilitation from the movement layer:

$$\tilde{\alpha}_{in}^t = \alpha_{in}^t + \sum_{k=1}^{K} (1 - \alpha_{ik}^t)\beta_{in,k}^t m_k. \tag{5}$$

Weight dynamics. The input weights ρ_{ij} are randomly initialized to values between 0 and 1, with a subsequent Euclidean normalization of each receptive field $\mathbf{r}_{w(t)} := (\rho_{i1}, \ldots \rho_{iJ})$. The receptive field of the winner unit approaches the presented input via the competitive learning rule with learning rate λ_1 (see Kohonen 1982)

$$\mathbf{r}_{w(t)}^{t+1} = \frac{\mathbf{r}_{w(t)}^t + \lambda_1 \mathbf{f}^t}{\|\mathbf{r}_{w(t)}^t + \lambda_1 \mathbf{f}^t\|}. \tag{6}$$

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

The input weights of the other units remain unchanged. Connections within the map layer are established between the last two winners and represent a transition between the last two views in the maze:

$$\alpha^t_{w(t),w(t-1)} = (1-\lambda_2)\alpha^{t-1}_{w(t),w(t-1)} + \lambda_2 \alpha_{max} \tag{7}$$

Here, λ_2 is a learning rate and α_{max} an upper limit for the weights.

Finally, the weights of the presynaptic connections $\beta_{in,k}$ are set to some constant $\phi \in (0,1]$ if movement m_k coincided with the last increase of α_{in}; otherwise, $\beta_{in,k}$ is zero.

Threshold control. Due to the intrinsic term in the activation dynamics, the network could converge to a state where just two strongly connected map units are the winner units for all view presentations. In order to overcome this problem, the thresholds of winner units increase according to the *threshold dynamics* (λ_3: learning rate, θ_{max}: maximal threshold):

$$\theta^t_{w(t)} = (1-\lambda_3)\theta^{t-1}_{w(t)} + \lambda_3 \theta_{max}. \tag{8}$$

4. Simulations

In this Section, we briefly summarize a number of simulation results obtained with the neural network model. View sequences were generated by a random walk through a 7–place (12–view) hexagonal maze. A network with $I = 20$ input units, $J = 64$ map units and $K = 4$ movement units was used. A more thorough account of the network's computational capabilities as well as a simple plath–planning algorithm have been presented elsewhere (Schölkopf and Mallot 1994).

4.1. Convergence

Convergence of the learning process is judged from the combination of two measures: *Neighborhood preservation rate* (NPR) is the frequency of view presentations in which the winner neuron receives a map layer connection from the winner neuron of the previous time step. If the network contains the map layer conections for all edges of the view graph and view recognition is correct,

Table 2: View graph topology convergence in the simulation. NPR: Neighbourhood preservation rate (cf. Sect. 4.1.).

Learning time	0	10	20	30	50	70	90	110	130	ideal
NPR (in %)	0	40	60	73	77	83	87	100	100	100
No. of connections	0	9	14	17	19	22	24	26	26	26

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

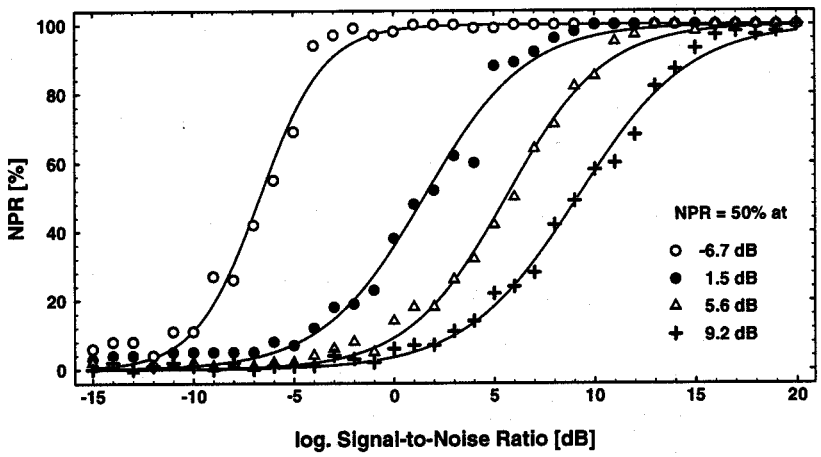

Figure 5: Neighbourhood preservation rates (NPR) measured in 200 testing steps for different amounts of Gaussian noise added to the input views. Learning time 110 steps. +: intrinsic connections cut. Δ: intrinsic connections without movement facilitation (topological biasing). •: intrinsic connections with movement facilitation. ○: additionally, the winner unit activity is set to one in each time step. The curves are logistic functions fitted to the data. The relative shift of the curves, i.e. the improvements achieved by adding the different features are given by the differences of the 50%–thresholds shown to the right.

neighborhood preservation rate becomes unity. While NPR increases during learning, the *total number of map layer connections* should be low since in a completely connected map layer, NPR would evaluate to one for trivial reasons. In Table 2, NPR and the number of map layer connections are shown for a number of learning steps. Learning is complete after 110 steps.

4.2. View Recognition

The intrinsic term of the map layer activation function (Eq. 4) helps recognize the views. In Fig. 5, view recognition as a function of signal–to–noise ratio is depicted by means of the neighborhood preservation rate, NPR (see Section 4.1.). If noise is added to the views, the map layer weights reduce the signal–to–noise ratio required for recognition by a factor of about 2 (3.6 dB). This indicates that the topological structure stored in the map layer weights is used to distinguish similar, but distant views. Further improvements are achieved by including the movement layer, indicating that knowledge about the last movement decision (i.e. the direction of approach) also helps distinguish similar views.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

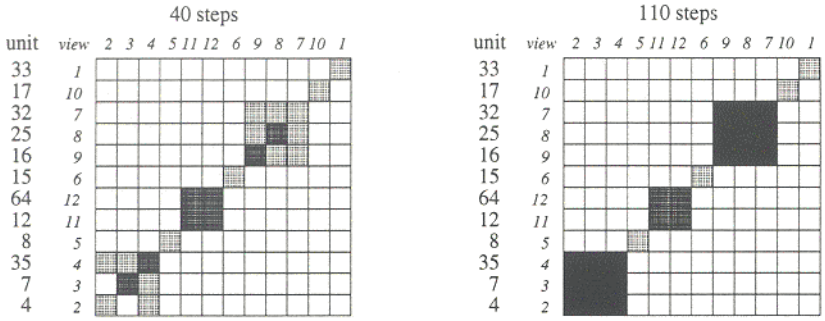

Figure 6: Completion matrices $\tilde{A}^\top \tilde{A}$ derived from the weight matrices after 40 and 110 learning steps. Grey–levels correspond to the values 0 (white) to 3 (black), unit numbers denote units in the map layer. The block structure is clearly visible after 40 steps, the underlying maze can be inferred already from this stage.

4.3. Maze Reconstruction

From the weight matrix, we derive an estimate of the adjacency matrix A of the view graph by deleting the all–zero rows and columns, thresholding the remaining entries, and suitable reordering of the rows and columns. The estimated completion matrix (cf. Section 2.3.) after 40 and 110 learning steps is shown in Fig. 6. The block structure is already fully developed after 40 learning steps. In addition, from the inhomogenities within the blocks, optimal strategies for further exploration of the maze can be derived.

5. Discussion

The examples presented in this paper indicate that view–based approaches can go a long way in the processing of spatial information. The information implicitly present in the images can support many behavioral competences without being transformed into explicit spatial representations.

The view–graph can also be used to represent three–dimensional objects by their various views. Edelman and Weinshall (1991) proposed a neural network model for rotation–invariant object recognition which recovers view–graphs of objects. The authors argue that views perceived in close temporal sequence are likely to belong to the same 3D object and should be linked in the view–graph. The predictions for view extrapolation made by view–based models have been confirmed in psychophysical experiments by Bülthoff and Edelman (1992). In maze reconstruction, the problem is slightly more complicated: again, views perceived in temporal sequence are connected, but do not in general belong to the same place. Object recognition is thus analogous to the reconstruction of the places of a maze from connections with "turn–labels". In order to recover the connections between places, "go–labels" are required as well (Sect. 2.2.).

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 277-290

In general, the amount, type, and explicitness of the information represented at a stage depends on the information processing streams interacting at this stage. If, for example, maze information could be gathered from additional sources such as global compass, general guidances and world knowledge (e.g.: "if you want to reach the water, go downward"), acoustic information, or even communication with other exploring systems, a common stage would be needed where all these inputs could be compared. A "good" representation is therefore not so much characterized by its explicitness but by its ability to integrate data from different streams. To what degree explicit representations of space are required to make the various inputs and outputs commensurable, is an open question.

References

[1] H. H. Bülthoff and S. Edelman. Psychophysical support for a two–dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, USA*, 89:60 – 64, 1992.

[2] S. Edelman and D. Weinshall. A self–organizing multiple–view representation of 3D objects. *Biological Cybernetics*, pages 209 – 219, 1991.

[3] C. R. Gallistel. *The organization of learning.* The MIT Press, Cambridge, MA, USA, 1990.

[4] T. Kohonen. Self–organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59 – 69, 1982.

[5] D. Marr and H. K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society (London) B*, 200:269 – 294, 1978.

[6] T. M. Martinetz and K. Schulten. Topology representing networks. *Neural Networks*, 7:507 – 522, 1994.

[7] B. L. McNaughton, C. A. Barnes, and J. O'Keefe. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely moving rats. *Experimental Brain Research*, 52:41 – 49, 1983.

[8] J. O'Keefe. The hippocampal cognitive map and navigational strategies. In J. Paillard, editor, *Brain and Space*, pages 273 – 295. Oxford University Press, Oxford, 1991.

[9] B. Schölkopf and H. A. Mallot. View–based cognitive mapping and path planning. Technical Report 007, Max–Planck–Institut für biologische Kybernetik, Tübingen, Germany, 1994. (Available as /pub/mpi-memos/TR-007.ps.Z via anonymous ftp from ftp@mpik-tueb.mpg.de.).

[10] K. Wagner. *Graphentheorie.* Bibliographisches Institut, Mannheim, Wien, Zürich, 1970.