Spatial Feature Binding and Learning in Competitive Neural Layer Architectures

Der Technischen Fakultät

der Universität Bielefeld

vorgelegt von

Heiko Wersing

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

März 2000

To my parents

Acknowledgements

This work was done in the Neuroinformatics group, headed by Prof. Helge Ritter, at the Faculty of Technology, University of Bielefeld and supported by the DFG Graduiertenkolleg Strukturbildungsprozesse.

First, I want to thank Helge for providing me with such an interesting topic with its interdisciplinary scope between biology, mathematics, computer science, and perhaps even physics. His creative support and encouragement gave a strong impetus to my scientific work. With his enthusiasm and optimism he reminded me many times, that science can be a truly exciting thing.

The Neuroinformatics Group was an excellent workplace, both with regard to the infrastructure which was very well maintained by Nils, and to the friendly atmosphere among my colleagues. I also want to thank Petra for her friendly ways of always willing to help with formal things.

Jochen accompanied all my time in the group, and the uncountable discussions with him have shaped many of the ideas which are written down in this work. I am most thankful to him for his competent criticism, his encouraging support, and simply for the nice time we had together during the last years. I want to thank Tim for the excellent collaboration, where he showed that my work can be really usefully applied. Our time together in the Graduiertenkolleg gained so much from his incorruptible sense of humour. Jörg contributed a lot to this work with his excellent Diploma thesis. I want to thank him for helping many times with the computer stuff, and for all the nice time we spent together outside the university. Jochen, Tim, Jörg, and Ute read the whole manuscript and helped me a lot to improve the presentation.

The Graduiertenkolleg Strukturbildungsprozesse provided an interdisciplinary forum with many interesting insights into neighbouring scientific disciplines. I want to thank the members of the Graduiertenkolleg for the friendly and informal atmosphere, which has led to many fruitful discussions. Special thanks go to Christian, Ralf and Martin for organizing everything so well, and to Prof. Wolf-Jürgen Beyn for the nice cooperation, which showed how much can be gained from the interdisciplinary approach.

Finally, I want to thank Ute, for many influential discussions on the topics I touched upon during this work, and what is most important, for her love.

I dedicate this work to my parents, for all the support they gave to me throughout my life.

Contents

Chapter 1

Introduction

One of the most fascinating scientific challenges of our time is to understand the brain, the biological basis of our mental capabilities like thought, perception, learning, and memory. It has taken the effort of a wide range of disciplines, such as biology, physics, mathematics, computer science, and psychology to contribute to this larger goal by offering different methods of analysis and levels of description. In an attempt to provide a coarse ordering of these levels we may place at the lower end the neurobiology of single nerve cells and the biochemical mechanisms governing their function. The upper end is then provided by the psychological concepts of the high-level functions of the brain such as memory, action, learning, behaviour, and consciousness. The relatively new research field of artificial neural networks aims at bridging the large gap between these levels by providing intermediate model descriptions which focus on the functional capabilities of abstracted neural networks, amenable to simulation and mathematical analysis. With concepts such as topographic feature mapping, associative memory and Hebbian learning, the field has shaped a number of current ideas of the principles of information processing in the brain. Nevertheless, compared to the universality of the physical laws that describe the phenomena of our physical surroundings, where our deep comprehension is reflected by their ubiquitous technical use, our current understanding of the relationship between the biological neural representation and high-level brain functions is at most fragmentary.

What is the reason for this gap in understanding the relations between the different levels of description of brain function? Nowadays the prevailing scientific tenet of the meaning of *understanding* a natural phenomenon or system is in the view of reductionism. This means that the system under study can be reduced to smaller subsystems, which are governed by laws that can be fully described and understood. The laws permit a unified description of the diverse behaviours of the system and allow the prediction of its behaviour given the states of its components. In this direction perhaps the most progress has been achieved in identifying the properties and dynamics of single neurons. This dates back to the works of Cajal who first used Golgi's staining technique to label individual nerve cells and developed the neuron doctrine — the principle that the nervous system is made up of many individual signalling elements, the neurons. Since then, the working principles of single neurons in isolation have been increasingly well understood, revealing more and more of the biochemical processes involved.

At the same time, when we consider higher brain functions, which are assumed to be generated from the interactions and concerted signalling of the vast number of neurons, we yet cannot formulate first principles with a similar degree of predictive power. One reason for this is that the brain is far from being decomposable into smaller subsystems which are only weakly interacting – usually considered a prerequisite for any feasible reductionist approach. There are about 10^{11} neurons in the brain, of which most cells make contacts with about 10^3 to 10^4 other neurons (Kandel, Schwartz, & Jessel 1991). This establishes a high degree of connectivity which also spreads over quite remote brain regions. There is also strong indication from recent brain imaging techniques that most mental activities in behaving subjects involve the simultaneous activation of many brain areas. Therefore any reductionist approach faces severe difficulties in linking low level descriptions and high level function, since it is not yet clear what could serve as an adequate subsystem, that actually allows a simpler and concise explanation. Although certainly no neuroscientist would disagree, that neurons provide such a subsystem, the step from single neurons to interacting networks results in an enormous increase in complexity. This can be seen from the experimental evidence, that while a single neuron in vitro shows regular and predictable activity, a similar neuron in vivo exhibits highly irregular and chaotic activity (Holt, Softky, Koch, & Douglas 1996).

The *binding problem* is an important example for our current lack of understanding the principles of brain function on the large-scale system level. Evidence from neurophysiology and neuroanatomy indicates that processing streams in the visual system are divergent, so that feature dimensions such as colour, motion, location and object identity are processed along different pathways, located in separate brain regions. This has raised the question, how the brain is able to combine or *bind* these remote neural representations, since any sensory, cognitive, or motor process must result from parallel interactions between large numbers of neurons, distributed over multiple cortical structures. The term "binding problem" then addresses the question, how this integration process could be implemented by the neural interactions in the brain. Since its clearest formulation as a general problem of brain theory with far-reaching implications by von der Malsburg (1981) there has been an ongoing controversial discussion on the necessity for such a binding process.

The *temporal correlation hypothesis* was suggested as a central binding mechanism in the brain (Milner 1974; von der Malsburg 1981). According to this hypothesis a binding between remote neural populations is represented by their synchronized activity. Apart from theoretical considerations on the advantages of such a mechanism, the concept has gained much support from experimental evidence on the presence of synchronized activity in the visual cortex and other brain regions, recently reviewed by Singer (1999). Nevertheless, others have heavily refuted the correlation hypothesis by arguing against a necessity of binding in general and the plausibility of its underlying physiological mechanisms. Also this opposite view is based on substantial experimental evidence which demonstrates the absence of synchronization effects in certain perceptual tasks, reviewed by Shadlen & Movshon (1999). Both supporters and opponents agree that a conclusive experimental test of the correlation hypothesis has not yet been done, since all recent experiments have assessed the question only indirectly. A decisive experiment would have to demonstrate either the presence or the absence of a clear causal relation between the perceptual impression, e.g. measured by behavioural performance, and the synchronization of the activated neurons. Since such an experimental design is at the limits of the currently available techniques, the issue of binding must be until now considered as unresolved.

The concept of binding is not necessarily tied to the correlation hypothesis and there may also exist other mechanisms to represent binding in the cortical networks. *Spatial coding* is an important and well-established principle of representation of sensory information in the brain. The most prominent example are cortical feature maps (see Ritter, Martinetz,

& Schulten 1992 for a comprehensive introduction), where neurons are topographically ordered according to their response properties such that the position of neural activity in the map carries relevant information. The layered organization of the cortex plays an important role for this spatial representation which is usually two-dimensional in nature. Due to the relevance of spatial processing in the brain, Ritter (1990) has suggested the principle of *spatial feature binding* as an alternative binding mechanism, where binding is represented by the spatial coactivation of feature-detecting neurons in a system of layers with topographically organized competitive and cooperative interactions. The exploration of this concept is the topic of this thesis.

1.1 Scope and Goals

The goal of this thesis is to contribute to the understanding of feature binding processes by investigating an artificial neural network model for spatial feature binding, the competitive layer model (CLM). The CLM is a recurrent neural network, which employs topographically structured competitive and cooperative interactions in a system of neural layers to represent binding by the layer-wise coactivation of feature-representing neurons. This model is explored by means of mathematical analysis and simulations on artificial and real-world data.

A special focus lies on the issue of neurodynamics. The dynamics of the CLM is formulated as an additive recurrent neural network with linear threshold neurons, which is a standard model for rate-coded activity dynamics in biological neural networks. The special dynamical form facilitates a deep mathematical analysis with respect to the attractors and the dynamics by using methods of eigensubspace analysis. The results show that the binding dynamics can be interpreted as a symmetry-breaking process in a multi-stable system, and therefore a thorough characterization of the conditions is given, which allow for such behaviour in arbitrary linear threshold networks. The simulation results show that the full capabilities of the CLM can be best exploited by operating close to the stability margins implied by these conditions.

Another important topic is the implementation of perceptual grouping according to Gestalt laws of visual perception. This is a main target of most current models of feature binding, however, the application to real-world image data is still rather exceptional due to the high computational demands of synchronization-based models. The aim of the CLM binding model for Gestalt perception is to demonstrate that the feature binding concept as a functional model can serve as a competitive approach to challenging real-world image segmentation problems. The application is largely facilitated by the thorough theoretical analysis of the model, which also provides explanations for the dynamical mechanisms which are responsible for the observed results.

The applications to Gestalt-based grouping show that interesting results can be obtained by heuristically choosing the explicit connection structure in the CLM. Most other recent feature binding approaches to segmentation are based on a similar heuristic design of the lateral interactions for a particular grouping problem under study. Nevertheless, there are many applications, where it is desirable to adapt or learn the connections which guide the binding process directly from the data which is to be segmented. In this manuscript a new approach based on supervised learning is explored, where a set of manually segmented training patterns is used to adapt the connection structure in the neural model. The goal of the presented learning method is to provide a means of choosing appropriate lateral interactions in complex feature spaces where the heuristic design approach becomes infeasible.

One aspect of perceptual grouping is that the interpretation of a visual scene is performed under the physical constraints of illumination, shading, occlusion, transparency, and various other factors. This raises the question how a binding process can achieve a solution to such multiple constraint satisfaction problems. Classical backtracking algorithms have been successfully applied to problems of this type, but their recursive nature does not allow a direct interpretation in terms of neural circuits. On the other hand, deterministic annealing provides a solution method to constraint satisfaction based on neural networks. A drawback of deterministic annealing is the lack of a mechanism for a later revision of partially violated constraints in the final network output. Therefore, in the last part of the thesis the new backtracking deterministic annealing algorithm is presented, which combines the effective search strategy of backtracking within a consistent neurodynamical model based on deterministic annealing.

1.2 Plan of the Manuscript

The first chapter motivates the issue of feature binding as one of the important questions regarding our understanding of brain function and defines the scope and goals of this thesis in relation to this concept.

The second chapter gives an overview of the controversial scientific discussion of the binding problem and reviews the different types of neural network model approaches to binding. The controversy on the necessity for a binding process in the brain is presented in terms of the two contrary concepts of convergent hierarchical coding and temporal correlation and their main arguments are summarized. Gestalt-based perceptual grouping is discussed as an important target of feature binding models. The review of the different model approaches distinguishes between temporal correlation and energy-based labeling models and focuses on their application in sensory segmentation and grouping applications.

In the third chapter some new theoretical results for conditions on the synaptic interactions are established which ensure the stability of linear threshold recurrent networks. The conditions provide a regime, where linear threshold networks are at the same time sensitive to small changes in their inputs and are capable of strong recurrent amplification without runaway activity. The results are of particular relevance to the competitive layer model, since they allow to operate the model in a mode of strong contextual activity modulation which improves the overall grouping quality. A reader mainly interested in feature binding and segmentation applications may skip the chapter at first reading, since the mathematical treatment is rather general and makes no particular reference to grouping.

In Chapter 4 the competitive layer model is introduced as a recurrent neural network architecture for feature binding, sensory segmentation and labeling problems. First the structure and dynamics of the model and the basic topographic competitive and cooperative interactions are described. A deterministic annealing mechanism is introduced, which is compared to related Potts-spin mean field models. In the theoretical analysis section the stability conditions of Chapter 3 are applied to ensure convergence of the CLM, and additional conditions are proved, which guarantee the Winner-Take-All behaviour of the competitive columnar interactions. The grouping dynamics is characterized in relation to the lateral interactions by performing an eigensubspace analysis. The efficient simulation on standard serial computers is described in terms of an explicit algorithmic formulation. Finally, the lateral coupling scheme of the competitive layer model is generalized by also allowing interlateral interactions between layers. With this generalization, the CLM can be applied to general labeling problems, and the chapter concludes with a comparison to other labeling approaches.

Chapter 5 presents the application of the CLM to model a wide range of Gestalt-based perceptual grouping laws. First, the grouping of contours according to the principle of continuity is considered by using oriented edge elements as features. A mechanism for combined figure-ground segmentation and grouping is introduced and the dependency of the grouping performance on the relevant model parameters is explored. The results are discussed for real images and compared to the results of a Potts model with same lateral interactions. An extension of the contour grouping approach to the task of cell segmentation is described. Other grouping principles which are considered include motion grouping, greyscale segmentation, and texture segmentation.

In Chapter 6 methods are presented to obtain appropriate lateral interactions for grouping and segmentation by learning processes from the feature input data. After a review of related approaches, new methods are developed which perform supervised learning form a set of manually labelled training patterns. The methods are evaluated on an artificial data set and cell images from fluorescence microscopy. A different learning approach is presented in Chapter 7, which aims at optimizing the lateral connection structure in order to improve the capabilities for solving complex constraint satisfaction problems. The proposed backtracking deterministic annealing method is interpreted as a heuristic approach to combine classical backtracking with neural deterministic annealing and successfully applied to the problem of complex tiling problems.

The concluding Chapter 8 summarizes the main results and discusses possible future research directions.

Chapter 2

Feature Binding and Perceptual Grouping

This chapter gives an overview of the controversial scientific discussion of the binding problem and reviews the different types of neural network model approaches to binding. The controversy on the necessity for a binding process in the brain is presented in terms of the two contrary concepts of convergent hierarchical coding and temporal correlation and their main arguments are summarized. Gestalt-based perceptual grouping is discussed as an important target of feature binding models. The review of the different model approaches distinguishes between temporal correlation and energy-based labeling models and focuses on their application in sensory segmentation and grouping applications.

2.1 Concepts of Binding

The binding problem has received widespread attention among theorists and experimentalists in the recent years. This considerable interest was triggered by the realization that there are few, if any, places in the nervous system where all the information to carry out a particular task is localized. This means that any sensory, cognitive, or motor process must result from parallel interactions between large numbers of neurons, distributed over multiple cortical structures. The concept of binding in its general sense addresses, how this distributed activity is integrated towards a coherent processing and perception of single entities. Perceptual grouping according to Gestalt laws is a specific example in sensory processing, where binding has been considered to provide a particularly elegant solution.

2.1.1 The General Binding Problem – A Controversy

Consider the following simple example, which is due to Rosenblatt (1962) (see Figure 2.1). Suppose there are two neurons for which one is active when a square appears somewhere in the visual field and the other is active if a triangle is present. A second pair of neurons is activated if an object is located in the top or bottom of the visual field, respectively. Suppose a particular task has to be carried out when the triangle is in the top position. If only single objects are presented, this poses no problem, since then this situation can be discriminated from the neurons' activations. If, however, both objects are present, it is not clear which of the objects is in the top position, since all neurons are activated. This is an example for a "superposition catastrophe", according to von der Malsburg (1981), where the information

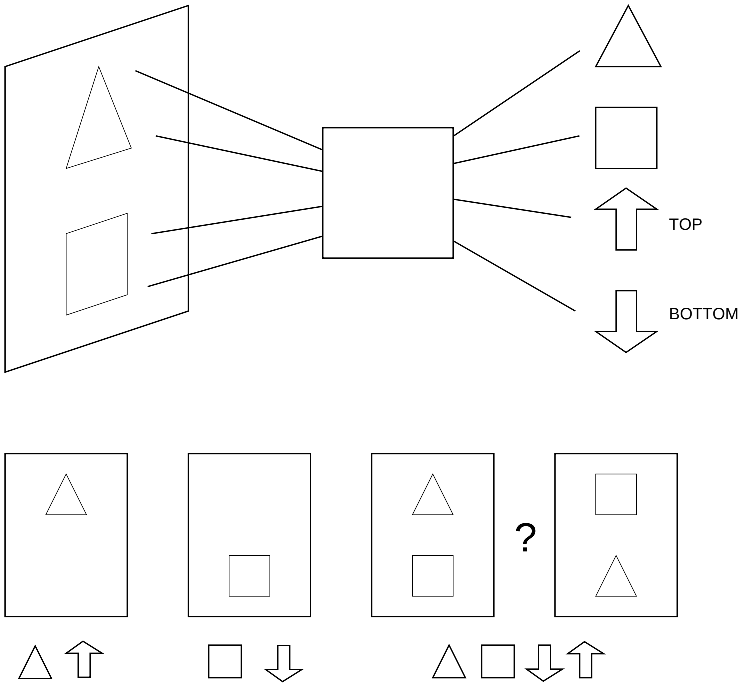


Figure 2.1: Rosenblatt's example. There are four neurons which individually code for the attributes {triangle,square,top,bottom}. If both triangle and square are present, it cannot be decided on the activations alone which is at the top. This requires a binding between the object type and the location attribute (after von der Malsburg 1999 and Rosenblatt 1962).

on the feature conjunctions is lost if no additional binding mechanism is present. This raises the question, which solution strategy is adopted by the cortical organization to deal with this problem. In the following the two contrary concepts of hierarchical coding and temporal correlation are discussed.

Convergent Hierarchical Coding

The concept of convergent hierarchical coding assumes that the brain is organized in hierarchical processing stages, where each stage performs specialized, parallel operations that depend on input from earlier stages. The hierarchical processing scheme in turn leads to neural representations which capture increasingly complex feature combinations, up to the so-called "grandmother cell", which fires only if a specific object is being recognized, perhaps even under specific viewing conditions. This concept is also known as the "neuron doctrine" of perception, postulated by Barlow (1972, 1985). It has been argued that this processing scheme avoids any binding problem, because the strong convergence of information performs the integration of segregated sensory channels. There is in fact substantial experimental evidence in favour of this notion of hierarchical processing. Proceeding from the neurons which receive the retinal input to neurons in higher visual areas, an increase in receptive field size and dependence on increasingly complex stimuli can be observed. A number of experiments have also identified neurons which respond to object-like stimuli (Tanaka 1993; Tanaka 1996; Logothetis & Pauls 1995).

The main criticism against hierarchical coding is, that it leads to a combinatorial explosion of the possibilities which must be represented, due to the large number of combinations of features which make up a particular object under different viewing conditions (von der Malsburg 1999; Gray 1999). This problem can be approached by using population or coarse coding, where a particular stimulus pattern, corresponding to an object, is represented by a distributed pattern of activity, which itself has a larger combinatorial capacity. This scheme, however suffers from the superposition problem outlined before, where multiple activations may exhibit strong interference and cross-talk without a binding mechanism.

A number of arguments were brought up in order to defend the neuron doctrine against this criticism. Ghose & Maunsell (1999) argue that the combinatorial explosion can be avoided due to the fact that the crucial number is not the number of objects which can be recognized with all position and lighting conditions, but the number of objects which can be distinguished. This number which has been estimated at about 100000 (Biederman 1987) may be far below the number of neurons in the visual cortex. Therefore, in representing only a finite set of objects and attributes, there could be small specialized groups of neurons for each discriminable object or attribute. This point of view was elaborated into an explicit model by Riesenhuber & Poggio (1999). Their model is capable of view-invariant recognition of artificial paper-clip objects in cluttered surroundings. It is based on a hierarchy from simple to complex to hypercomplex cells, which code for combinations of line elements with increasing complexity and receptive field size. The basic concept to avoid a superposition problem as outlined previously, is to allow each complex cell to focus on the input of the presynaptic cell providing the strongest input. This maximum selection operation effectively isolates features representative for an object from the surrounding clutter. A related mechanism of maximum selection was suggested by Reynolds & Desimone (1999) which propose that it is guided by attentional mechanisms.

The Temporal Correlation Hypothesis

The temporal correlation hypothesis (von der Malsburg 1981) postulates that spatially segregated neural activations are selectively bound together by the synchronized activity of the corresponding neurons. This offers an elegant approach to avoid the superposition problem, and introduces the concept of rapid assembly formation, which can be interpreted as flexible means of population coding. For Rosenblatt's example this means that instead of requiring an explicit neuron which represents the common activation of the attributes "top" and "triangle", this is represented by the synchronized activity of neurons for the single attributes. This synchronized neuron group can then selectively activate later processing stages. Von der Malsburg has emphasized the dynamical nature of this binding process and its advantages for implementing learning mechanisms in the brain. While for the combination coding cells an instance for each combination must exist, in temporal binding a concept can be rapidly formed and represented by a synchronized assembly. This offers advantages with respect to plasticity and generalization, when confronted with new situations (von der Malsburg 1999).

The presence of correlated and synchronized neural activity has been established in the recent years by numerous experiments. As Usrey & Reid (1999) have pointed out in their review, the range of such phenomena is very broad, taking place at different time-scales and spatial ranges, however, their significance is still an unsettled question. For example, strong correlation between the activity of two neurons can be observed, if they are both driven by a common presynaptic neuron. This is a simple consequence of the phenomenon of anatomical divergence, and not the type of correlation which realizes emergent assembly formation (Usrey & Reid 1999). The main experimental evidence that is brought in favour of the correlation hypothesis concerns stimulus-dependent synchronization of particular neurons in the visual cortex of the cat and the monkey, measured with multi-electrode recordings. Kreiter & Singer (1996) showed that cells in the middle temporal (MT) area, associated with motion processing, show selective synchronization if the cells are stimulated by a single bar stimulus and no synchronization if the cells are stimulated by two coherently moving, but separate bars. Engel, König, & Singer (1991) demonstrated that cells in striate cortex form separate synchronous assemblies when two drifting bars are presented together but move in orthogonal directions over the same region of the visual field. Similar experimental paradigms were investigated on a lot of studies (reviewed by Gray 1999 and Singer 1999).

Criticism on the correlation hypothesis has been addressed towards the stringency of its theoretical concept and the plausibility of the postulated underlying neurophysiological processes. The supporters of the "neuron doctrine" have expressed their doubts whether there is any binding problem, since the hierarchical coding concept can circumvent this problem, as discussed in the previous section. Shadlen & Movshon (1999) have criticised the significance of correlated firing on physiological grounds, and reviewed a range of experimental evidence where no significant synchronization was observed. They emphasize the point that conclusive evidence for the correlation hypothesis will only be based on a demonstration of the clear causal relation between synchronized neural state and the perceptual impression, measured by some behavioural task.

2.1.2 Sensory Segmentation and Gestalt Perception

The segmentation problem in perception is an important application where binding has been proposed to provide particularly elegant solution strategies. One of the main motivations for postulating binding processes in sensory processing is based on the huge amount of psychophysical experimental data, which demonstrates the presence of grouping phenomena in vision or other sensory modalities. The foundations of this research were laid by the Gestalt school of psychology, founded in the early twenties of this century. The Gestaltists investigated and formulated the phenomenological laws of proximity, similarity, continuity, closure, and common fate (see Figure 2.2), which describe how simpler visual features interact to generate larger percepts (Wertheimer 1923; Wertheimer 1938; Köhler 1929; Koffka 1935). They proposed that perceptual grouping is a largely automatic process, which precedes higher perceptual stages like object recognition. This notion of *preattentive* grouping was supported by experiments, which showed that under certain conditions the time required to identify a target in a display did not depend on the number of items in the display (Neisser 1963). The conclusion was that grouping is not a serial process mediated by focused attention, but is to a large extent carried out in parallel to provide an initial segmentation of a visual scene into its component objects on the basis of Gestalt properties (Neisser 1967). This notion was later supported by numerous studies on preattentive grouping processes (Julesz 1981; Pomerantz 1981; Treisman 1982; Bergen & Adelson 1988; Elder & Zucker 1993).

Preattentive grouping has been suggested as a main target for feature binding for a number of reasons. It is a very flexible process, which takes into account multiple simultaneous grouping cues across different feature dimensions. If different grouping laws compete for alternative interpretations of a scene, those which provide the simplest "explanation" of a feature configuration will dominate; this is the Gestalt concept of *Prägnanz.* Since the final perceived grouping depends heavily on the whole feature arrangement, and the number of possible feature combinations seems effectively infinite, the combinatorial capacity of

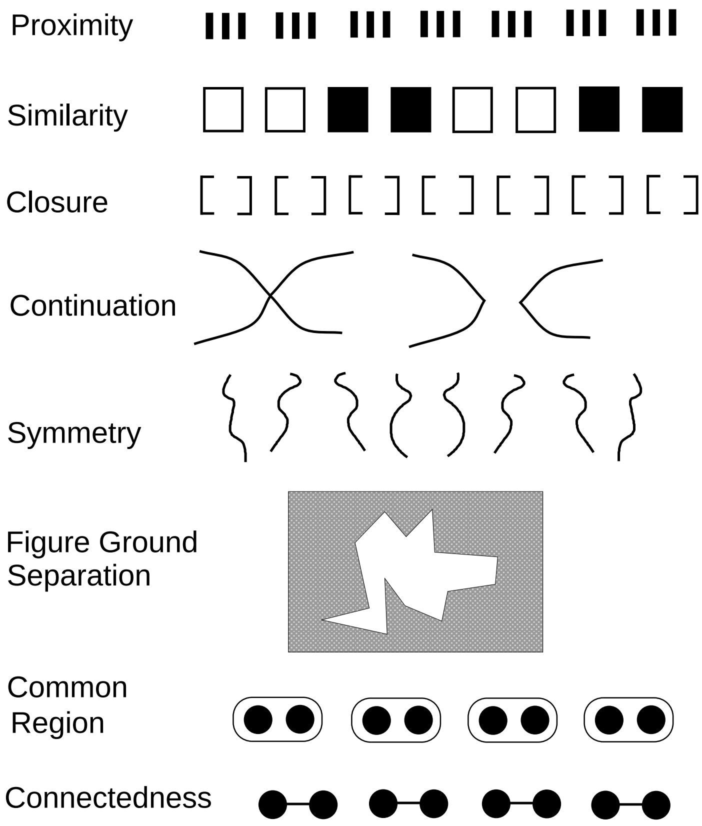


Figure 2.2: Table of Gestalt laws. The Gestalt laws express a phenomenological description of the principles which guide the grouping of simple visual stimuli into larger percepts (after McCafferty 1990 and Rock & Palmer 1990).

perceptual grouping must be very high. This has led the supporters of binding to reject convergent hierarchical coding as an appropriate model for perceptual grouping, since it may suffer from a combinatorial explosion of the number of cells required (von der Malsburg 1981; Gray 1999). On the contrary, the rapid assembly formation due to feature binding could provide these complex grouping mechanisms.

Key mechanisms of perceptual grouping have been attributed to the rich horizontal or lateral connection structure observed in the visual cortex. Prominent examples are the excitatory connections linking orientation sensitive cells of similar stimulus preference (Gilbert 1992; Rockland & Lund 1983; Weliky, Kandler, Fitzpatrick, & Katz 1995). Interestingly enough, before experimental evidence on such connections was available, the theoretical advantages of binding in the brain have led von der Malsburg (1981) to postulate similar horizontal connections between cells with similar receptive field properties. These lateral interactions are assumed to influence the synchronization or desynchronization of neurons or neuron pools representing a bound group of features and are an essential component in almost all models of feature binding for sensory processing.

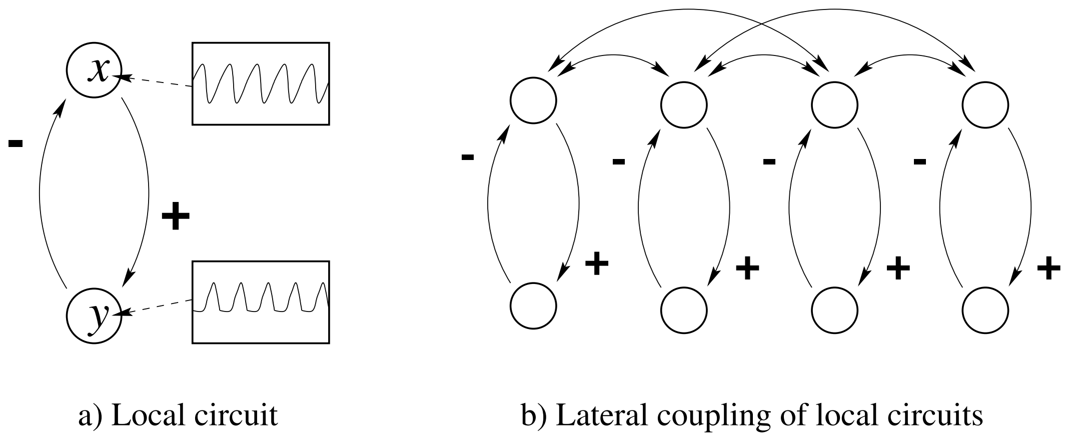


Figure 2.3: Activity-Based Oscillator Models. a) Sketch of a local circuit which consists of an excitatory and an inhibitory neuron to produce nonlinear oscillations. b) Lateral interactions between local modules, corresponding to individual features, cause synchronized or desynchronized oscillations to represent a binding of the features.

2.2 Synchronization Models

The experimental support for synchronous activity in the visual cortex raised a lot of scientific interest in models of synchronized activity in the recent years. There is now a variety of modeling approaches, differing mainly on the level of abstraction from biology. Models of spiking neurons provide a very fine-grained level of detail with respect to the time-resolved activity of neurons, to discuss the plausibility and qualitative nature of neural synchronization in close connection to the known neuronal physiology (Gerstner, van Hemmen, & Cowan 1996). Since these models rather focus on the intrinsic physiological processes than on performing specific grouping tasks,[1] their discussion is beyond the scope of this thesis. The application of synchronization-based binding in models of perceptual grouping and segmentation, however, has also led to stronger abstractions from biology, to facilitate the large-scale simulation and the theoretical analysis of the resulting neural networks.

2.2.1 Activity-Based Oscillator Models

A large class of synchronization-based models uses local activity-based oscillatory circuits which are coupled by long-range horizontal interactions to facilitate the synchronization of the single oscillatory modules. The generic form of the local circuit (see Figure 2.3) is given by an excitatory and an inhibitory unit with activities denoted by x and y respectively, coupled according to the following nonlinear dynamics

$$\begin{aligned} \dot{x} &= F(x,y) + I^x, \\ \dot{y} &= G(x,y) + I^y, \end{aligned} \tag{2.1}$$

where I^x and I^y provide external inputs. The nonlinear functions F and G satisfy $\partial F/\partial y < 0$ and $\partial G/\partial x > 0$ which ensures that the excitatory unit exerts positive coupling onto the inhibitory unit and the inhibitory unit gives negative feedback. Depending on the particular

[1] An application of spiking neurons to associative memory was investigated by Ritz, Gerstner, Fuentes, & van Hemmen (1994).

form of the nonlinearities, equations (2.1) give rise to diverse types of oscillatory behaviours ranging from smoothly varying sinusoidal to rapidly switching relaxation oscillations (Wilson & Cowan 1972; Li 1998; Terman & Wang 1995).

While a single module of this type corresponds to a particular sensory feature, the binding of features is represented by the synchronized oscillation of multiple modules. Synchronization is facilitated by lateral horizontal interactions between modules which introduce a coupling of local oscillators of the form

$$\begin{aligned}\dot{x}_r &= F(x_r, y_r) + F_r^{\mathrm{lat}}(\mathbf{x}, \mathbf{y}, \mathbf{f}) + I^x,\\ \dot{y}_r &= G(x_r, y_r) + G_r^{\mathrm{lat}}(\mathbf{x}, \mathbf{y}, \mathbf{f}) + I^y,\end{aligned} \tag{2.2}$$

where $F_r^{\mathrm{lat}}(.)$ and $G_r^{\mathrm{lat}}(.)$ represent the effects of lateral coupling. The activities of the local circuits for feature $r = 1, \ldots, N$ are denoted by x_r and y_r; $\mathbf{x} = (x_1 \ldots x_N)$ and $\mathbf{y} = (y_1 \ldots y_N)$ denote the vectors of all activities, and $\mathbf{f}$ is a vector of weights which characterizes the lateral interactions. In some models these weights are also time-dependent (von der Malsburg 1981; von der Malsburg & Schneider 1986).

The early work on such nonlinear oscillator models for feature binding considered local (von der Malsburg & Buhmann 1992) and global (Schillen & König 1994) connection schemes, but simulations were only carried out on small networks and highly simplified test images. Later it was shown, that sinusoidal oscillators have limitations to support fast and long-range synchronization with local interactions (Terman & Wang 1995; Somers & Kopell 1995). This has led to the development of the LEGION model which uses relaxation oscillators to achieve fast synchronization and desynchronization with local excitatory and global inhibitory coupling schemes (Terman & Wang 1995; Wang & Terman 1997). While Wang & Terman show that the consistent dynamical model of the type (2.2) can stably represent only up to 5 segments, they also suggest an algorithmic abstraction of the original model which does not suffer from this limitation. The capabilities of the LEGION model when confronted with real-world sensory processing tasks were successfully demonstrated for greyscale image segmentation (Wang & Terman 1997; Shareef, Wang, & Yagel 1999), range image segmentation (Liu & Wang 1999) and auditory processing (Wang & Brown 1999).

2.2.2 Phase Models

Phase models provide an even stronger abstraction from the explicit time-dependent neural activity by assuming that the oscillatory behaviour of a local neural circuit can be entirely described by a single phase variable (see Figure 2.4). The coupling between oscillators is then expressed by formulating a dynamics for the phase variables, that takes their interactions into account. An example is given by

$$\dot{\theta}_r = \omega_r + \eta_r + \sum_{r'} \mathsf{f}_{rr'} \sin(\theta_r - \theta_{r'}), \tag{2.3}$$

where θ_r denotes the phase of feature r, ω_r is the constant oscillation speed of the local oscillator, η_r provides a noise term to desynchronize separate groups, and $\mathsf{f}_{rr'}$ denotes the lateral coupling between features r and r'. The binding of two features is represented by a sufficiently small difference in their phase values.

Baldi & Meir (1990) have investigated the application of the dynamical phase model (2.3) to texture-based figure-ground segmentation. They show that two segments, corre-

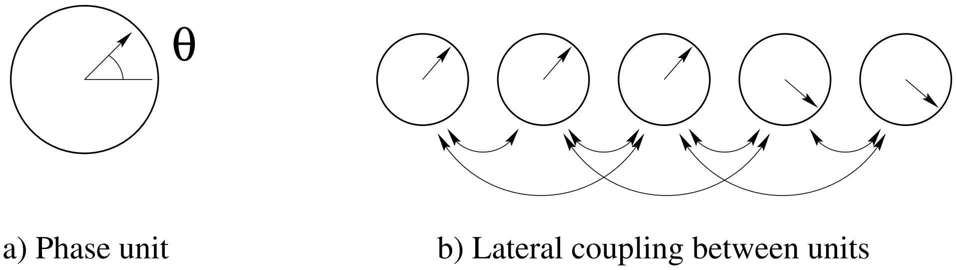


Figure 2.4: Phase Models. a) The oscillatory state of a local circuit is completely characterized by a single phase variable θ. b) The binding of features is represented by groups of oscillators separated according to their phase variables.

sponding to figure and ground can be stably represented. Sompolinsky, Golomb, & Kleinfeld (1991) performed a thorough theoretical analysis of the model (2.3). A certain drawback of these models is that robust synchronization can only be achieved by requiring all-to-all connections between units, which has limited their application to small and synthetic examples.

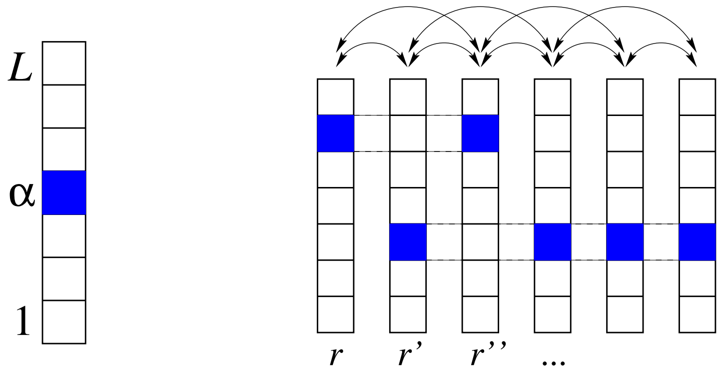
Gislén, Peterson, & Söderberg (1992) and Zemel, Williams, & Mozer (1993) proposed a model of a network of stochastic directional units, for which the mean field approximation gives rise to the coupled dynamics of units characterized by phase and amplitude. The units are represented as complex variables with imaginary and real part which allows to formulate an energy function for the dynamics which is an extension of the Hopfield model to complex weights and activation variables. The application of the model to feature binding was investigated by Mozer, Zemel, Behrmann, & Williams (1992), who also suggested a recurrent learning approach to feature binding.

Fellenz (1997) and Fellenz & Hartmann (1995) have considered a model for contour-based image segmentation by a phase diffusion process, where the dynamics (2.3) is augmented by additional nonlinearities. The output is a phase distribution across the image which is continuously varying across smooth contours, and has discontinuities at object borders. This approach, however, can be only used to obtain boundaries between segments since it requires an additional segmentation step on the phase variables and offers no clear representation of a grouping with distinct segments.

The contour integration model of Yen & Finkel (1998) uses a phase-averaging rule between local groups of interconnected neural oscillators to compute contour salience. They discuss a significant preference of their model towards contour closure in relation to the psychophysical study of Kovács & Julesz (1993) and also present results for real images. A certain drawback is that the model requires rule-based interactions and mainly focuses on figure-ground segmentation, without an explicit representation of multiple groupings.

2.3 Energy-Based Labeling Models

Instead of using a continuous-valued phase variable, the state of a local neuron group may also be described by a finite number of labels. These labels may be identified with particular phase-intervals of a local oscillator, since most oscillator models have only a limited capac-



a) Label unit b) Lateral coupling between units

Figure 2.5: Labeling Models. a) The binding state of a local feature unit is represented by the assignment of a particular label α from a set of L labels. b) The binding of features can be represented by groups sharing the same labels.

ity of 5-8 separable channels. A binding of features is then represented by both features sharing the same label. It should be noted that this makes the readout of multiple bindings much easier than in a continuous phase model, where additional problems are posed by phase shifts and the necessity to define arbitrary tolerance intervals for group discrimination.

The methods that are discussed in this section all share a common approach of solving labeling problems by the minimization of a related energy function. In this framework contextual constraints are represented by compatibilities between label assignments of different features. Apart from modeling feature binding by interpreting equal label assignments as binding, the general labeling approach, however, also covers more general problems like clustering and constraint satisfaction problems. The compatibilities may also extend beyond pairwise interactions and include higher order relations between three and more features like in the case of Markov random fields.

2.3.1 Relaxation Labeling

Relaxation labeling denotes a family of iterative procedures which have become a standard technique in pattern recognition and computer vision domains. They were pioneered by Rosenfeld, Hummel, & Zucker (1976) and have since then found applications in a variety of different problems (see Kittler & Illingworth 1985 for a review). The task of relaxation labeling is to solve the following labeling problem:

Suppose we are given a set of N features indexed by $r = 1, \ldots, N$. Each feature r can be labelled by one of a set of L labels indexed by $\alpha = 1, \ldots, L$. Find a labeling which embodies contextual knowledge in an optimal way, where the contextual information is given by a set of pairwise compatibilities $\mathsf{f}_{rr'}^{\alpha\beta}$ between assigning label α to feature r and label β to feature r'.

The coefficients $\mathsf{f}_{rr'}^{\alpha\beta}$ may be derived by heuristic arguments, statistical considerations or learning (Peleg & Rosenfeld 1978; Pelillo & Refice 1994). A distinction can be made if the compatibilities do not depend on the explicit label indices α and β, i.e. $\mathsf{f}_{rr'}^{\alpha\beta} = \mathsf{f}_{rr'}$ for all α, β. This is the generic case for the application to *grouping*, since the labels are then only used to represent the membership of the features in the same group and carry no additional information.

If we define $x_{r\alpha} \geq 0$ as the certainty of the assignment of label α to feature r, the space of weighted labelings is defined by the condition $\sum_\alpha x_{r\alpha} = 1$. The task of a relaxation labeling algorithm is then to find an unambiguous and consistent labeling, that is

$$x_{r\hat{\alpha}(r)} = 1, \quad x_{r\beta} = 0, \quad \text{for all} \quad r, \beta \neq \hat{\alpha}(r) \;, \tag{2.4}$$

where $\hat{\alpha}(r)$ denotes the unambiguously assigned label to feature r, and consistency of the unambiguous labeling is defined as

$$F_{r\hat{\alpha}(r)} > F_{r\beta}, \quad \text{for all} \quad r, \beta \neq \hat{\alpha}(r), \quad \text{where} \quad F_{r\alpha} = \sum_{r'\beta} \mathsf{f}^{\alpha\beta}_{rr'} x_{r'\beta} \;. \tag{2.5}$$

The $F_{r\alpha}$ are called the linear support functions which accumulate the pairwise contributions of all other certainties weighted with compatibility.

For symmetric compatibilities with $\mathsf{f}^{\alpha\beta}_{rr'} = \mathsf{f}^{\beta\alpha}_{r'r}$ conditions (2.4) and (2.5) are necessary and sufficient for a local minimum of the negative average local consistency (Hummel & Zucker 1983)

$$E = -\sum_{r\alpha} x_{r\alpha} F_{r\alpha} = -\sum_{r\alpha} \sum_{r'\beta} \mathsf{f}^{\alpha\beta}_{rr'} x_{r\alpha} x_{r'\beta}. \tag{2.6}$$

A relaxation labeling algorithm starts with an initial weighted labeling assignment $x_{r\alpha}(0)$ and iteratively adjusts it to obtain a local minimum which provides a consistent solution. Rosenfeld, Hummel, & Zucker (1976) proposed the iteration scheme

$$x_{r\alpha}(t+1) = \frac{x_{r\alpha}(t) F_{r\alpha}(t)}{\sum_\beta x_{r\beta}(t) F_{r\beta}(t)}, \tag{2.7}$$

where the coefficients $\mathsf{f}^{\alpha\beta}_{rr'}$ are required to be positive. Hummel & Zucker (1983) have suggested and analyzed an alternative iteration algorithm which poses no restriction on the compatibilities, but requires a computationally expensive projection operator to confine the state space to proper weighted labelings.

Conventional relaxation labeling algorithms as that of Hummel & Zucker and Rosenfeld et al. converge to local minima of the energy and achieve good results, if it is either not necessary to find a global optimum, or the corresponding energy function has not too many local minima. Parent & Zucker (1989) have demonstrated the application of relaxation labeling to the inference of local curvature information in noisy and complex images. In their model, for each image location r the labels α represent curvature classes and orientations, therefore it is a model for contour integration and not for contour grouping. The coefficients $\mathsf{f}^{\alpha\beta}_{rr'}$ characterize their mutual compatibility, which is positive when the local orientations allow for good continuation. If the certainties are initialized with a broad estimate of the local orientation (Zucker, David, Dobbins, & Iverson 1988), only few iterations are necessary to arrive at a globally consistent interpretation of the image.

2.3.2 Spin Models

Spin models provide an approach to labeling problems, that is conceptually related to relaxation labeling by a very similar energy function. The focus is, however, generally shifted towards a greater interest in obtaining global minima of the corresponding energy. In the framework, as it was introduced in the previous section, the different labels $\alpha = 1, \ldots, L$ correspond to different binary states $x_{r\alpha} \in \{0, 1\}$ of a Potts spin variable at site r. Since the spin is assumed to be in a defined state, the variables satisfy $\sum_\alpha x_{r\alpha} = 1$. In the following the notation $\hat{\alpha}(r)$ is used to denote this active spin. The compatibilities $\mathsf{f}_{rr'}^{\alpha\beta}$ describe the coupling strength of two spin states $x_{r\alpha}$ and $x_{r\beta}$. The energy of the Potts system (Wu 1983) is then given by

$$E = -\frac{1}{2} \sum_{r\alpha} \sum_{r'\beta} \mathsf{f}_{rr'}^{\alpha\beta} x_{r\alpha} x_{r'\beta} - \sum_{r\alpha} g_{r\alpha} x_{r\alpha}. \tag{2.8}$$

The additional linear components weighted with $g_{r\alpha}$ correspond to external fields in the physical interpretation. If we choose $g_{r\alpha} = 0$, then this is identical to the quadratic relaxation labeling energy (2.6).

If the spins are in thermal equilibrium, then the probability P of a spin configuration $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}(1), \ldots, \hat{\alpha}(N))$ is given by the Boltzmann distribution

$$P(\hat{\boldsymbol{\alpha}}) = \frac{1}{Z} e^{-E(\hat{\boldsymbol{\alpha}})/T}, \tag{2.9}$$

where Z is a normalizing factor and T is the temperature. For sufficiently small T this distribution peaks around the minima of E. To achieve the goal of finding global minima of (2.8), many different approaches exist, which either try to solve the binary problem by a search in the configuration space, or turn the spins themselves into probabilistic and continuous variables.

Simulated Annealing

In simulated annealing (Kirkpatrick, Gelatt, & Vecchi 1983), the search is done by a controlled series of local updates, which is accompanied by a lowering of a global temperature parameter. Starting from an initial configuration $\mathbf{x}_t$, a new state $\mathbf{x}_{t+1}$ is randomly chosen, which obeys the assignment constraints. If the energy is lowered, due to $E(\mathbf{x}_{t+1}) - E(\mathbf{x}_t) < 0$, the system is directly updated to the new state. If the energy is increased, the update to the new state is done with a probability

$$p(\mathbf{x}(t) \to \mathbf{x}(t+1)) = \frac{1}{1 + \exp\left(\frac{E(\mathbf{x}_{t+1}) - E(\mathbf{x}_t)}{T}\right)}. \tag{2.10}$$

This allows for a possibility to escape from poor local minima. During iterations the temperature parameter T is lowered or "annealed"; if this annealing is sufficiently slow, then the probability to obtain the global minimum converges to 1 with increasing number of iterations. The main problem of this approach is its slow convergence, since usually very slow annealing is required to achieve satisfactory results.

Mean Field Annealing

In the mean-field annealing approximation to the system (2.9), the binary spin variables $x_{r\alpha} \in \{0, 1\}$ are replaced by their expectation values which are then continuous variables with $0 \leq x_{r\alpha} \leq 1$. Instead of minimizing the energy directly, the function to be minimized is the mean-field free energy (Peterson & Soderberg 1989) given by

$$E^F = -\frac{1}{2}\sum_{r\alpha}\sum_{r'\beta} \mathsf{f}_{rr'}^{\alpha\beta} x_{r\alpha}x_{r'\beta} - \sum_{r\alpha} g_{r\alpha}x_{r\alpha} + T\sum_{r\alpha} x_{r\alpha}\log x_{r\alpha}. \tag{2.11}$$

subject to the constraints $\sum_\alpha x_{r\alpha} = 1$. At high temperature T, the free energy is dominated by the convex entropy terms and has a unique minimum. This convex term serves to smooth out local minima, by gradually lowering T more and more structure of the original energy at $T = 0$ is recovered. A local minimum solution to (2.11) satisfies

$$x_{r\alpha} = \frac{\exp(-F_{r\alpha}/T)}{\sum_\beta \exp(-F_{r\beta}/T)}, \qquad \text{where} \quad F_{r\alpha} = \sum_{r'\beta} \mathsf{f}_{rr'}^{\alpha\beta} x_{r'\beta} + g_{r\alpha}. \tag{2.12}$$

A mean field algorithm then works by parallel or iteratively solving the mean field equations (2.12) together with a gradual lowering of the temperature T. Numerical studies have shown (Peterson & Soderberg 1989; Li, Soh, & Teoh 1998) that the mean field approach can obtain solutions to combinatorial optimization problems like the travelling salesman problem about 100 times faster than simulated annealing at a comparable solution quality. In image processing, the Potts mean field model has been applied to figure-ground segmentation (Herault & Horaud 1993) and boundary detection (Geman, Geman, Graffigne, & Dong 1990).

Cluster Updates

Cluster updates in Potts spin models have been considered for data clustering (Blatt, Wiseman, & Domany 1997) and image segmentation applications (Eckes & Vorbrüggen 1996; Vorbrüggen 1995; Opara & Wörgötter 1998). The main idea is to change several spin states simultaneously, to achieve faster convergence of the system to an equilibrium state.

In the following the ECU (energy-based cluster update) algorithm (Opara & Wörgötter 1998) is discussed in more detail, since it will be used for comparison in the chapter 5 on perceptual grouping applications section. It is assumed that the lateral interactions are of the form $\mathsf{f}_{rr'}^{\alpha\beta} = \delta_{\alpha\beta}\mathsf{f}_{rr'} - k\delta_{\alpha\beta}$, where $\mathsf{f}_{rr'} > 0$ and $k > 0$. The ECU algorithm consists of two steps which are iteratively repeated. In the cluster step, links are placed between pairs of spins for which $\mathsf{f}_{rr'} > 0$ with a probability

$$p(r, r') = 1 - \exp\left(\frac{1}{2T}\mathsf{f}_{rr'}\delta_{\hat\alpha(r)\hat\alpha(r')}\right), \tag{2.13}$$

where $\hat\alpha(r)$ denotes the single active spin with $x_{r\hat\alpha(r)=1}$ and $x_{r\beta\neq\hat\alpha(r)} = 0$, and T is a temperature parameter. Clusters are then generated by identifying all disconnected subgraphs in the set of links. Let $c_i = \{r_1^i, \ldots, r_k^i\}$ denote the set of k spins in cluster i. Now the energy of the subconfiguration is considered, if all the spins in cluster i are put into state β, which is given by

$$E_\beta^{c_i} = \sum_{r\in c_i}\Big(\sum_{r'} k\delta_{\beta\hat\alpha(r')} + \sum_{r'\notin c_i} \mathsf{f}_{rr'}\delta_{\beta\hat\alpha(r')}\Big). \tag{2.14}$$

The probability of changing the spins of this cluster into state β is then chosen as

$$P(c_i \to \beta) = \frac{e^{E_\beta^{c_i}/T}}{\sum_\alpha e^{E_\alpha^{c_i}/T}}. \tag{2.15}$$

As Opara & Wörgötter (1998) have proved, this update converges to a detailed balance at temperature T according to the Boltzmann distribution. They have tested the algorithm for greyscale segmentation and successfully applied it to real world images.

2.3.3 Markov Random Fields

Markov random fields are based on a global probability model, that can be completely specified in terms of local interaction or clique potentials. The stochastic formulation is similar to spin models, but is more general with respect to the interactions which may also include higher order terms beyond pairwise interactions. Another focus in Markov random fields is the proper modeling of a priori knowledge.

In a Markov random field the label state $\hat{\alpha}(r)$ of a feature is given by a random variable that is only dependent on the values of its neighbours, or more general of the members of a local clique. According to the Hammersley-Clifford theorem (Besag 1974) the labels are then distributed according to a Boltzmann distribution with

$$P(\hat{\boldsymbol{\alpha}}) = \frac{1}{Z} e^{-E(\hat{\boldsymbol{\alpha}})/T}, \tag{2.16}$$

where Z is a normalizing constant and T is a temperature parameter which controls the flatness of the distribution. The energy is given by

$$E(\hat{\boldsymbol{\alpha}}) = \sum_{c \in \mathcal{C}} V_c(\hat{\boldsymbol{\alpha}}), \tag{2.17}$$

where $\mathcal{C}$ is a set of cliques and V_c are the associated clique potentials. Usually only pairs of sites are considered, but also terms involving three or more sites are possible. In image interpretation we are interested in the a posteriori probability $P(\hat{\boldsymbol{\alpha}}|\mathbf{d})$ given the image data $\mathbf{d} = (d_1, \ldots, d_n)$ as observations for a set of images. If we assume conditional independence between the observations, which is an often used approximation (Chou & Brown 1990; Posch & Schlüter 1998), then the a posteriori energy is given by

$$E(\hat{\boldsymbol{\alpha}}|\mathbf{d}) = \sum_{c \in \mathcal{C}} V_c(\hat{\boldsymbol{\alpha}}) - T \sum_i \log P(d_i|\hat{\boldsymbol{\alpha}}). \tag{2.18}$$

Minimizing this a posteriori energy over the possible labelings results in a maximum a posteriori estimate of the random field and can be used to obtain a globally consistent interpretation of the image data. For the minimization of this energy algorithms such as stochastic relaxation (Geman & Geman 1984), regularization techniques (Terzopoulos 1986), or mean field annealing (Geiger & Yuille 1991) were considered. While those authors have focused on low-level image restoration and pixel-based segmentation, others have also shown the application of Markov random fields to modeling high-level grouping according to Gestalt laws of perceptual grouping (Krishnamachari & Chellappa 1996; Murino, Regazzoni, & Foresti 1996; Massmann, Posch, Sagerer, & Schlüter 1997; Posch & Schlüter 1998).

2.4 Summary

The binding problem is still a controversial issue in the field of neuroscience, since the question whether binding processes play a significant role for neural information processing in the brain cannot yet be decided on empirical grounds. Therefore, the investigation of neural network models for feature binding has become increasingly important over the recent years, to explore the advantages and usefulness of binding in sensory processing applications. The model approaches have centered around synchronization-based and labeling models. Synchronization is an elegant concept, because additional information in the temporal domain can be used to represent binding, without requiring additional processing elements. The complex temporal interactions, however, also pose problems in applications with respect to the stability and robustness of the desired groupings. A particular advantage of labeling models is that they generally possess an energy function, which allows a characterization of the system's attractors and dynamics. While the early work on binding models has centered around theoretical issues and simple test patterns, the recent years have seen a growing amount of work, where feature binding models have provided new algorithms to deal with difficult problems of segmentation in pattern recognition and computer vision.

Chapter 3

Linear Threshold Networks

This chapter discusses the special recurrent neural network type of linear threshold (LT) networks, since it is the underlying model of the recurrent dynamics in the competitive layer model. A reader mainly interested in feature binding and segmentation applications may skip this chapter at first reading, since the mathematical treatment is rather general and makes no particular reference to grouping.

Two new conditions are established, which ensure the non-divergence[1] of additive recurrent networks with unsaturating piecewise linear transfer functions, also called linear threshold or semilinear transfer functions. The conditions establish a regime, where linear threshold networks are at the same time sensitive to small changes in their inputs and are capable of strong recurrent amplification without runaway activity (Wersing, Beyn, & Ritter 2000). The first condition can be applied to nonsymmetric networks and has a simple interpretation of requiring that the strength of local inhibition must match the sum over excitatory weights converging onto a neuron. The second condition is restricted to symmetric networks, but can also take into account the stabilizing effect of non-local inhibitory interactions. The conditions are applied to a simple example and the orientation-selectivity model of Ben-Yishai et al. (1995). It is shown that for their model these conditions can be used to identify regions of maximal orientation-selective amplification and symmetry breaking. As the application results given in Chapter 5 demonstrate, the conditions are of particular relevance for exploiting the full capabilities of the competitive layer model. In the final section the relation to other formulations of piecewise linear dynamics is discussed and an efficient simulation algorithm is given.

3.1 Additive Recurrent Networks

A prominent model of the coupled dynamics of networks of neurons is the additive recurrent model

$$\dot{x}_i = -x_i + \sigma_i\Big(\sum_j w_{ij} x_j + h_i\Big), \tag{3.1}$$

where x_i denotes the activity of neuron i. The synaptic weights w_{ij} denote the strength of the synaptic connection from neuron j onto neuron i. The weights can be positive or negative corresponding to an excitatory or inhibitory synapse respectively. The function $\sigma_i(.)$

[1] or boundedness in the usual mathematical terminology

is called the transfer or gain function of the neuron i, and can be interpreted as computing the cell's average spike rate depending on its membrane potential. The constants h_i denote external inputs.

For the case of symmetric synaptic interactions satisfying $w_{ij} = w_{ji}$, the dynamics is largely characterized by a corresponding energy function (Cohen & Grossberg 1983; Hopfield 1984)

$$E = -\frac{1}{2}\sum_{ij} w_{ij}x_i x_j - \sum_i h_i x_i + \sum_i \int_0^{x_i} \sigma^{-1}(y)dy. \tag{3.2}$$

This energy is nonincreasing under the dynamics (3.1), i.e. $d/dtE \leq 0$, and serves therefore as a global Lyapunov function. A global Lyapunov function assures convergence to a bounded set of equilibria, if either $E \to \infty$ for $x \to \infty$ or the variables x_i are dynamically constrained to a bounded region (Hirsch 1989). The form of the transfer function $\sigma(.)$ (see Fig. 3.1) plays an important role for bounding the dynamics (3.1). If the transfer function has lower and upper saturation limits according to $a \leq \sigma(x) \leq b$, then the dynamics is obviously bounded, since $\dot{x}_i \geq 0$ for $x_i \leq a$ and $\dot{x}_i \leq 0$ for $x_i \geq b$. Examples for such saturating transfer functions are the logistic or Fermi function

$$\sigma(x) = \frac{1}{1 + e^{-2\beta x}}, \tag{3.3}$$

and the limiter function (also referred to as the saturating linear function or the piecewise linear sigmoidal function)

$$\sigma(x) = \begin{cases} a, & \text{if} \quad x < a, \\ x, & \text{if} \quad a \leq x \leq b, \\ b, & \text{if} \quad x > b. \end{cases} \tag{3.4}$$

The existence of an energy function, together with the boundedness property, offers the systematic method of designing the synaptic weights such that desired attractor states correspond to local minima of the energy. Since the high-dimensional energy function may have multiple local minima, the model may have multiple locally stable attractors and is thus a multistable dynamical system. Due to these well-established stability properties, symmetric additive recurrent networks with saturating transfer functions have found a wide range of applications in models of associative memory (Hopfield 1982; Hopfield 1984), combinatorial optimisation (Hopfield & Tank 1985) (see also Hertz, Krogh, & Palmer 1991 for related references), cognitive and psychological models (Anderson 1983), or Hebbian development processes (Linsker 1992; Feng, Pan, & Roychowdhury 1996), to name just a few.

In recent years there has been a growing amount of literature on biologically-based models (Hartline & Ratliff 1958; von der Malsburg 1973; Douglas, Mahowald, & Martin 1994; Ben-Yishai, Lev Bar-Or, & Sompolinsky 1995; Salinas & Abbott 1996; Adorjan, Levitt, Lund, & Obermayer 1999; Bauer, Scholz, Levitt, Obermayer, & Lund 1999), where the transfer function, denoted as semilinear or linear threshold (LT) function, is non-saturating and of the form

$$\sigma_i(x) = \begin{cases} 0 & \text{if} \quad x < \theta_i, \\ k_i(x - \theta_i) & \text{if} \quad x \geq \theta_i. \end{cases} \tag{3.5}$$

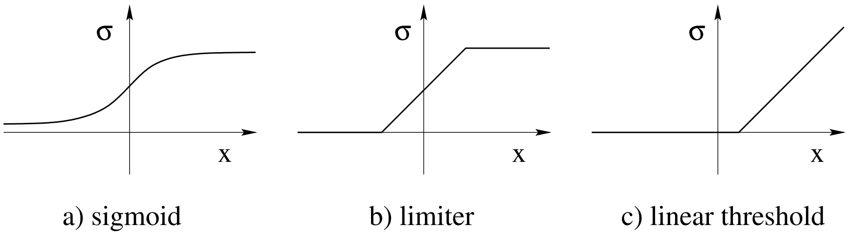


Figure 3.1: Transfer functions. The sigmoidal function a) is continuous and saturates for $x \to \infty$. The limiter function b) is linear between its saturation points. The linear threshold or semilinear transfer function does not saturate as $x \to \infty$.

where θ_i is the threshold and $k_i > 0$ is the gain of the transfer function of neuron i. It has been argued (Douglas, Koch, Mahowald, Martin, & Suarez 1995) that this transfer function is more appropriate, because cortical neurons rarely operate close to saturation, despite strong recurrent excitation. This indicates that the upper saturation may not be involved in the actual computations of the recurrent network, since the activation is dynamically bounded due to the effective net interactions. Another important field for non-saturating linear threshold transfer functions are winner-take-all (WTA) networks (Lippmann 1987; Hahnloser 1998; Ritter 1990; Wersing, Steil, & Ritter 1997,2000), where the lack of an upper saturation is necessary to exclude spurious ambiguous states.

Some properties of neural networks with piecewise linear transfer functions have been investigated since. Feng & Hadeler (1996) and Kuhn & Loehwen (1987) proved conditions which ensure the uniqueness of a stationary state. Feng, Pan, & Roychowdhury (1996) analyzed the general fixed point structure of networks with piecewise linear sigmoidal functions and a convergence condition for asynchronous update procedures was given in (Feng 1997). Hahnloser (1998) considered non-saturating linear threshold (LT) neurons for winner-take-all (WTA) networks and discussed their piecewise linear analysis. He emphasized the necessity of achieving both non-divergent and multistable dynamics in WTA applications and proved a corresponding condition for LT networks which is based on global inhibition with all-to-all connections.

To summarize, in spite of their wide application in biological models, there has been paid little attention so far to the general analysis of the conditions which ensure non-divergence in linear threshold networks. Therefore, this chapter gives an in-depth analysis of this issue.

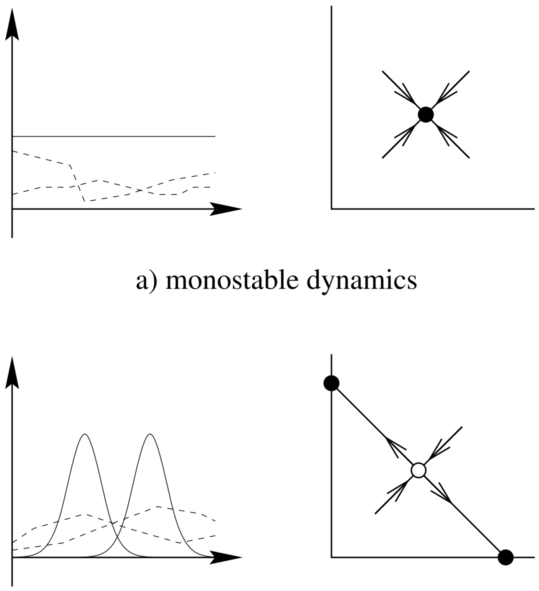
3.2 Monostable and Multistable Dynamics

In the following only the case where $k_i = 1$ and $\theta_i = 0$ is considered, since any network can be reduced to this form by an affine transformation of the activities that leaves the boundedness property of the dynamics invariant[2]. The dynamics is then given by

$$\dot{x}_i = -x_i + \sigma\Big(\sum_j w_{ij} x_j + h_i\Big), \tag{3.6}$$

where the transfer function is of the form $\sigma(x) = \max(0, x)$.

[2]If $\sigma_i(x_i) = k_i \max(0, x_i - \theta_i)$, the transformation is given by $x_i' = \sqrt{k_i}(x_i - \theta_i)$ and the corresponding coefficients are $w_{ij}' = \sqrt{k_i k_j} w_{ij}$ and $h_i' = (h_i - \theta_i)/\sqrt{k_i}$. Note also that for $W = \{w_{ij}\}$ being invertible, (3.6) is dynamically equivalent to $\dot{m}_i = -m_i + \sum_j w_{ij}\sigma(m_j) + h_i$, where $m_i = \sum_j w_{ij} x_j + h_i$.



a) monostable dynamics

b) multistable dynamics

Figure 3.2: Difference between monostable and multistable dynamics. In a monostable system a) different initial activations (dashed lines) converge to the same and unique attractor (straight line and filled circle on the right). In a multistable system b) many attractors can coexist and the dynamics also has unstable fixed points (open circle) which allow for unstable modes and symmetry breaking.

The standard approach to obtain a non-diverging dynamics (3.6) in the general case of nonsymmetric weights is to choose the combined gains of the transfer function and weights sufficiently small (see Steil 1999 for a review of related conditions). A simple example is the condition given by Hirsch (1989) which is based on the property that all eigenvalues of the symmetrical parts of the Jacobians of the vector field in (3.6) must be negative. This gives for (3.6) the conditions

$$w_{ii} + \frac{1}{2}\sum_{j\neq i} |w_{ij}| + |w_{ji}| < 1 \quad \text{for all} \quad i. \tag{3.7}$$

These conditions, however, imply *global* asymptotic stability, i.e. convergence to a single and unique attractor, and can be therefore only employed if the application requires a monostable dynamical behaviour (see Fig. 3.2). Nevertheless, there are many applications where *multistable* dynamics is essential to the model's operation. One example are WTA networks, where, depending on the external input (or the initial value) only the neuron with strongest input (or highest initial value) should remain active. If arbitrarily small differences can be amplified by this competitive process, then the system is said to exhibit symmetry-breaking behaviour. In the piecewise linear LT systems this is only possible if there are unstable directions in the state space, corresponding to local Jacobians with eigenvalues with positive real parts. The conditions that are presented in the following allow for this multistable behaviour.

3.3 Conditions for Non-Divergence

3.3.1 Nonsymmetric Networks

Hahnloser (1998) has proved a condition for non-divergence of LT networks that is based on a principle of global inhibition. For a given excitatory weight matrix with components $a_{ij} \geq 0$, he assumes a global inhibition with a strength equal to the sum over all excitatory weights diverging from a neuron. This leads to an effective weight matrix W of the form

$$w_{ij} = a_{ij} - \sum_k a_{kj}, \qquad \text{for all} \quad i, j. \tag{3.8}$$

Although the result is rather restrictive, and global all-to-all interactions are prohibitive in large networks, it is an example for a condition which does not imply monostability[3] and can be applied to some WTA networks (Hahnloser 1998).

In the following a more general criterion is proved, which shows that in fact local inhibition is sufficient to achieve boundedness of the LT dynamics. The condition is based on a reduced connection matrix W^+ with entries

$$w_{ij}^+ = \begin{cases} w_{ij} & \text{if} \quad i = j, \\ \max(0, w_{ij}) & \text{if} \quad i \neq j, \end{cases} \tag{3.9}$$

which is constructed from the diagonal and the positive off-diagonal entries of W.

Theorem 1. *If there exists a vector $\hat{\mathbf{v}}$ with positive components $\hat{v}_i > 0$ satisfying*

$$\sum_j w_{ij}^+ \hat{v}_j < \hat{v}_i \tag{3.10}$$

for all i, then the LT dynamics

$$\dot{x}_i = -x_i + \sigma\Big(\sum_j w_{ij} x_j + h_i\Big) \tag{3.11}$$

is bounded: If $0 \leq x_i(0) \leq c\hat{v}_i$ for all i, where $c > 0$ is sufficiently large to satisfy $c\big(\hat{v}_i - \sum_j w_{ij}^+ \hat{v}_j\big) > h_i$ for all i, then $0 \leq x_i(t) \leq c\hat{v}_i$ for all i and $t > 0$.

Proof. If the activities are positive for $t = 0$, $x_i(0) \geq 0$, then they stay positive since $x_i = 0$ implies $\dot{x}_i \geq 0$, which means that the activities cannot leave the positive domain. The boundedness is proved by constructing a "box" in the positive domain which cannot be left by the dynamics. The i-th face of the box is defined by

$$x_i = c\hat{v}_i, \qquad x_{j \neq i} \leq c\hat{v}_j. \tag{3.12}$$

Now we have either $\dot{x}_i = -x_i \leq 0$ or

$$\dot{x}_i = -x_i + \sum_j w_{ij} x_j + h_i \tag{3.13}$$

$$\leq -x_i + \sum_j w_{ij}^+ x_j + h_i. \tag{3.14}$$

[3] Suppose the matrix A with components a_{ij} is symmetric positive definite and has an eigenvector of $1's$. Since the row-constant shift in (3.8) affects only this eigenvalue, the resulting matrix W still has positive eigenvalues, causing unstable dynamical modes.

Due to (3.12) and the positiveness of both activities and $\hat{\mathbf{v}}$ we obtain

$$\dot{x}_i \leq -x_i + \sum_j w_{ij}^+ c\hat{v}_j + h_i \tag{3.15}$$

$$= c\Big(\sum_j w_{ij}^+ \hat{v}_j - \hat{v}_i\Big) + h_i. \tag{3.16}$$

Therefore, if we choose $c > 0$ sufficiently large such that $c > \max_i(h_i/(\hat{v}_i - \sum_j w_{ij}^+ \hat{v}_j))$, then $\dot{x}_i \leq 0$ on all faces i and the dynamics is bounded. □

Using the theory of nonnegative matrices (Berman & Plemmons 1979), condition (3.10) may be expressed in terms of eigenvalues of the matrix W^+. It is known that the maximal real part $Re(\lambda_{\max})$ of the eigenvalues of W^+ is an eigenvalue itself with corresponding nonnegative eigenvector $\hat{\mathbf{v}}$, $\hat{v}_i \geq 0$ for all i. Condition (3.10) is then equivalent to $\lambda_{\max}\{W^+\} < 1$. This shows that condition (3.10) is sharp if the eigenvector $\hat{\mathbf{v}}$ is positive and $W = W^+$ (no off-diagonal inhibitory weights). In fact, if W has an eigenvalue $\lambda > 1$ with positive eigenvector then the dynamics obviously tends to infinity after initialization with a suitable multiple of this eigenvector.

Another well-known characterization of condition (3.10) (see Berman & Plemmons 1979) states that the linear system

$$\hat{v}_i - \sum_j w_{ij}^+ \hat{v}_j = 1 \tag{3.17}$$

has a unique solution which has positive components $\hat{v}_i > 0$. Thus condition (3.10) can be tested by solving this system. Finally we notice that the special case $\hat{v}_i = 1$ for all i in Theorem 1 yields an especially simple criterion:

Corollary 1. *If the self-coupling weights satisfy $w_{ii} < 1 - \sum_{j \neq i} w_{ij}^+$ for all i then the LT dynamics is bounded.*

If we compare this to the global inhibition principle of Hahnloser, expressed in (3.8), we see that starting from the same a_{ij}, local inhibition mediated by a self-inhibitory term is sufficient for non-divergence with

$$w_{ij} = a_{ij} - \delta_{ij} \sum_k a_{ik}. \tag{3.18}$$

Since self-interacting synapses (autapses) are rare in cortical circuits, the local inhibition should be mediated by local interneurons in a biologically more realistic setting. If this inhibition is sufficiently fast, there is, however, no qualitative difference in the dynamics (Li & Dayan 1999).

3.3.2 Symmetric Networks

If the weights are symmetric, an additional criterion can be given that allows to improve the stability margins by taking into account also the non-local inhibitory contribution. The result is based on the energy function of the system which can be constructed for symmetric weights.

Theorem 2. *The LT dynamics is bounded if there exists a symmetric matrix $\hat{W}$ which satisfies $\hat{w}_{ij} \geq w_{ij}$ for all i, j and for which $\lambda_{\max}\{\hat{W}\} < 1$.*

Proof. The LT dynamics has an energy function of the form

$$E = -\frac{1}{2}\sum_{ij}(w_{ij} - \delta_{ij})x_i x_j - \sum_j h_j x_j. \tag{3.19}$$

If we take $E_i = \partial E/\partial x_i = \sum_j w_{ij}x_j - x_i + h_i$ then we see that E is nonincreasing along the trajectories of the system:

$$\dot{E} = \sum_j E_j \dot{x}_j = \sum_j E_j(-x_j + \sigma(-E_j + x_j)) \leq 0, \tag{3.20}$$

where we have used

$$E_j(-x_j + \sigma(-E_j + x_j)) = \begin{cases} -E_j^2 \leq 0 & \text{if} \quad E_j < x, \\ -x_j E_j \leq 0 & \text{if} \quad E_j \geq x_j \geq 0. \end{cases} \tag{3.21}$$

Since $x_i \geq 0$ we simply obtain

$$E = -\frac{1}{2}\sum_{ij}(w_{ij} - \delta_{ij})x_i x_j - \sum_j h_j x_j \tag{3.22}$$

$$\geq -\frac{1}{2}\sum_{ij}(\hat{w}_{ij} - \delta_{ij})x_i x_j - \sum_j h_j x_j \tag{3.23}$$

$$\geq -\frac{1}{2}(\lambda_{\max}\{\hat{W}\} - 1)||\mathbf{x}||^2 - \sum_j h_j x_j \to \infty \tag{3.24}$$

for $\mathbf{x} \to \infty$. Hence the dynamics must be bounded.

□

The matrix W^+ used in Theorem 1 provides an example for a possible choice of $\hat{W}$ with nonnegative off-diagonal elements. The components of $\hat{W}$, however, need not be positive. In the following an argument is given based on linear perturbation theory, which shows that negative entries can be used to infer larger stability margins.

Suppose the weight matrix W is decomposed into the excitatory part W^+ and an inhibitory off-diagonal part W^- according to $w_{ij} = w_{ij}^+ + \xi w_{ij}^-$, where ξ controls the strength of the off-diagonal inhibitory part. The condition of Theorem 1 amounts to choosing $\xi = 0$, taking $\hat{W} = W^+$ and neglecting the off-diagonal inhibition. The eigenvector of W^+ with largest eigenvalue $\lambda_{\max}$ has nonnegative components $\hat{v}_i \geq 0$. According to linear perturbation theory, the eigenvalue of the dominant eigenvector changes to first order in ξ by $\lambda'_{\max} = \lambda_{\max} + \xi\hat{v}^T W^- \hat{v}$. Since W^- has only nonpositive entries, the inequality $\lambda'_{\max} < \lambda_{\max}$ is satisfied. By raising ξ from zero the stability margin is increased by a decrease of the maximal eigenvalue. Since this, however, may in turn increase other eigenvalues for which then $\lambda_i > 1$, the increase in stability can be inferred only for small values of ξ.

What happens if the strength of the non-local inhibition is increased to large values of ξ ? For such a strong inhibition, only pools of neurons can remain active, which share only excitatory interactions. Since then only these pools contribute to the dynamics, we can reconsider Theorem 1, applied to all possible active pools. Since now the number of excitatory interactions which must be considered is smaller, strong inhibition has the expected effect of decreasing the local self-inhibition that is necessary to avoid runaway excitation.

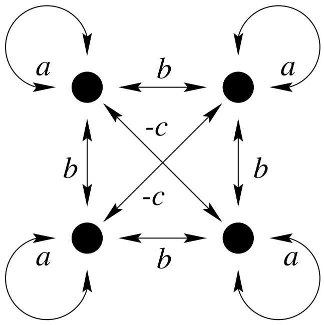


Figure 3.3: A simple symmetric network consisting of four neurons.

3.4 Examples

3.4.1 A Simple Example

Let us consider a simple example which consists of only four neurons, and which allows a complete discussion of the different parameter regimes of the presented conditions for non-divergence. The interaction of neural activities x_i, $i = 1, 2, 3, 4$ (see Fig. 3.3) is given by

$$W = \begin{pmatrix} a & b & -c & b \\ b & a & b & -c \\ -c & b & a & b \\ b & -c & b & a \end{pmatrix}, \tag{3.25}$$

where $b \geq 0$ and $c \geq 0$ give the strength of mutual excitation and inhibition respectively and a is the self-coupling weight. This circulant matrix has the corresponding eigenvectors $\mathbf{v}_i$ and eigenvalues λ_i:

$$\begin{aligned} \mathbf{v}_1 &= (1, 1, 1, 1), & \lambda_1 &= a + 2b - c, \\ \mathbf{v}_2 &= (1, 0, -1, 0), & \lambda_2 &= a + c, \\ \mathbf{v}_3 &= (0, 1, 0, -1), & \lambda_3 &= a + c, \\ \mathbf{v}_4 &= (1, -1, 1, -1), & \lambda_4 &= a - 2b - c. \end{aligned} \tag{3.26}$$

We now address the following question: Given $b \geq 0$ and $c \geq 0$, how small must be the self-coupling a to keep non-divergence of the dynamics?

If Corollary 1 is applied, which neglects all off-diagonal inhibitory contributions, we obtain the non-divergence condition

$$a < 1 - 2b, \tag{3.27}$$

which is not dependent on c.

According to Theorem 2 the non-divergence of a system with parameters (a, b, c) can be inferred from the non-divergence of a system with parameters (a', b', c'), where $a < a'$, $b < b'$, and $c > c'$. The maximum eigenvalue is given by

$$\lambda_{\max} = \begin{cases} a + 2b - c & \text{if} \quad b \geq c, \\ a + c & \text{if} \quad b < c. \end{cases} \tag{3.28}$$

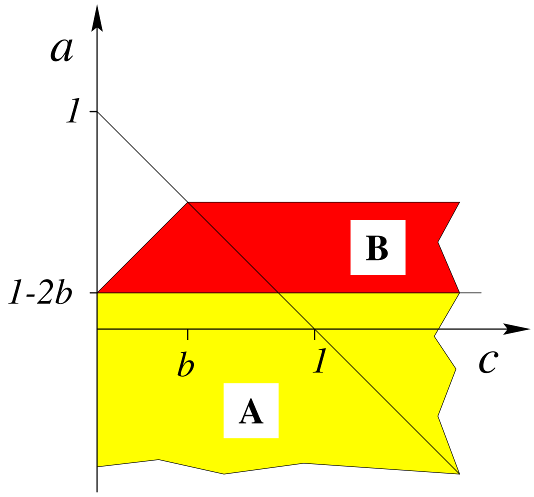


Figure 3.4: Non-divergence domains. The figure shows the non-divergent domains depending on the values of parameters a and inhibition strength $c > 0$. The system is bounded for parameters in region A according to Theorem 1. Theorem 2 covers boundedness in the additional region B. Note that due to Theorem 2 the non-divergence for a pair of parameters (a, c) implies non-divergence for all (a', c') with $a' < a$ and $c' > c$.

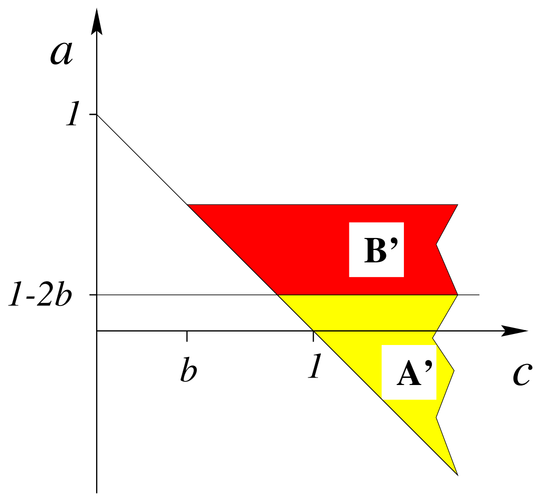


Figure 3.5: Non-divergence domains with multi-stable dynamics. Since the system is globally asymptotically stable if $a < 1 - c$, multistable dynamics only occurs above this line. Regions A' and B' show these sub-regions as indicated by Theorems 1 and 2 respectively.

The condition $\lambda_{\max} < 1$ then results in

$$i) \quad a < 1 - 2b + c \quad \text{for} \quad b > c, \tag{3.29}$$

$$ii) \quad a < 1 - c \quad \text{for} \quad b \leq c. \tag{3.30}$$

Case $i)$ improves the stability margin towards larger values of a with a maximum when $c = b$ to $a < 1 - b$. Case $ii)$ offers no advantage since $b \leq c$ implies that $1 - c < 1 - 2b$. The domains of non-divergence which can be obtained from these inequalities are shown in Figures 3.4 and 3.5. The diagrams demonstrate, that by taking into account off-diagonal inhibitory contributions Theorem 2 covers a larger area of the parameter space.

3.4.2 A Model for Orientation Selectivity

Ben-Yishai, Lev Bar-Or, & Sompolinsky (1995) have considered a one-dimensional continuum model for orientation selectivity which is of the form

$$\dot{x}(\phi) = -x(\phi) + \sigma(F(\phi)), \qquad -\pi/2 \leq \phi \leq \pi/2, \tag{3.31}$$

where

$$F(\phi) = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} f(\phi - \phi') x(\phi') d\phi' + h^{\text{ext}}(\phi - \phi_0), \tag{3.32}$$

with $f(\psi) = -J_0 + J_2 \cos(2\psi)$ and

$$\sigma(x) = \begin{cases} 0, & \text{if} \quad x \leq 1, \\ k(x-1), & \text{if} \quad 1 < x < k^{-1} + 1, \\ 1, & \text{if} \quad x \geq k^{-1} + 1. \end{cases} \tag{3.33}$$

The parameters $J_0 > 0$ and $J_2 > 0$ characterize on "On-Center-Off-Surround" interaction between a continuum of neurons in a 1-D orientation space with closed boundary conditions. The external input is of the form

$$h^{\text{ext}}(\phi) = c(1 - \eta + \eta \cos(2\phi)), \tag{3.34}$$

where c denotes the contrast and $\eta \ll 1$ is the amplitude of a weakly biased input stimulus. The parameter k is the gain of the transfer function. Of particular interest is a region in the $J_0 \times J_2$ space, where the model exhibits a so-called marginal phase. In this phase the model spontaneously generates orientation selectivity, by recurrently amplifying arbitrarily small anisotropies in the input for $\eta \to 0$. As the results of Ben-Yishai et.al. demonstrate, this occurs in a regime where the upper saturation of the transfer function is not reached, which corresponds to the case of non-divergent linear threshold dynamics. Therefore, we use in the following the presented conditions for non-divergence with multi-stable dynamics to identify this regime in the parameter space.

The linear integral operator in (3.32) has two continuous eigenfunctions x_0, x_2 with eigenvalues λ_0, λ_2, where

$$\begin{aligned} x_0(\phi) &= 1, & \lambda_0 &= -1 - J_0 k, \\ x_2(\phi) &= \cos(2\phi), & \lambda_2 &= -1 + \frac{k}{2} J_2. \end{aligned} \tag{3.35}$$

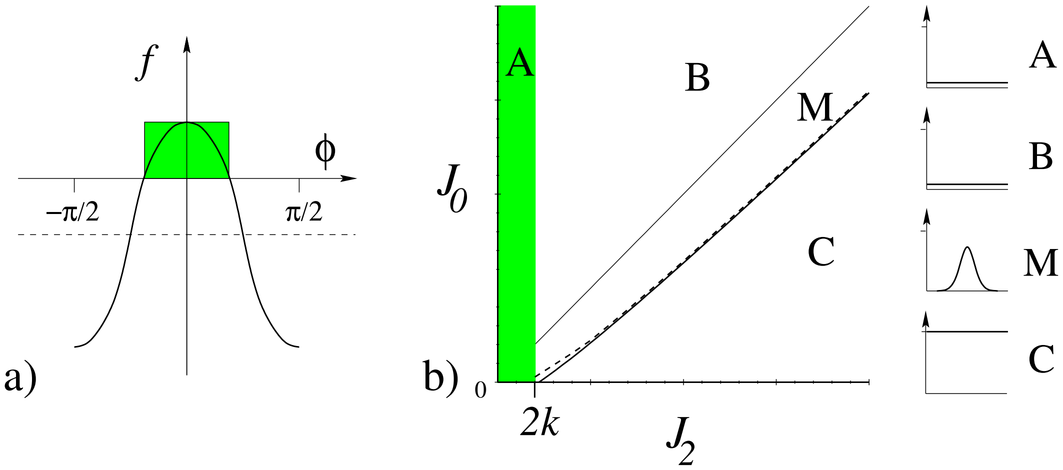


Figure 3.6: Angular On-Center-Off-Surround interaction and phase diagram of the orientation selectivity model by Ben-Yishai et.al. The shaded rectangle in a) is an estimate to the sum over excitatory interactions. With this estimate a phase diagram shown in b) can be obtained. In region A (shaded area) all eigenvalues are negative and the system always converges to an isotropic state. In region B the activity remains small due to large inhibition with $J_0 > J_2$ with the same unique isotropic attractor. Region M is the marginal phase, where a sharp and peaked response can be generated from isotropic input. In region C the network saturates due to strong excitation. The border between M and C is given by the non-divergence condition, the bold straight line gives the theoretical prediction for the critical inhibition J_0^c, while the dashed line was obtained from a simulation of the system. Within region M there is a gradual increase of peak height towards this boundary.

As Ben-Yishai et.al. have stated, a necessary condition for the system being in the marginal phase is that $J_2 > 2k^{-1}$, which in turn makes the x_2-mode unstable with $\lambda_2 > 0$.

We can now use Corollary 1 to infer a condition which ensures the non-divergence of the system, even if unstable modes are present. The straightforward generalization to the continuum case is given by the condition

$$0 > -1 + \frac{k}{\pi} \int_{-\pi/2}^{\pi/2} \max\big(0, f(\phi)\big) d\phi. \tag{3.36}$$

This can be geometrically interpreted as the condition that the area under the positive part of f is smaller than π/k. If $J_0 > J_2$ the lateral interaction is completely negative, and the condition is trivially satisfied. Due to the absence of excitatory interactions the activity settles down to zero for a weak unspecific input. For $J_0^c < J_0 < J_2$ the area can be estimated from above by the larger rectangle with height $J_2 - J_0$ and width between the zeroes of f (see Figure 3.6). Using the approximations $J_0 = J_2 - \epsilon$ with ϵ small and $\epsilon \arccos(1 - a\epsilon) \approx \sqrt{2a}\epsilon^{3/2}$ we obtain the condition

$$J_0 > J_0^c = J_2 - \frac{\pi^{2/3}}{2^{1/3}} \frac{1}{k^{2/3}} J_2^{1/3}. \tag{3.37}$$

This condition can be used to effectively identify the region in the $J_0 \times J_2$ space, where the lateral interaction is capable of recurrently amplifying small fluctuations in the activity distribution, without saturating at the upper saturation limit (see Figure 3.6). This leads to localized peaks with increasing amplification for $J_0 \to J_0^c$. For $J_0 > J_0^c$ the network runs into saturation. A comparison with a numerical simulation of the dynamics shows that (3.37) captures the border between maximal amplification and saturation rather precisely.

3.5 Dynamical Implementation

The linear threshold transfer function provides a very simple nonlinearity which facilitates alternative dynamical implementations. In the following two different formulations of the continuous time dynamics are compared and an effective iterative simulation algorithm is given.

3.5.1 Continuous Dynamics

The conventional LT dynamics, as it was introduced in Section (3.2) is given by

$$\dot{x}_i = F_i^\sigma(\mathbf{x}) = -x_i + \sigma\Big(x_i + \sum_j w_{ij}x_j + h_i\Big), \tag{3.38}$$

where F_i^σ denotes the vector field defining the dynamical system. In (3.38) a self-excitatory term is added which simplifies the following discussion and can be compensated for by taking $w'_{ij} = w_{ij} - \delta_{ij}$.

Earlier formulations of the competitive layer model (Wersing 1996; Wersing, Steil, & Ritter 1997) employed an alternative implementation of the form

$$\dot{x}_i = F_i^\omega(\mathbf{x}) = \omega_i \cdot \Big(\sum_j w_{ij}x_j + h_i\Big), \tag{3.39}$$

which avoids the occurrence of the $\sigma(.)$-nonlinearities and instead enforces the constraint $x_i \geq 0$ by a set of "switching variables"

$$\omega_i = \begin{cases} 0 & \text{if} \quad x_i = 0\ ,\ \sum_j w_{ij}x_j + h_i < 0, \\ 1 & \text{else.} \end{cases} \tag{3.40}$$

A closer inspection reveals that both vector fields F^σ and F^ω implement very similar dynamical systems. They share the same fixed points, since $F^\sigma(\mathbf{x}^F) = 0$ if and only if $F^\omega(\mathbf{x}^F) = 0$. For symmetric weights they share the same energy function, which according to (3.2) is given by

$$E = -\frac{1}{2}\sum_{ij} w_{ij}x_ix_j - \sum_i h_ix_i. \tag{3.41}$$

If we define $E_i = -\partial E/\partial x_i = \sum_j w_{ij}x_j + h_i$, then we see that E is nonincreasing under (3.38) (see proof of Theorem 2) , since

$$\frac{d}{dt}E = \sum_i \frac{\partial E}{\partial x_i}F_i^\sigma = -\sum_i E_i\big(-x_i + \sigma(E_i + x_i)\big) \leq 0. \tag{3.42}$$

The same holds for dynamics (3.39),

$$\frac{d}{dt}E = \sum_i \frac{\partial E}{\partial x_i}F_i^\omega = -\sum_i \omega_i E_i^2 \leq 0. \tag{3.43}$$

Since both dynamics share the same energy function they converge to the same attractors which are local minima of (3.41) and may be considered as dynamically equivalent. While the formulation (3.38) is more appropriate when referred to biological models, the formulation (3.39) has advantages for the linear analysis, shifting the nonlinearity to the boundary of the state space. Another difference is that (3.39) can produce faster transients with faster approach to attractors, since the condition $F_i^\sigma \geq -x_i$ limits the possible rate of activity decay.

3.5.2 Iterative Solution

The LT dynamics (3.38) or (3.39) can be simulated in principle in parallel by any differential equation integrator like the Euler or Runge-Kutta method. Also, due to the piecewise linearity, an analog VLSI implementation may be even simpler than classic circuits incorporating sigmoid transfer functions. If simulated on a serial processor, there is an alternative approach which replaces the explicit trajectory integration by a rapid search for fixed point attractors. This can be done by iteratively solving the fixed point equations which is largely facilitated by the piecewise linearity of the model (Ontrup & Ritter 1998). This iterative solution procedure, also known as a Gauss-Seidel approach has also been extensively used for Markov Random Field approaches to image segmentation (Besag, Green, Higdon, & Mengersen 1995). The algorithm can be implemented in the following way:

1. Initialize all x_i with small random values

$$x_i(t=0) \in [0, \epsilon].$$

2. Do N times: Choose i randomly and update $x_i = \max(0, \xi)$, where

$$\xi = \frac{h_i + \sum_{j \neq i} w_{ij} x_j}{1 - w_{ii}} \quad .$$

3. Go to step 2 until convergence.

The single activity update in step 2 corresponds to solving the fixed point equation for this activity with all other activities held constant. This asynchronous dynamics converges (Ontrup & Ritter 1998) due to a convergence result on asynchronous iteration in neural networks by Feng (1997). We observed a speed gain of roughly a factor of ten compared to Euler integration at comparable solution quality.

3.6 Summary

The two presented conditions allow to ensure the non-divergence of linear threshold networks without restricting them to be monostable. Although, at least for symmetric systems, the second condition covers a larger parameter regime, it requires the knowledge of eigenvalues which may not be readily available. On the contrary, the first condition can be easily applied by choosing appropriate local (self-inhibitory) interactions matching the sum over excitatory weights. Although the conditions are strictly speaking only sufficient and not necessary, the discussion of the orientation-selectivity model shows that they can be used to effectively approach the regime of critical amplification in LT networks, where small stimulus contrasts can be largely enhanced through dynamical symmetry breaking. The discussion of the dynamical implementation shows that older formulations of the piecewise linear dynamics for the competitive layer model are very similar to the biologically more plausible linear threshold formulation. The proposed iterative solution algorithm makes effective usage of the simple nonlinearity of the linear threshold dynamics.

Chapter 4

The Competitive Layer Model

In this chapter the competitive layer model (CLM) is introduced as a recurrent neural network architecture for feature binding and labeling problems. First the structure of the model with its basic topographic competitive and cooperative interactions is described. It is shown that the differential equations defining the model dynamics correspond to gradient descent in an energy function. A mechanism is presented which implements deterministic annealing in the CLM by dynamically varying the self-inhibitory connections. The properties of the resulting competitive grouping model are discussed referring to the energy function and the approach is compared to related Potts-spin mean field models.

In the theoretical analysis section the general stability conditions for linear threshold networks established in Chapter 3 are applied to the CLM to ensure convergence of the model. Additional conditions are proved, which guarantee that only unambiguously assigned states are attractors of the model. An eigensubspace analysis is done which allows to characterize the grouping dynamics in relation to the lateral coupling schemes. The efficient simulation on standard serial computers is described in terms of an explicit algorithmic formulation.

Finally, the competitive layer model is slightly generalized by considering more general coupling schemes, where the basic extension is given by also allowing interlateral interactions between layers. This generalization is necessary for the application of the CLM to labeling problems where more general compatibilities between label assignments are considered. The chapter concludes with a discussion of the relation of the CLM to other labeling models.

4.1 The CLM feature binding model

4.1.1 The CLM architecture

The CLM consists of a set of L layers of feature-selective neurons (see Figure 4.1). The activity of a neuron at position r in layer α is denoted by $x_{r\alpha}$, and a *column* r denotes the set of the neuron activities $x_{r\alpha}$, $\alpha = 1, \ldots, L$ that share a common position r in each layer. With each column a particular "feature" is associated, which is described by a parameter vector $\mathbf{m}_r$. A typical feature example are local edge elements which are characterized by position and orientation, $\mathbf{m}_r = (x_r, y_r, \theta_r)$. More complex features were used by Ontrup & Ritter (1998) for texture segmentation, where a texture feature consisted of a vector of local Gabor filter responses at different spatial frequencies and orientations.

A binding between two features, represented by columns r and r', is expressed by simul-

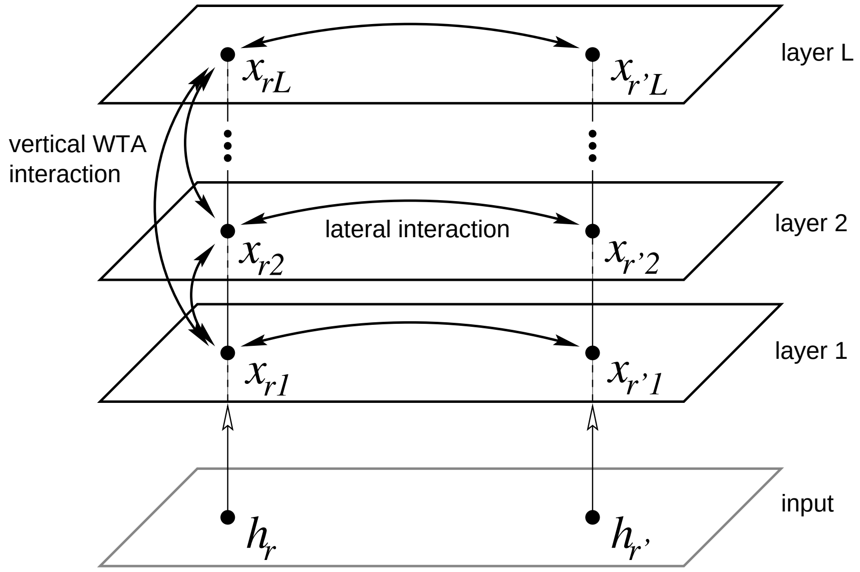


Figure 4.1: The CLM architecture. For each input feature, there is in each layer a responding neuron. A vertical columnar WTA circuit implements a topographic competition between layers. Lateral interactions characterize compatibility between features and guide the binding process.

taneous activities $x_{r\hat{\alpha}} > 0$ and $x_{r'\hat{\alpha}} > 0$ that share a common layer $\hat{\alpha}$. Therefore, binding is achieved by having each (activated) column r assign its feature to one of the layers α, interpreting the activity $x_{r\alpha}$ as a measure for the certainty of that assignment. All the neurons in a column r are equally driven by an external input h_r which is to be interpreted as the significance of the detection of feature r by a preprocessing step. The afferent input h_r is fed to the activities $x_{r\alpha}$ with a connection weight $J_r > 0$.

Within each layer α the activities are coupled by the lateral interaction $\mathsf{f}^{\alpha}_{rr'}$ which corresponds to the degree of compatibility between features r and r' and which is a symmetric function of the feature parameters, $\mathsf{f}^{\alpha}_{rr'} = \mathsf{f}^{\alpha}(\mathbf{m}_r, \mathbf{m}_{r'}) = \mathsf{f}^{\alpha}(\mathbf{m}_{r'}, \mathbf{m}_r)$. The interactions $\mathsf{f}^{\alpha}_{rr'}$ determine which pattern configurations, if elicited as activity pattern within a single layer α, will be mutually supporting among their constituent parts ($\mathsf{f}^{\alpha}_{rr'} > 0$) or instead suffer mutual inhibition ($\mathsf{f}^{\alpha}_{rr'} < 0$). Two examples for lateral interactions motivated by Gestalt laws of perceptual grouping are visualized in Figure 4.2. For the applications to perceptual grouping considered in this thesis, the lateral interaction pattern is either identical in all layers or one layer is changed into a so-called ”ground” layer to perform additional figure-ground segmentation.

The purpose of the layered arrangement in the CLM is to enforce a dynamical assignment of the input features to the layers, using the contextual information stored in the lateral interactions. The unique assignment to a single layer is realized by a columnar Winner-Take-All (WTA) circuit, which uses mutual symmetric inhibitory interactions with strength $I^{\alpha\beta}_r = I^{\beta\alpha}_r > 0$ between neural activities $x_{r\alpha}$ and $x_{r\beta}$ that share a common column r. Due to the WTA coupling, a stable equilibrium state of the CLM can have only one active layer within each column. The subset of features which are active in a particular layer α are also denoted as a *group* α. The number of layers does not determine the number of groups, since for sufficiently many layers only those are active that carry a salient group.

The combination of afferent inputs and lateral and vertical interactions is combined into

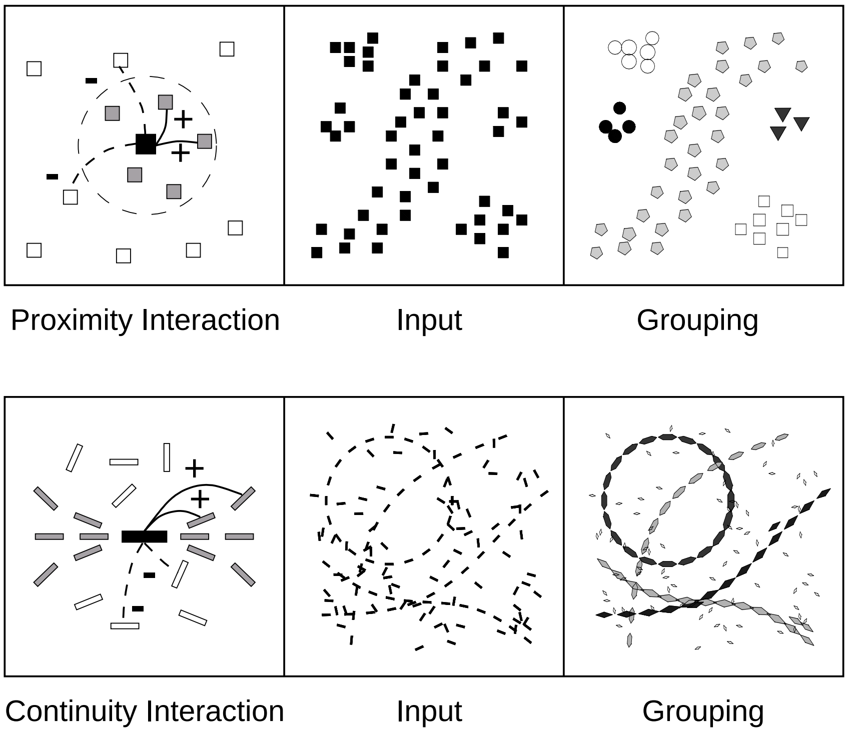


Figure 4.2: Examples of lateral interactions for perceptual grouping. An "On-Center-Off-Surround" interaction which is excitatory for short distances and weakly inhibitory for larger distances is capable of clustering according to proximity of points. Different symbols denote activity in different layers. The Gestalt principle of continuity can be modelled by a local cocircular interaction of edge elements that lie along curves of constant curvature combined with a weak long-range inhibition to separate different segments.

the standard linear threshold additive activity dynamics

$$\begin{aligned} \dot{x}_{r\alpha} &= -x_{r\alpha} + \sigma\Big(J_r h_r - \sum_{\beta} I_r^{\alpha\beta} x_{r\beta} + \sum_{r'} \mathsf{f}_{rr'}^{\alpha} x_{r'\alpha} + x_{r\alpha}\Big) \\ &= -x_{r\alpha} + \sigma(E_{r\alpha} + x_{r\alpha}), \end{aligned} \tag{4.1}$$

where $\sigma(x) = \max(0, x)$ is a non-saturating linear threshold transfer function, and $E_{r\alpha} + x_{r\alpha}$ is the total input to neuron $x_{r\alpha}$. The additional self-excitatory term $+x_{r\alpha}$ in (4.1) simplifies the following computations and can be compensated for by taking $\mathsf{f'}_{rr'}^{\alpha} = \mathsf{f}_{rr'}^{\alpha} - \delta_{rr'}$.

4.1.2 Energy Formulation and Binding Dynamics

The dynamics (4.1) has an energy function of the form

$$E = -\sum_{r\alpha} J_r h_r x_{r\alpha} + \frac{1}{2}\sum_{r}\sum_{\alpha\beta} I_r^{\alpha\beta} x_{r\alpha} x_{r\beta} - \frac{1}{2}\sum_{\alpha}\sum_{rr'} \mathsf{f}_{rr'}^{\alpha} x_{r\alpha} x_{r'\alpha}. \tag{4.2}$$

The energy satisfies $\partial E/\partial x_{r\alpha} = -E_{r\alpha}$, which makes E nonincreasing under the dynamics (4.1),

$$d/dt\, E = -\sum_{r\alpha} E_{r\alpha}(-x_{r\alpha} + \sigma(E_{r\alpha} + x_{r\alpha})) \leq 0, \tag{4.3}$$

since (4.1) confines the activities to be nonnegative (compare Section 3.3). Thus the attractors of the dynamics (4.1) are the local minima of (4.2) under constraints $x_{r\alpha} \geq 0$. While the vertical interactions $I_r^{\alpha\beta}$ establish the columnar WTA process, the lateral interactions $\mathsf{f}^{\alpha}_{rr'}$ contribute in the quadratic energy as a sum over all pairwise compatibilities within groups, weighted with the corresponding activity values. If we interpret the negative value of the energy as the overall quality of the grouping, the aim of the dynamics (4.1) is to reach a globally minimal or almost minimal energy state.

The quadratic summation approach can be compared to the corresponding Potts spin mean field free energy (Peterson & Soderberg 1989)

$$E_{\text{PMF}} = -\frac{1}{2}\sum_{\alpha}\sum_{rr'} \mathsf{f}^{\alpha}_{rr'} x_{r\alpha} x_{r'\alpha} + T\sum_{r\alpha} x_{r\alpha}\log x_{r\alpha}, \tag{4.4}$$

which is to be minimized subject to the weighting constraint $\sum_\alpha x_{r\alpha} = 1$. The additional convex entropy term that is weighted by the temperature parameter T biases the local minima of (4.4) towards soft columnar assignments, since it attains its minimum for the "maximally undecided" state with $x_{r\alpha} = 1/L$. Minimal solutions to (4.4) can then be obtained by the recurrent dynamics

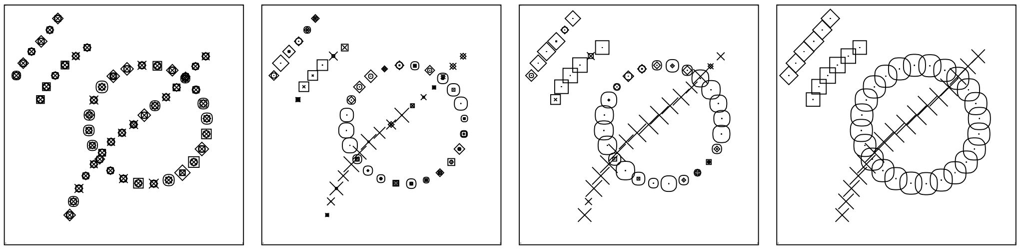
$$\dot{x}_{r\alpha} = -x_{r\alpha} + \frac{e^{F_{r\alpha}/T}}{\sum_\beta e^{F_{r\beta}/T}}, \qquad F_{r\alpha} = \sum_{r'} \mathsf{f}^{\alpha}_{rr'} x_{r'\alpha}, \tag{4.5}$$

which implements a columnar WTA circuit that is gradually sharpened by decreasing (or "annealing") $T \to 0$.

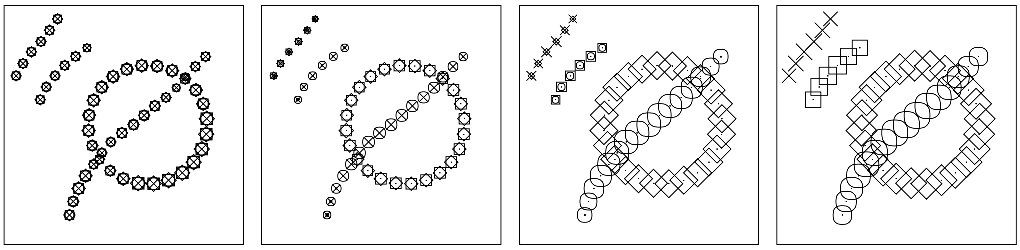
The CLM dynamics (4.1) provides a similar soft competition scheme between all possible groupings in the different layers of the model. The hard weighting constraint and divisive nonlinearity (4.5), however, are replaced by a more flexible WTA circuit, that is coupled dynamically to the afferent input h_r and permits a contextual modulation of the input according to the salience of the grouping, but also requires an additional stability analysis. Perceptual context effects of this kind have been observed in a wide range of physiological (Gilbert 1992; Kapadia, Ito, & Westheimer 1995) and psychophysical (Kovács & Julesz 1993; Field, Hayes, & Hess 1992; Polat & Sagi 1994) studies. As the later eigensubspace analysis of the CLM dynamics reveals, the columnar WTA dynamics is driven by eigenmodes that depend on the lateral interaction pattern and whose eigenvalues characterize the rate of the group formation. The analysis motivates a mechanism for slowing down the lateral mode dynamics for increased grouping quality by adding a self-inhibitory coupling at each neuron of the form $\mathsf{f}'^{\alpha}_{rr'} = \mathsf{f}^{\alpha}_{rr} - T\delta_{rr'}$, where $T \geq 0$ is the strength of this self-inhibition. This leads to a new CLM energy

$$E' = E + T\sum_{r\alpha} x^2_{r\alpha}, \tag{4.6}$$

which adds an analog convex term that biases the local minima towards graded assignments and thus makes the WTA more soft. Similar to annealing in the Potts model (Peterson &



a) Activity dynamics without annealing



b) Activity dynamics with self-inhibitory annealing

Figure 4.3: Activity dynamics and group formation in the CLM. The neural activity is displayed by different symbols in each layer with symbol size representing activity value. The lateral interaction is locally cocircular with a weak global inhibition as illustrated in Figure 4.2. A typical time-course of the dynamics is shown in a). After initialization with random activity values first all activities within in a column are equally active in all layers. After that, dynamical modes, which break the symmetry between layers, cause formation of groups. The "soft competition" between the group assignments lasts until the columnar WTA circuit has caused an assignment to one of the layers. The modes are given by the eigenmodes of the lateral interaction pattern. The main problems are long-range coherence with local interactions and a proper separation of different groups. Fragmented groupings can occur especially for very short-ranged interactions due to fast expression of symmetry-breaking modes which lead to fragmented groups. b) shows how the mechanism of self-inhibitory annealing can be used to suppress these fragmenting modes. During the dynamics the strength of self-inhibition is lowered, which has the effect of making the columnar WTA circuit less strict and increases the "soft competition" between grouping assignments of a single feature. The groups are then differentiated in a sequence, where the most salient structures appear first.

Soderberg 1989), we can then use a gradual lowering of the inhibitory self-coupling T to sweep from graded assignments to the final unique assignment as the final grouping result.

A time course of the dynamics with and without self-inhibitory annealing performing contour grouping is shown in Figure 4.3. Annealing results in a much more regular group formation process, which produces groups proceeding hierarchically from more to less salient groups. The slowing down of the dynamics, however, results in a tradeoff for convergence time. Whether annealing is necessary depends on the complexity of the input pattern with respect to the lateral interactions. Two problems that have been emphasized by Wang & Terman (1997) are long-range coherence with local interactions and the proper separation of different segments. Self-inhibitory annealing increases long-range coherence, and also reduces the necessary strength of global inhibition for achieving separation, by a reinforcement of the corresponding dynamical modes, which is discussed in detail in the following section.

4.1.3 Earlier Work on the CLM

The competitive layer model was suggested as a recurrent neural network model for spatial feature binding by Ritter (1990), where the idea of topographically coupled competitive layers was introduced. An "On-Center-Surround-Interaction" was suggested for implementing the Gestalt law of proximity and conditions were stated which ensure convergence and the assignment property for that lateral interaction. The conditions were later proved in (Wersing 1996), where also an eigensubspace analysis was developed for the case of equal lateral interactions which will be given in the following section. Another focus in (Wersing 1996) was the investigation of a continuum approximation in terms of integral operators, for which the dynamics induced by the proximity grouping interaction could be analyzed with linear perturbation theory.

4.2 Theoretical Analysis

The dynamics (4.1) is of the linear threshold type discussed in Chapter 3. The index i is replaced by a corresponding double index (r, α). Therefore, if we consider the following correspondences

$$x_i \leftrightarrow x_{r\alpha}, \qquad w_{ij} \leftrightarrow \mathsf{G}^{\alpha\beta}_{rr'}, \qquad h_i \leftrightarrow J_r h_r,$$

where the $(NL \times NL)$ matrix G with entries $\mathsf{G}^{\alpha\beta}_{rr'}$ establishes the combined vertical and lateral interaction between neural activities $x_{r\alpha}$ and $x_{r'\beta}$, the CLM dynamics can be written in the form

$$\dot{x}_{r\alpha} = -x_{r\alpha} + \sigma\Big(\sum_{r'\beta} \mathsf{G}^{\alpha\beta}_{rr'} x_{r'\beta} + J_r h_r\Big), \tag{4.7}$$

with

$$\mathsf{G}^{\alpha\beta}_{rr'} = -\delta_{rr'} I^{\alpha\beta}_r + \delta_{\alpha\beta}(\mathsf{f}^{\alpha}_{rr'} + \delta_{rr'}). \tag{4.8}$$

The theoretical analysis begins with a discussion of the model attractors and the conditions for convergence. Then an eigensubspace analysis is used to discuss which of the attractors are preferred by the recurrent dynamics. With a synthetic example, it is illustrated how this relates to the properties of the binding process implemented by the CLM.

4.2.1 Convergence

Networks composed of non-saturating linear threshold neurons as in (4.1) may be divergent if the excitatory interactions are not balanced by sufficient inhibition. If we apply the simple non-divergence condition of Corollary 1 (see Section 3.3), we obtain the following condition to ensure boundedness and convergence for the CLM system of layerwise coupled WTA columns.

Theorem 3. *The CLM dynamics is bounded and convergent, if the stability margins $\kappa_{r\alpha}$ satisfy*

$$\kappa_{r\alpha} = I_r^{\alpha\alpha} - \mathsf{f}_{rr}^{\alpha} - \sum_{r' \neq r} \max(0, \mathsf{f}_{rr'}^{\alpha}) > 0 \quad \textit{for all} \quad r, \alpha. \tag{4.9}$$

Then a bound $M = \max_{r\alpha}(J_r h_r / \kappa_{r\alpha})$ on the activities can be given, for which $0 \leq x_{r\alpha}(t) \leq M$ for all r, α and $t > 0$, if the activities are initialized with values smaller then M according to $0 \leq x_{r\alpha}(0) \leq M$.

Proof. If we define G^+ by the diagonal and nonnegative off-diagonal entries of G, (see Section 3.3.1), the conditions of Corollary 1 are given by

$$0 < 1 - \sum_{r'\beta} \mathsf{G}_{rr'}^{\alpha\beta +} = I_r^{\alpha\alpha} - \mathsf{f}_{rr}^{\alpha} - \sum_{r' \neq r} \max(0, \mathsf{f}_{rr'}^{\alpha}) \quad \text{for all} \quad r, \alpha. \tag{4.10}$$

If this condition is satisfied, then the dynamics is bounded. Together with the existence of the energy function (4.2) this ensures that the dynamics converges to the set of equilibria. □

The factors $\kappa_{r\alpha}$ control the stability margin and the maximal amplification of the inputs h_r. They are positive, if the self-interaction strength $I_r^{\alpha\alpha} - \mathsf{f}_{rr}^{\alpha}$ of a neuron is larger than the sum of the excitatory connections $\sum_{r' \neq r} \max(0, \mathsf{f}_{rr'}^{\alpha})$ converging onto it. In the simple case of equal vertical interactions with $J_r = I_r^{\alpha\beta} = J$ for all r, α, β, the afferent input and vertical interaction terms can be combined in the energy[1] to the term $J/2 \sum_r \left(h_r - \sum_\beta x_{r\beta}\right)^2$. We can interpret this as a penalty term with J as a constraint multiplier enforcing the condition, that the summed activity in a column is equal to the input h_r. Unlike as the usual penalty function approach to combinatorial optimisation problems, in the framework of sensory segmentation which is considered here, a strict enforcement of this constraint is undesirable. If we choose J close to the stability margin of Theorem 3, the penalty term derived from the vertical interactions no longer dominates over the lateral interactions and the columnwise activity is strongly influenced by lateral effects.

For the case of equal vertical interactions, a second result can be obtained from the second non-divergence condition for symmetric LT networks, given in Theorem 2.

Theorem 4. *Let the vertical interactions be equal to $I_r^{\alpha\alpha} = J_r = J$ for all r, α, β. If $J > \lambda_{\max}\{\mathsf{f}_{rr'}^{\alpha}\}$ for all α, where $\lambda_{\max}\{\mathsf{f}_{rr'}^{\alpha}\}$ is the largest eigenvalue of the lateral interaction matrix within a layer α, then the CLM dynamics is bounded and convergent.*

Proof. Let a matrix $\hat{\mathsf{G}}$ (see Theorem 2) be constructed by omitting all off-diagonal vertical interactions with

$$\hat{\mathsf{G}}_{rr'}^{\alpha\beta} = -\delta_{rr'}\delta_{\alpha\beta}(J-1) + \delta_{\alpha\beta}\mathsf{f}_{rr'}^{\alpha}. \tag{4.11}$$

The matrix $\hat{\mathsf{G}}$ satisfies $\hat{\mathsf{G}}_{rr'}^{\alpha\beta} \geq \mathsf{G}_{rr'}^{\alpha\beta}$ for all r, r', α, β. Since $\hat{\mathsf{G}}$ contains no cross-layer coupling it has a layer-wise block structure, where all entries corresponding to a coupling of neurons which are in different layers are zero. The eigenvalues of $\hat{\mathsf{G}}$ are then given by the eigenvalues of the lateral interaction within layers which are given by $\lambda_i\{\mathsf{f}_{rr'}^{\alpha}\} - J + 1$.

[1] If a constant term $\frac{1}{2} J \sum_r h_r^2$ is added to the energy, which, however, has no relevance for the gradient descent dynamics.

Therefore, if $J > \lambda_{\max}\{\mathsf{f}^{\alpha}_{rr'}\}$, then $\lambda_{\max}\{\hat{\mathsf{G}}\} < 1$ and the dynamics is bounded according to Theorem 2. Since the dynamics also has an energy function, it converges to the set of equilibria. □

The two conditions of Theorem 3 and Theorem 4 are independent, since cases can be constructed where either one or the other gives a lower resulting bound.

4.2.2 Assignment

Essential to the columnwise WTA approach is that we obtain a proper assignment of the features, that is, a state where each column r contains at most a single nonvanishing activity $x_{r\hat{\alpha}} > 0$. The conditions on the vertical interactions for this assignment property to hold are stated in the following theorem:

Theorem 5. *Let $F_{r\alpha} = \sum_{r'} \mathsf{f}^{\alpha}_{rr'} x_{r'\alpha}$ denote the lateral feedback of neuron r, α. If the lateral interaction is self-excitatory, $\mathsf{f}^{\alpha}_{rr} > 0$ for all r, α, and the vertical interactions satisfy $I^{\alpha\alpha}_r I^{\beta\beta}_r \leq (I^{\alpha\beta}_r)^2$ for all α, β, then an attractor of the CLM has in each column r either*

i) at most one positive activity $x_{r\hat{\alpha}}$ with

$$x_{r\hat{\alpha}} = \frac{J_r}{I^{\hat{\alpha}\hat{\alpha}}_r} h_r + \frac{F_{r\hat{\alpha}}}{I^{\hat{\alpha}\hat{\alpha}}_r}, \quad x_{r\beta} = 0 \text{ for all } \beta \neq \hat{\alpha}, \tag{4.12}$$

where $\hat{\alpha} = \hat{\alpha}(r)$ is the index of the maximally supporting layer characterized by $F_{r\hat{\alpha}} > F_{r\beta}$ for all $\beta \neq \hat{\alpha}$, or

ii) all activities $x_{r\alpha}$, $\alpha = 1, \ldots, L$ in a column r vanish and $F_{r\alpha} \leq -J_r h_r$ for all $\alpha = 1, \ldots, L$.

The proof of this Theorem is carried out in Section 4.4, where a more general interaction pattern is considered. This theorem states that as long as there are self-excitatory interactions within each layer, and the vertical cross-inhibition is sufficiently strong, the CLM converges to a unique assignment of features to the layer with maximal lateral feedback $F_{r\alpha}$. A central result is that, due to the layered topology of the network, this does not require arbitrarily large vertical couplings. Also note that the lack of an upper saturation is essential for the WTA behaviour, because it allows to exclude spurious ambiguous states. Nonzero activities are only stable due to a dynamical equilibrium and not due to saturation. Douglas et al. (1995) and Hahnloser (1998) have argued for the plausibility of this form of dynamical stability since cortical neurons rarely operate close to saturation.

The lateral and vertical feedback modulates the input intensity by an amount that is dependent on the ratios of $I^{\alpha\alpha}_r, F_{r\alpha}$ and J_r. Consider again the simple example $J_r = I^{\alpha\beta}_r = J$, for all r, α, β where, according to Theorem 5, the output of the only active neuron in a column is then given by

$$x_{r\hat{\alpha}} = h_r + \frac{F_{r\hat{\alpha}}}{J}. \tag{4.13}$$

This shows that by lowering J the lateral effects can be increased and we obtain a context-dependent activity distribution, which still remains sensitive to the input intensities. Note that the additional self-inhibitory annealing by choosing $\mathsf{f}'^{\alpha}_{rr'} = \mathsf{f}^{\alpha}_{rr} - T\delta_{rr'}$ for T sufficiently large violates the condition of Theorem 2 and leads to stable states with multiple nonzero activity within a column, as the following eigensubspace analysis shows.

4.2.3 Eigensubspace Analysis

After having discussed the possible attractors of the model, an interesting question is which of them will be preferred by the dynamics. Since the CLM time development is defined by the piecewise linear differential equation (4.7), the method of eigensubspace analysis can be applied to obtain a solution in the linear domain, where the constraints are inactive. In the following the special case is considered of a lateral interaction $\mathsf{f}^{\alpha}_{rr'} = \mathsf{f}_{rr'}$ for all α that is identical in all layers. For simplicity $J_r = I^{\alpha\beta}_r = J$ is taken as constant for all r, α, β. This is also the case that will be considered for the applications of the CLM to perceptual grouping. As will be shown in this section, the special form of two topological "orthogonal" interactions allows us to characterise the global eigenmodes completely in terms of the lateral eigenmodes.

Apart from constraints, the system is described by the linear dynamics

$$\dot{x}_{r\alpha} = Jh_r + \sum_{r'\beta} \mathsf{G}^{\alpha\beta}_{rr'} x_{r'\beta}, \tag{4.14}$$

where $\mathsf{G}^{\alpha\beta}_{rr'} = -J\delta_{rr'} + \delta_{\alpha\beta}\mathsf{f}_{rr'}$. Introducing the $N \times L$ vectors

$$\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_L) \quad \text{with} \quad \mathbf{x}_\alpha = (x_{1\alpha}, \ldots, x_{N\alpha}) \quad \text{and} \tag{4.15}$$

$$\mathbf{h} = (\mathbf{h}_0, \ldots, \mathbf{h}_0) \quad \text{with} \quad \mathbf{h}_0 = (h_1, \ldots, h_N) \tag{4.16}$$

we can write this in matrix notation as

$$\dot{\mathbf{x}} = J\mathbf{h} + \mathsf{G}\mathbf{x} \tag{4.17}$$

where G is a linear transformation in the space $\mathbb{R}^{N\times N} \otimes \mathbb{R}^{L\times L}$ and can be decomposed as

$$\mathsf{G} = \mathsf{f} \otimes \mathsf{Id}^{L\times L} - J\,\mathsf{Id}^{N\times N} \otimes \mathbb{I}^{L\times L} \tag{4.18}$$

with Id as the identity and $\mathbb{I}^{L\times L}$ as an $L \times L$-matrix of 1's. Equation (4.18) shows that an orthonormal eigenvector basis $\{\mathbf{v}^{i\gamma}, \Lambda_{i\gamma}\}$ for G can be obtained from orthonormal eigenvector bases [2] $\{\mathbf{b}^i, \lambda_i\}_{i=1\ldots N}$ and $\{\mathbf{q}^\gamma, \mu_\gamma\}_{\gamma=1\ldots L}$ for f and $\mathbb{I}^{L\times L}$ respectively:

$$\mathbf{v}^{i\gamma} = \mathbf{b}^i \otimes \mathbf{q}^\gamma, \qquad i = 1\ldots N,\ \gamma = 1\ldots L, \tag{4.19}$$

$$\Lambda_{i\gamma} = \lambda_i - J\mu_\gamma. \tag{4.20}$$

$\mathbf{q}^\gamma, \mu_\gamma$ can be calculated analytically. Here, only the property is used that the first eigenvector obviously is given by

$$\mathbf{q}^1 = 1/\sqrt{L}(1, \ldots, 1), \qquad \mu_1 = L, \tag{4.21}$$

while the remaining orthogonal eigenvectors have zero eigenvalues $\mu_2 = \mu_3 = \cdots = \mu_L = 0$. This has a direct geometrical interpretation. We can divide the eigenmodes of the linear system into two classes:

- DC-Subspace: This is the subspace spanned by the eigenmodes

  $$\mathbf{v}^{i1} = 1/\sqrt{L}(\mathbf{b}^i, \ldots, \mathbf{b}^i), \quad \Lambda_{i1} = \lambda_i - JL \tag{4.22}$$

  which contains the coherent eigenmodes which have equal components in all layers. For sufficiently large J all the corresponding eigenvalues Λ_{i1} are negative.

[2] We assume that the λ_i are sorted in descending order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$

- AC-Subspace: This is the orthogonal subspace spanned by the remaining eigenvectors

$$\mathbf{v}^{i\gamma\neq 1} = (q_1^\gamma \mathbf{b}^i, \ldots, q_L^\gamma \mathbf{b}^i), \quad \Lambda_{i\gamma} = \lambda_i \tag{4.23}$$

which contain the eigenmodes causing differences in the activity patterns between layers and which have a zero mean summed over columns. Their eigenvalues $\Lambda_{i\gamma\neq 1}$ are given by the eigenvalues λ_i of the lateral interaction $\mathsf{f}_{rr'}$ alone and are $L-1$ degenerate due to the symmetry between layers.

The linear dynamics can be completely characterized by the eigensubspaces and the location of the fixed point which is determined by the inhomogeneous input $\mathbf{h}$. Relative to the fixed point, the components

$$c_{i\gamma}(t) = (\mathbf{x}(t) - \mathbf{x}^F) \cdot \mathbf{v}^{i\gamma} \tag{4.24}$$

develop according to

$$c_{i\gamma}(t) = c_{i\gamma}(0)e^{\Lambda_{i\gamma} t}. \tag{4.25}$$

This corresponds to exponential growth for positive eigenvalues, while negative eigenvalues lead to attracting affine constraint surfaces which pass through the fixed point.

By expanding the fixed point equation $\dot{\mathbf{x}} = 0$ in the basis $\mathbf{v}^{i\gamma}$ we obtain the following expansion of the fixed point $\mathbf{x}^F$ in powers of λ_i/JL (note that $\mathbf{h}$ and thus also $\mathbf{x}^F$ has no component in the AC space):

$$\mathbf{x}_\alpha^F = \frac{1}{L}\mathbf{h}_0 + \frac{1}{L}\sum_i \frac{\lambda_i}{JL}(\mathbf{b}^i \cdot \mathbf{h}_0)\mathbf{b}^i + \mathcal{O}\Big(\frac{\lambda_i}{JL}\Big)^2. \tag{4.26}$$

Therefore, if $JL \gg \lambda_i$ we can approximate the fixed point by $\mathbf{x}_{r\alpha}^F = h_r/L$. We can now draw a sketch (see Fig.4.4) of the initial CLM dynamics referring to the eigensubspace analysis.

Suppose $h_r > 0$ for all r and the system is initialized with small positive random values $0 < x_{r\alpha}(0) \ll h_r$. For sufficiently large $JL \gg \lambda_i$ all eigenmodes in the DC-subspace have large negative eigenvalues. Therefore, the initial state will be rapidly projected onto the affine subspace $\sum_\alpha x_{r\alpha} = \sum_\alpha \mathbf{x}_{r\alpha}^F$ orthogonal to the DC-subspace which passes through the fixed point. The rapid projection dynamics in the DC space moves the activities away from the zero boundary and thus keeps the initial dynamics linear. Now, the modes in the AC-subspace with positive eigenvalues λ_i develop differences between layers on a slower timescale, since their eigenvalues have smaller absolute value. The principal AC-mode with the largest positive eigenvalue dominates the AC-subspace dynamics in this initial phase and determines the timescale of the WTA dynamics which drives some activities to zero in a column while others are increased. After some of the activities reach zero level, they no longer contribute to the linear dynamics and the nonlinearity is taking effect by introducing a new segment of the piecewise linear dynamics with different resulting eigenmodes. Since then the symmetry between layers is broken, the coupled WTA dynamics gets more complex in the general case, by subsequently driving the unassigned columns towards an assigned state as can be seen by comparison with Figure 4.3. For regular and highly symmetric patterns, however, which is discussed in the following section, the final grouping result can be already obtained from the dominant mode in the initial linear phase of the dynamics.

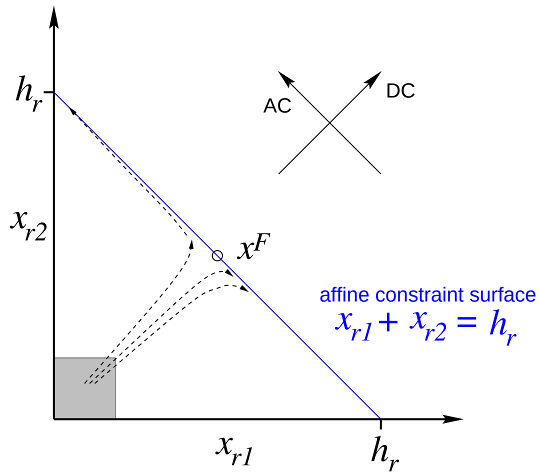


Figure 4.4: Sketch of the linear dynamics for two layers. Shown are the activity trajectories for the two activities x_{r1}, x_{r2} of a single column r. Starting from small initial values (grey square) the activities quickly approach the fixed point $\mathbf{x}^F$, which represents the "totally undecided" state, and the constraint surface $\sum_{\alpha=1,2} x_{r\alpha} = h_r$ in the DC subspace. Then the dynamics in the orthogonal AC subspace drives the WTA process until only one layer is active.

Now suppose one layer (i.e. for $\alpha = 1$) is changed into a ground layer with $\mathsf{f}^1_{rr'} = m\delta_{rr'}$ and the others remain unchanged $\mathsf{f}^{\alpha>1}_{rr'} = \mathsf{f}_{rr'}$ which results in a breaking of the symmetry between layers. However, the overall eigensubspace structure remains approximately the same. If we consider the lateral interactions as a first order perturbation to the vertical interaction component, we obtain the following (not normalized) eigenvectors with eigenvalues:

$$\mathbf{v}^{i1} \approx (\mathbf{b}^i, \ldots, \mathbf{b}^i), \qquad \Lambda_{i1} \approx -JL + \frac{(L-1)}{L}\lambda_i + \frac{m}{L}, \tag{4.27}$$

$$\mathbf{v}^{i2} \approx (-(L-1)\mathbf{b}^i, \mathbf{b}^i, \ldots, \mathbf{b}^i), \qquad \Lambda_{i2} \approx \frac{L-1}{L}m + \frac{\lambda_i}{L}, \tag{4.28}$$

$$\mathbf{v}^{i\gamma>2} = (\mathbf{0}^N, q_1^\gamma \mathbf{b}^i, \ldots, q_L^\gamma \mathbf{b}^i), \qquad \Lambda_{i\gamma} = \lambda_i, \tag{4.29}$$

where $\mathbf{0}^N$ is a zero vector of dimension N. The ground layer mainly causes a further subdivision among the AC modes, where the ground layer strength parameter m controls the relative activity growth of the ground layer versus the other layers.

The suggested mechanism of self-inhibitory annealing by choosing $\mathsf{f}'^{\alpha}_{rr'} = \mathsf{f}^{\alpha}_{rr} - T\delta_{rr'}$ shifts all eigenvalues by $\Lambda'_{i\gamma} = \Lambda_{i\gamma} - T$. If T is greater than the critical value $T_c = \lambda_1$, then all eigenvalues are negative and the unassigned fixed point $\mathbf{x}^F$ is asymptotically stable, which is analogous to Potts mean field models (Peterson & Soderberg 1989). If then T is lowered from the critical value, at first only the $\mathbf{b}^1$-AC modes acquire a positive eigenvalue and grow exponentially away from the fixed point. The following section discusses how this can be employed to increase the grouping quality.

4.2.4 Coherence, Separation, and Self-Inhibitory Annealing

Let us consider two simple stimulus examples for the contour grouping lateral interaction to discuss the binding problems of long-range coherence and separation referring to the

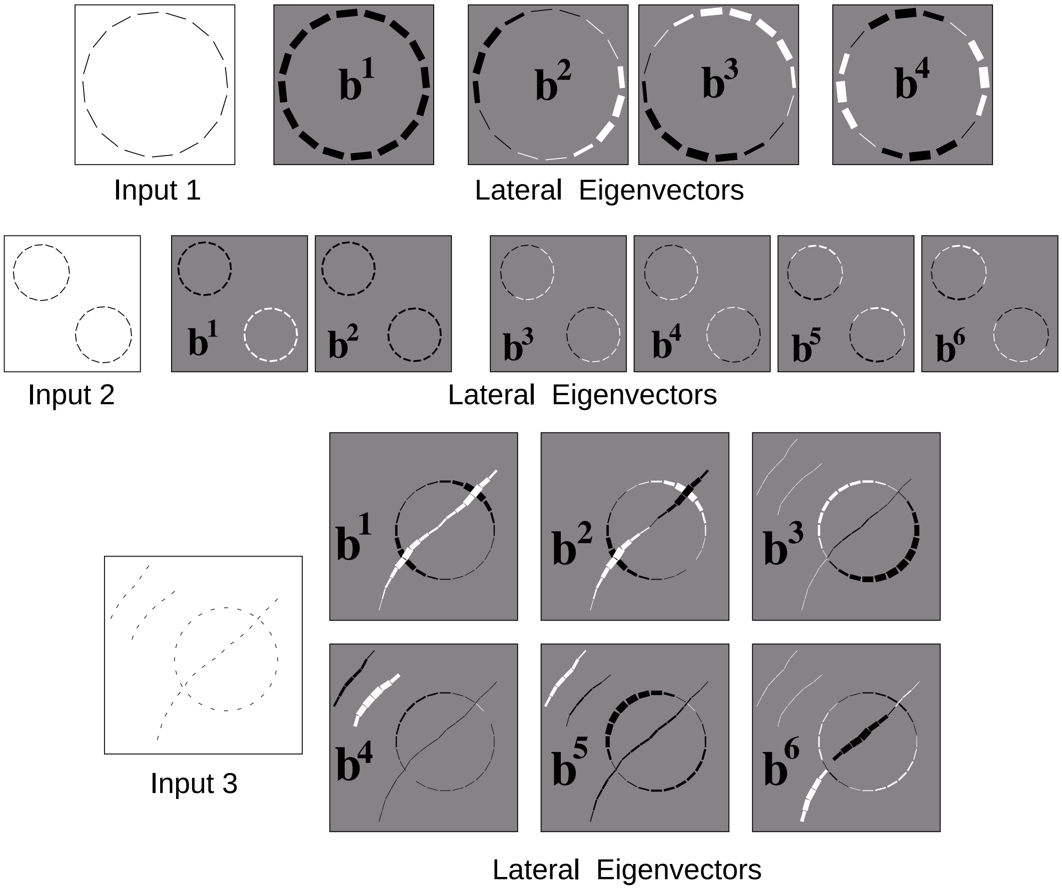


Figure 4.5: Lateral eigenvectors of the contour grouping interaction. The top row shows the lateral eigenvectors of the lateral interactions on the active features of this circular input pattern. Vector components are shown as black (positive) and white (negative) edges with thickness proportional to magnitude. To achieve a coherent binding on the circle, the eigenvalue λ_1 of the coherent mode $\mathbf{b}^1$ must be sufficiently large compared to the eigenvalues $\lambda_{2,3}$ of the circle-splitting modes $\mathbf{b}^{2,3}$. For two well-separated circles (middle row) the eigenvectors are symmetric and antisymmetric combinations of the single circle eigenvectors. Without global inhibition, the modes $\mathbf{b}^1$ and $\mathbf{b}^2$ are degenerate. A weak global inhibition suffices to lower the eigenvalue λ_2 and thus suppress the unwanted $\mathbf{b}^2$ mode. For more complex inputs (lower row) the succession of eigenvectors constitutes a multi-scale hierarchy. The principal eigenmode $\mathbf{b}^1$ separates the circle from the crossing line and the short lines. $\mathbf{b}^4$ and $\mathbf{b}^5$ drive the segregation of the short lines. The simple self-inhibitory annealing scheme proposed in section 4.1.2 can be interpreted as hierarchically suppressing subsequent eigenmodes resulting in a hierarchical group formation process.

previous eigensubspace analysis. The first stimulus consists of a circle, composed of N regularly placed edge elements (see Figure 4.5). We assume that we can neglect the effect of all other columns r' which are at subthreshold activity (i.e. $h_{r'} < 0$), and therefore consider the lateral eigenvectors only from the resulting $N \times N$ lateral interaction matrix $\mathsf{f}_{rr'}$ of the active features. The contour grouping lateral interaction has two contributions, a local cocircular interaction $f^{\mathrm{cocirc}}_{rr'}$ (see Appendix A) and a global inhibition with strength k, giving together $\mathsf{f}^{\alpha}_{rr'} = f^{\mathrm{cocirc}}_{rr'} - k$. Since the overall interaction pattern is rotationally invariant on the circle, the eigenvectors of the interaction are the sine and cosine waves on the circle with corresponding spatial frequencies.

First suppose the global inhibition is switched off by choosing $k = 0$. Each edge shares positive interactions in a local neighbourhood range and the zeroth order (constant) cosine with $b^1_r = 1$ is the eigenvector with largest eigenvalue $\lambda_1 = \sum_{r'} f^{\mathrm{cocirc}}_{rr'}$, equal to the sum over the local cocircularity values. The corresponding global AC modes coherently increase and decrease the activity on the circle in the different layers. The next two modes with degenerate eigenvalues $\lambda_{2,3} < \lambda_1$ are the first order sine and cosine waves, which split the circle into two halves. The gap between λ_1 and $\lambda_{2,3}$ determines whether the coherent mode dominates over the fragmenting modes. It is a general property of local interactions that this gap tends to zero, if the range of the interactions tends to zero.[3] Since the difference between the modes grows exponentially, also the time during which the system stays in the initial linear mode determines the degree of coherence. By the shift of the spectrum of the lateral interaction with the self-inhibitory addition $\mathsf{f}'_{rr'} = \mathsf{f}_{rr'} - T\delta_{rr'}$, this time is increased. This can be used to dynamically suppress the spurious $\mathbf{b}^{2,3}$ modes, however also reduces the overall convergence rate.

Now suppose we add a second equal and well-separated circle, such that there are no local cocircular interactions between the circles (see Fig. 4.5). Now the task is not only to achieve coherence on the circles, but also proper separation. Since the lateral interaction pattern is exchange-symmetric between the circles, the eigenvectors of the composite system are symmetric and antisymmetric combinations of the single circle eigenvectors with the same eigenvalues. The modes $\mathbf{b}^1$ and $\mathbf{b}^2$, responsible for grouping together the circles and separating them, respectively, are degenerate with $\lambda_1 = \lambda_2$ without any global inhibition. If we now add the global lateral inhibition by choosing $k > 0$, only the $\mathbf{b}^2$ mode undergoes a change in eigenvalue, since all other modes have a zero component within this global constant inhibition. The effect of k is then given by shifting the eigenvalue λ_2 of the spurious mode to the value $\lambda_2 - 2kN$. Therefore, if k is sufficiently large, the mode which separates the two circles coherently is dominant. There is, however, an obvious limitation to increasing k, since a strong global inhibition destabilizes the coherence within a circle. Similarly, for the single circle the inhibition lowers the eigenvalue of the single coherent mode by $\lambda'_1 = \lambda_1 - kN$. Therefore, to achieve both long-range coherence and separation only small values of k are possible, where an appropriate choice is $k = z/N$, with $0 < z \leq 1$.

The picture gets considerably more complex if there are more complex interacting groups in the feature input as the lower example in Figure 4.5 shows. The sign-distribution of the lateral eigenvectors constitutes a multi-scale hierarchy of the structures in the input. By

[3] Suppose the spatial exponential decay of the cocircular interaction is replaced by a hard cut-off function, such that the interaction of two perfectly cocircular features on the circle at angular positions $\phi_r, \phi_{r'} \in [0, 2\pi]$ is $\mathsf{f}_{rr'} = 1$ if $|\phi_r - \phi_{r'}| < D$ and $\mathsf{f}_{rr'} = 0$ otherwise. The difference between the coherent mode $b^1_r = 1$ and the first fragmenting modes $b^2_r = \sin(\phi_r), b^3_r = \cos(\phi_r)$ is in the limit of sufficiently small feature spacing given by $2(D - \sin(D)) \to 0$ for $D \to 0$.

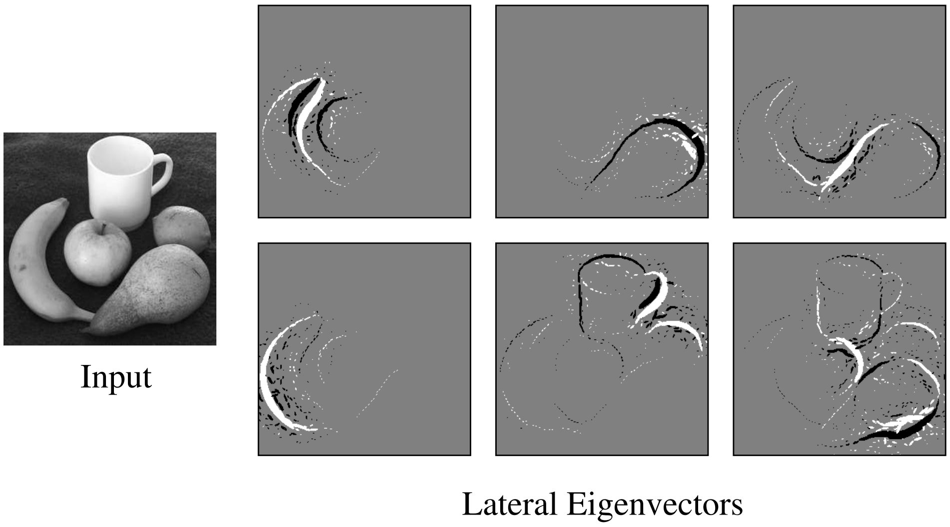


Figure 4.6: Eigenvectors of a contour grouping lateral interaction on edge features derived from a real image. The eigenvectors were numerically computed with a Lanczos algorithm (see Section 4.3.2). Vector components are shown as black (positive) and white (negative) edges with thickness proportional to magnitude. The eigenvectors are sorted according to their eigenvalues in descending order and the most salient contour groups correspond to the dominant eigenvectors.

gradually lowering the self-inhibition T, which can be understood as a type of annealing process with regard to the energy function of the dynamics, these modes are expressed in a hierarchical sequence, as can be seen by looking back at Figure 4.3. The dominant eigenvectors for a feature set derived from a real image are shown in Figure 4.6.

The eigensubspace analysis reveals an interesting parallel between the grouping dynamics in the CLM and the oscillatory synchronization in the contour integration model of Li (1998). If we restrict ourselves to lateral interactions which are only excitatory, then the dominant oscillation mode in that model has the same eigenvector components as the dominant grouping mode in the CLM. Therefore, a similar mechanism of suppressing spurious eigenmodes could be used to improve the synchronization properties in that model. Therefore a combination of synchronization-based and spatial mechanisms could provide a promising approach to more realistic models of cortical feature binding, which have enough computational power to be used for sensory segmentation tasks.

4.3 Simulation

4.3.1 General Algorithm

The CLM dynamics can be effectively simulated by the asynchronous stochastic update procedure for LT networks that was given in Section 3.5.2. The proposed mechanism of self-inhibitory annealing can be directly implemented in this scheme since it is only a modification of the self-coupling coupling weights which are gradually changed during iterations. The algorithm in conjunction with an exponential annealing schedule for the temperature parameter T can be implemented in the following way:

1. Initialize all $x_{r\alpha}$ with small random values around

$$x_{r\alpha}(t=0) \in [h_r/L - \epsilon, h_r/L + \epsilon].$$

Initialize T with $T = T_c$.

2. Do $N \cdot L$ times: Choose (r, α) randomly and update $x_{r\alpha} = \max(0, \xi)$, where

$$\xi = \frac{J_r h_r - \sum_{\beta \neq \alpha} I_r^{\alpha\beta} x_{r\beta} + \sum_{r' \neq r} \mathsf{f}_{rr'}^{\alpha} x_{r'\alpha}}{I_r^{\alpha\alpha} - \mathsf{f}_{rr}^{\alpha} + T} \quad .$$

3. Decrease T by $T := (1-\eta)T$, with $0 < \eta < 1$. Go to step 2 until convergence.

For simulation without self-inhibitory annealing simply set $T = 0$. The estimation of the critical temperature T_c is important with respect to the quality of the grouping and the speed of convergence. The optimal value is given by $T_c = \lambda_{\max}\{\mathsf{G}\}$. If T_c is estimated too small then the annealing is not fully effective since fragmenting grouping modes may be expressed. If it is estimated too large then the system remains in the high-temperature unassigned fixed point $x_{r\alpha} = h_r/L$ where the iterations are useless until $T < \lambda_{\max}\{\mathsf{G}\}$.

4.3.2 Algorithm for Gestalt-Based Perceptual Grouping

For the implementation of Gestalt grouping only lateral interactions are considered which are equal in all layers and thus of the form

$$\mathsf{f}_{rr'}^{\alpha} = \mathsf{f}_{rr'}. \tag{4.30}$$

In this case the largest eigenvalue is exactly given by $\lambda_{\max}\{\mathsf{G}\} = \lambda_{\max}\{\mathsf{f}_{rr'}\}$ (see Section 4.2.3). If additionally one layer is taken as a ground layer with

$$\mathsf{f}_{rr'}^{1} = m\delta_{rr'}, \qquad \mathsf{f}_{rr'}^{\alpha>1} = \mathsf{f}_{rr'}, \tag{4.31}$$

then the maximum eigenvalue can be estimated by $\lambda_{\max}\{\mathsf{G}\} \approx \max(m, \lambda_{\max}\{\mathsf{f}_{rr'}\})$. If the matrix $\mathsf{f}_{rr'}$ is large due to a large number N of features, the application of a standard eigensolver like the Jacobi algorithm (Press, Teukolsky, Vetterling, & Flannery 1992) which computes all eigenvalues and scales with $O(N^3)$ is very inefficient. If only some of the largest eigenvalues are required, the Lanczos method is a very powerful algorithm, which scales only with $O(KN) + O(KC(N))$, where K is the number of iterations and $C(N)$ is the computational cost of computing the product of the matrix $\mathsf{f}_{rr'}$ with an N-component vector (Golub & van Loan 1989). In the presented perceptual grouping applications in the following chapter a value of $K = 30$ is sufficient to obtain a very good estimate of $\lambda_{\max}\{\mathsf{f}_{rr'}\}$, and the interaction matrices are effectively[4] sparse for local interactions, leading to $C(N)$ being of the order of $O(N^{1/2})$. This makes it possible to estimate the optimal value of T_c even for simulations with $N = 20000$ features within a second on a standard Pentium III personal computer.

[4]The globally constant inhibitory component of the perceptual grouping interactions needs to be evaluated only once for each layer since it is equal for all neurons.

The Theorems 3 and 4 provide two alternatives for choosing the appropriate self-coupling to ensure convergence of the CLM. For the application to Gestalt grouping the vertical interactions were equally chosen as $J = J_r = I_r^{\alpha\beta}$ for all r, α, β. The two conditions for convergence can then be satisfied by

$$\text{i)} \quad J > J^c = \max\big(m, \sum_{r'} \max(0, \mathsf{f}^{\alpha}_{rr'})\big) \quad \text{or} \tag{4.32}$$

$$\text{ii)} \quad J > J^c = \max\big(m, \lambda_{\max}\{\mathsf{f}_{rr'}\}\big), \tag{4.33}$$

respectively. If no eigenvalues are computed, because the CLM is operated without annealing, then i) can be easily evaluated. If the maximum eigenvalue is available, then condition ii) usually allows to operate the CLM closer to the critical value of J where divergence occurs, as the later application results show. The results also demonstrate that this leads to a considerable gain in the grouping quality. If now annealing is done by performing the diagonal shift with $\mathsf{f}^{\alpha}_{rr'}{}' = \mathsf{f}^{\alpha}_{rr'} - T\delta_{rr'}$, the critical value J^c is shifted accordingly to the value $J^{c\prime} = J^c - T$. Therefore, to use the smallest possible value of J for a particular value of T to produce strong contextual effects, one should choose $J(T) = \mu J^c - T$ with $\mu = 1 + \epsilon$, $0 < \epsilon \ll 1$ and adapt J to the particular value of T during annealing. This can be summarized to the following simulation algorithm for a CLM grouping model with strong contextual effects:

1. Compute $\lambda_{\max}\{\mathsf{f}_{rr'}\}$.
   Initialize T with $T = \max\big(m, \lambda_{\max}\{\mathsf{f}_{rr'}\}\big)$.
   Set $J = \mu J^c = \mu \max\big(m, \lambda_{\max}\{\mathsf{f}_{rr'}\}\big)$, where $\mu = 1 + \epsilon$ with $0 < \epsilon \ll 1$
   Initialize all $x_{r\alpha}$.
2. Do $N \cdot L$ times: Choose (r, α) randomly and update
   $x_{r\alpha} = \max(0, \xi)$, where
   $$\xi = \frac{(J - T)\big(h_r - \sum_{\beta \neq \alpha} x_{r\beta}\big) + \sum_{r' \neq r} \mathsf{f}^{\alpha}_{rr'} x_{r'\alpha}}{J - \mathsf{f}^{\alpha}_{rr}} .$$
3. Decrease T by $T := \eta T$, with $0 < \eta < 1$. Go to step 2 until convergence.

Stated in this form, the update step 2 already incorporates the temperature-dependent shift for J.

4.4 General Labeling Problems

For the application of the CLM to feature binding only lateral interactions within layers were considered. This was motivated by the structure of the perceptual grouping problem, where the layer index only represents membership information for a particular group. In a labeling problem which was defined in Section 2.3.1 for the discussion of relaxation labeling algorithms, the value of the label index is relevant. Therefore, the compatibilities are also defined between different label alternatives of a pair of features. For the CLM this means that also interlateral interactions are allowed with $\mathsf{f}^{\alpha\beta}_{rr'} \neq 0$ for $\alpha \neq \beta$. The generalization of the strict layered organization to arbitrary couplings between labels then leads to a system of coupled winner-take-all columns, for which similar assignment and convergence properties are established in this section.

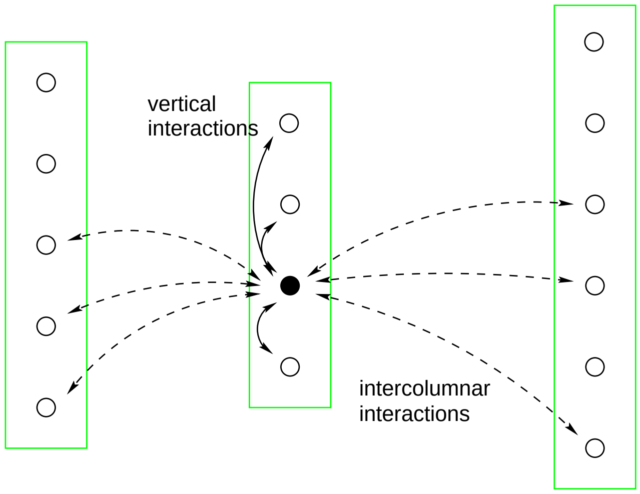


Figure 4.7: Vertical interactions within columns and intercolumnar interactions between columns.

4.4.1 CLM for General Labeling Problems

Suppose we are given a set of N columns $r = 1, \ldots, N$ which are composed of L_r neural activities $x_{r\alpha}$ each, where $\alpha = 1, \ldots, L_r$. The activity $x_{r\alpha}$ represents the certainty of the assignment of label α to feature r, where each feature may allow a different number L_r of labels. We assume that there are two types of interactions between neurons, intracolumnar or vertical interactions within columns and intercolumnar interactions between neurons in different columns (see Figure 4.7). The weight of the competitive vertical interaction between neurons $x_{r\alpha}$ and $x_{r\beta}$ is given by $I_r^{\alpha\beta} = I_r^{\beta\alpha} > 0$. The intercolumnar interaction between neuron $x_{r\alpha}$ and $x_{r'\beta}$ is given by the symmetric weight $\mathsf{f}_{rr'}^{\alpha\beta} = \mathsf{f}_{r'r}^{\beta\alpha}$. The weight coefficients are to be interpreted as the mutual compatibility of assigning label α to feature r and label β to feature r'. Except for the self-coupling components $\mathsf{f}_{rr}^{\alpha\alpha}$, the intercolumnar interactions are assumed to have no intracolumnar contribution, i.e. $\mathsf{f}_{rr}^{\alpha\beta} = 0$ for all r and $\alpha \neq \beta$. The corresponding LT dynamics is given by

$$\dot{x}_{r\alpha} = -x_{r\alpha} + \sigma\Big(J_r h_r - \sum_{\beta} I_r^{\alpha\beta} x_{r\beta} + \sum_{r'\beta} \mathsf{f}_{rr'}^{\alpha\beta} x_{r'\beta} + x_{r\alpha}\Big), \tag{4.34}$$

where we have adopted the convention that the combined summation over r' and β is carried out only over feasible pairs (r', β) with $\beta \leq L_{r'}$ (the same convention holds for summation over r, α). The additional self-excitatory term simplifies the following computations and can be compensated for by taking $\mathsf{f}_{rr'}^{\alpha\beta\,\prime} = \mathsf{f}_{rr'}^{\alpha\beta} - \delta_{\alpha\beta}\delta_{rr'}$. The energy function is of the form

$$E = -\sum_{r\alpha} J_r h_r x_{r\alpha} + \frac{1}{2}\sum_{r}\sum_{\alpha\beta} I_r^{\alpha\beta} x_{r\alpha} x_{r\beta} - \frac{1}{2}\sum_{r\alpha}\sum_{r'\beta} \mathsf{f}_{rr'}^{\alpha\beta} x_{r\alpha} x_{r'\beta}. \tag{4.35}$$

The interaction structure of the layer model presented in the previous sections of this chapter is a special case of the arrangement of WTA columns, where all columns have the same dimension with $L_r = L$ for all r, and the lateral interactions are restricted to the same label and layer with $\mathsf{f}_{rr'}^{\alpha\beta} = \delta_{\alpha\beta}\mathsf{f}_{rr'}^{\alpha}$.

4.4.2 Convergence and Assignment

The convergence conditions for the layered CLM architecture in Section 4.2.1 can be generalized in a straightforward way for this more general interaction pattern.

Theorem 6. *The generalized CLM dynamics (4.34) is bounded and convergent if*

i) $I_r^{\alpha\alpha} > \mathsf{f}_{rr}^{\alpha\alpha} + \sum_{(r'\beta)\neq(r\alpha)} \max(0, \mathsf{f}_{rr'}^{\alpha\beta})$ *or*

ii) $I_r^{\alpha\alpha} > \lambda_{\max}\{\mathsf{f}_{rr'}^{\alpha\beta}\}$.

Proof. The two conditions are obtained by omitting the vertical interactions $I_r^{\alpha\beta}$ for $\alpha \neq \beta$ from the global interaction matrix, like in the proofs of Theorem 3 and Theorem 4. □

In the following sufficient conditions are stated on the lateral and vertical interactions that ensure the convergence to an unambiguous state, with only one active neuron within a column. The intercolumnar feedback $F_{r\alpha}$ of a neuron $x_{r\alpha}$ is defined by $F_{r\alpha} = \sum_{r'\beta} \mathsf{f}_{rr'}^{\alpha\beta} x_{r'\beta}$. The proof of this theorem also includes the proof of Theorem 5 in Section 4.2.2, which is a special case of the more general intercolumnar coupling scheme considered here.

Theorem 7. *If the lateral interaction is self-excitatory,* $\mathsf{f}_{rr}^{\alpha\alpha} > 0$ *for all* r, α*, and the vertical interactions satisfy* $I_r^{\alpha\alpha} I_r^{\beta\beta} \leq (I_r^{\alpha\beta})^2$ *for all* α, β *then an attractor of the CLM has in each column* r *either*

i) at most one positive activity $x_{r\hat{\alpha}}$ *with*

$$x_{r\hat{\alpha}} = \frac{J_r}{I_r^{\hat{\alpha}\hat{\alpha}}} h_r + \frac{F_{r\hat{\alpha}}}{I_r^{\hat{\alpha}\hat{\alpha}}}, \quad x_{r\beta} = 0 \text{ for all } \beta \neq \hat{\alpha}, \tag{4.36}$$

where $\hat{\alpha} = \hat{\alpha}(r)$ *is the index of the maximally supported neuron in column* r *characterized by* $F_{r\hat{\alpha}} > F_{r\beta}$ *for all* $\beta \neq \hat{\alpha}$ *or*

ii) all activities $x_{r\alpha}$*,* $\alpha = 1, \ldots, L$ *in a column* r *vanish and* $F_{r\alpha} \leq -J_r h_r$ *for all* $\alpha = 1, \ldots, L$.

Proof. Suppose an equilibrium $\mathbf{x}^F$ has two positive activities $x_{r\alpha,\beta}^F > 0$ in a column r at two positions α and β. Due to the form of the linear threshold dynamics, the fixed point condition $\dot{x}_{r\alpha,\beta}^F = 0$ can be only satisfied for positive activities $x_{r\alpha,\beta}^F > 0$ if the condition $\frac{\partial E}{\partial x_{r\alpha,\beta}} = 0$ is satisfied. Now consider a small perturbation within this column of the form $\epsilon_{r\alpha} = \eta_\alpha,\ \epsilon_{r\beta} = \eta_\beta,\ \epsilon_{r'\alpha'} = 0$ for all other r', α'. Expanding the quadratic function E about the equilibrium we obtain for ΔE under this perturbation with all linear components vanishing at the equilibrium

$$2\Delta E = \sum_{rr'\alpha\beta} (I_r^{\alpha\beta}\delta_{rr'} - \delta_{\alpha\beta}\mathsf{f}_{rr'}^{\alpha\beta})\epsilon_{r\alpha}\epsilon_{r'\beta} \tag{4.37}$$

$$= I_r^{\alpha\alpha}\eta_\alpha^2 + I_r^{\beta\beta}\eta_\beta^2 + 2I_r^{\alpha\beta}\eta_\alpha\eta_\beta - \mathsf{f}_{rr}^{\alpha\alpha}\eta_\alpha^2 - \mathsf{f}_{rr}^{\beta\beta}\eta_\beta^2 \tag{4.38}$$

$$\leq I_r^{\alpha\alpha}\eta_\alpha^2 + I_r^{\beta\beta}\eta_\beta^2 + 2I_r^{\alpha\beta}\eta_\alpha\eta_\beta \ . \tag{4.39}$$

If we can state a perturbation $(\eta_\alpha, \eta_\beta)$ for which $\Delta E < 0$ then $\mathbf{x}^F$ can be no local minimum and therefore be no attractor of the gradient-descent dynamics. The quadratic (2×2) form (4.39) has at least one negative eigenvalue λ_1 or λ_2 if $\lambda_1\lambda_2 = I_r^{\alpha\alpha}I_r^{\beta\beta} - (I_r^{\alpha\beta})^2 < 0$. Then, however, exists a perturbation $(\eta_\alpha, \eta_\beta)$ with $\Delta E < 0$. If equality holds, $I_r^{\alpha\alpha} I_r^{\beta\beta} = (I_r^{\alpha\beta})^2$ then $\lambda_1 = 0$ or λ_2=0 and we can state a perturbation from the null-space for which then $\Delta E < 0$ due to $\mathsf{f}_{rr}^{\alpha\alpha}, \mathsf{f}_{rr}^{\beta\beta} > 0$. So, we have shown by contradiction that an attractor can have at most one positive activity in a column.

Now consider case $i)$: Let $\hat{\alpha}$ denote the layer index of the positive activity in column r. The constraint is then inactive for $x_{r\hat{\alpha}}$ and active for $x_{r\beta\neq\hat{\alpha}}$. Hence

$$0 = E_{r\hat{\alpha}} = J_r h_r - I_r^{\hat{\alpha}\hat{\alpha}} x_{r\hat{\alpha}} + F_{r\hat{\alpha}}, \tag{4.40}$$

$$0 > E_{r\beta\neq\hat{\alpha}} = J_r h_r - I_r^{\hat{\alpha}\hat{\alpha}} x_{r\hat{\alpha}} + F_{r\beta\neq\hat{\alpha}}, \tag{4.41}$$

which proves case $i)$. Case $ii)$ gives just the remaining possible stable equilibria which are not covered by $i)$. □

We can interpret Theorems 6 and 7 together as a prescription how to choose the competitive vertical interactions for a given set of compatibility or interaction coefficients $\mathsf{f}_{rr'}^{\alpha\beta}$. As a first step the vertical component $I_r^{\alpha\alpha}$ of the self-coupling interaction must be chosen either larger than the sum over positive intercolumnar interactions or larger than the maximum eigenvalue of the intercolumnar interaction matrix. In the second step the intracolumnar competitive interactions $I_r^{\alpha\beta}$ for $\alpha \neq \beta$ must be chosen larger or equal to the self-coupling components $I_r^{\alpha\alpha}$.

4.4.3 Relation to other Labeling Approaches

The CLM energy with the constant vertical interaction scheme $I_r^{\alpha\beta} = J_r = J$ for all r, α, β and equal inputs $h_r = 1$ for all r is given by

$$E^{\mathrm{CLM}} = \frac{1}{2} \sum_r \Big(1 - \sum_\beta x_{r\beta}\Big)^2 - \frac{1}{2} \sum_{r\alpha} \sum_{r'\beta} \mathsf{f}_{rr'}^{\alpha\beta} x_{r\alpha} x_{r'\beta}. \tag{4.42}$$

In this form, the CLM energy function can be compared to other labeling approaches.

Relaxation Labeling

The relaxation labeling energy corresponding to the CLM energy (4.2) is given by

$$E^{\mathrm{RL}} = -\frac{1}{2} \sum_{r\alpha} \sum_{r'\beta} \mathsf{f}_{rr'}^{\alpha\beta} x_{r\alpha} x_{r'\beta} \tag{4.43}$$

subject to the "hard" weighted labeling constraint $\sum_\alpha x_{r\alpha} = 1$. This constraint is replaced by a softly enforced constraint in the CLM, where the constraint multiplier J controls the strength compared to the lateral interactions. According to Theorem 7 a stable attractor of the CLM satisfies

$$x_{r\hat{\alpha}} = 1 + \frac{F_{r\hat{\alpha}}}{J}, \tag{4.44}$$

where $\hat{\alpha}$ is the index of the maximally supporting layer with $F_{r\hat{\alpha}} > F_{r\beta}$ for all $\beta \neq \hat{\alpha}$. If J is sufficiently large, then $x_{r\hat{\alpha}} \approx 1$ and any stable state of the CLM is approximately equal to a consistent labeling in the sense of relaxation labeling, i.e. has maximal linear support (see Section 2.3.1). Therefore the CLM dynamical system and its simulation algorithms can be used to obtain feasible solutions to relaxation labeling problems if J is sufficiently large.

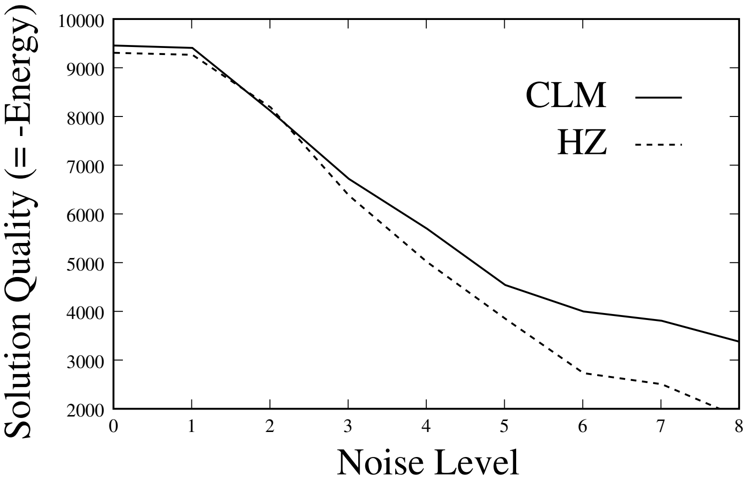


Figure 4.8: Comparison of labeling performance for the relaxation labeling algorithm by Hummel & Zucker (HZ) and CLM.

A comparison of different relaxation labeling algorithms on a benchmark problem was recently done by Li, Soh, & Teoh (1998). They suggested a problem of weighted graph matching for the generation of a set of compatibilities and evaluated the performance of the algorithms by the energy levels of the obtained labeling solutions. The problem setup is as follows: A set of M points is generated randomly with uniform distribution in a square of 100×100. Then, a second set of M points is generated, where 2/3 of the points are given by adding Gaussian noise of $N(0, \sigma^2)$ to points in the first set and 1/3 is chosen random uniformly distributed in the box. The problem is to establish the optimal node correspondences between the two weighted graphs given by the points as nodes and their distances as weights between nodes. The activity $x_{r\alpha}$ corresponds to the certainty of assigning node r of the first set to node α of the second set, where an additional label $\alpha = 0$ is introduced to cater the matching of outlier nodes. The compatibilities are defined by

$$\mathsf{f}_{rr'}^{\alpha\beta} = \begin{cases} 1 & \text{if } r = r' \text{ and } \alpha = \beta, \\ \max\left(-\left[d(r,r') - d(\alpha,\beta)\right]^2 ,\ -255\right) & \text{if } \alpha \neq 0 \text{ and } \beta \neq 0, \\ -k & \text{otherwise ,} \end{cases} \tag{4.45}$$

where $d(r, r')$ and $d(\alpha, \beta)$ denote the euclidian distances of the corresponding points and $k = 0.3 * 255$ in our simulations. We have compared the performance of the CLM with the algorithm of Hummel & Zucker (1983), which is one of the best RL schemes according to the results of Li et al., however, requires an elaborate gradient projection operation. The performance is evaluated for different noise levels σ by the energy value of the converged state averaged over 200 different point configurations where $M = 18$. Figure 4.4.3 shows the solution quality plotted as negative energy of the output configuration shifted by a common constant for CLM and HZ. With a comparable number of iterations for the algorithm outlined in the previous section, the CLM achieves consistently lower average energies for larger noise levels.

Markov Random Fields

Let $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}(1), \dots, \hat{\alpha}(N))$ denote an assignment of labels to all features. The probability distribution of a Markov random field is given by

$$P(\hat{\boldsymbol{\alpha}}) = \frac{1}{Z} e^{-E(\hat{\boldsymbol{\alpha}})/T}, \tag{4.46}$$

where Z is a normalizing factor, T is a temperature which controls the flatness of the distribution and $E(\hat{\boldsymbol{\alpha}})$ is the energy value of the configuration $\hat{\boldsymbol{\alpha}}$. If T is sufficiently low, then the distribution peaks around the local minima of the energy function. A common approach (Posch & Schlüter 1998) is to consider pairwise and singleton cliques, where E is of the form

$$E^{\mathrm{MRF}}(\hat{\boldsymbol{\alpha}}) = \frac{1}{2} \sum_r V_r^{(1)}\big(\hat{\alpha}(r)\big) + \frac{1}{2} \sum_{r' \neq r} V_{rr'}^{(2)}\big(\hat{\alpha}(r), \hat{\alpha}(r')\big). \tag{4.47}$$

The singleton potential $V_r^{(1)}(\alpha)$ describes the probability of assigning label α to feature r and the pairwise clique potential $V_{rr'}^{(2)}(\alpha, \beta)$ characterizes the mutual compatibility of assigning label α to feature r and label β to feature r'. Let $\mathbf{x}$ be the corresponding assignment vector for the labeling given by $\hat{\boldsymbol{\alpha}}$ with

$$x_{r\alpha} \approx 1 \text{ if } \alpha = \hat{\alpha}(r), \quad x_{r\alpha} = 0 \text{ if } \alpha \neq \hat{\alpha}(r). \tag{4.48}$$

For this state, the vertical coupling terms can be neglected in the CLM energy to obtain

$$E^{\mathrm{CLM}}(\mathbf{x}) \approx \frac{1}{2} \sum_{r\alpha} \mathsf{f}_{rr}^{\alpha\alpha} x_{r\alpha}^2 + \frac{1}{2} \sum_{r\alpha} \sum_{(r'\beta) \neq (r\alpha)} \mathsf{f}_{rr'}^{\alpha\beta} x_{r\alpha} x_{r'\beta} \tag{4.49}$$

$$= \frac{1}{2} \sum_r \mathsf{f}_{rr}^{\hat{\alpha}(r)\hat{\alpha}(r)} + \frac{1}{2} \sum_{r' \neq r} \mathsf{f}_{rr'}^{\hat{\alpha}(r)\hat{\alpha}(r')}. \tag{4.50}$$

Therefore, if we identify the corresponding Markov random field potentials with the compatibilities of the CLM according to

$$V_r^{(1)}(\alpha) = \mathsf{f}_{rr}^{\alpha\alpha}, \qquad V_{rr'}^{(2)}(\alpha, \beta) = \mathsf{f}_{rr'}^{\alpha\beta}, \tag{4.51}$$

then the minima of the Markov random field energy are approximately equal to the corresponding CLM energy minima.

4.5 Summary

The competitive layer model was introduced in this chapter as a recurrent neural network model for spatial feature binding. The binding of features is represented in the CLM by the layer-wise coactivation of feature-representing neurons in a system of neural layers, driven by a set of topographically organized competitive and cooperative interactions. The dynamics is formulated as an additive linear threshold recurrent network. Contextual relations among features are coded by pairwise compatibilities which define an energy function to be minimized by the neural dynamics. The general conditions for non-divergence of linear threshold networks derived in Chapter 3 were used to establish conditions on the vertical competitive interactions which ensure convergence of the CLM. It was shown that for appropriate self-coupling an attractor of the CLM has only one active layer for each feature, which ensures an unambiguous assignment of the features to the layers. The eigensubspace analysis of the linearized dynamics was used to discuss the properties of the binding dynamics for some simple input stimuli and to motivate a mechanism of self-inhibitory annealing for improving the grouping quality. The effective simulation of the CLM was discussed with a description of appropriate choices of the parameters for achieving strong contextual modulation in the model. The extension of the CLM to more general coupling schemes for the solution of general labeling problems was introduced and the corresponding conditions for convergence and assignment were established. The comparison to relaxation labeling and Markov random fields revealed that the energy function of the CLM is similar to the effective energy functions which are minimized in those two approaches, if the contextual modulation in the CLM is suppressed by strong vertical interactions.

Chapter 5

Perceptual Grouping with the CLM

This chapter presents the application of the competitive layer model architecture to the implementation of a number of Gestalt-based perceptual grouping laws. First, the grouping of contours according to the principle of continuity is considered by using oriented edge elements as features. A mechanism for combined figure-ground segmentation and grouping is introduced which is based on choosing one layer as a background layer with pure self-coupling. The background layer attracts noisy features which are not part of a salient group. The effects of the most important parameters of the model are explored and the results are compared to simulations with a Potts spin model with same lateral coupling scheme. The results show that operating the CLM in a regime of strong contextual modulation, characterized by the conditions derived in Chapter 3, improves the overall grouping performance. The contour grouping approach was extended by Nattkemper et al. (2000a) to the segmentation of cell images obtained from fluorescence microscopy and the results of that study are summarized. A set of artificially generated images is used to demonstrate how the layered architecture of the CLM can be used to solve the aperture problem of motion segmentation. For the application to greyscale segmentation the features are chosen as image pixels characterized by position and intensity, and the lateral interactions are based on the local intensity differences in the image. The segmentation results are compared to other recent binding models. As a final application, the main results of the study on texture segmentation with the CLM by Ontrup & Ritter (1998) are summarized. The chapter concludes with a discussion of the CLM binding approach and its biological relevance.

5.1 Contour Grouping

It is long known that in the early stages of visual information processing in visual cortex area V1 many neurons can be found that respond to local oriented edge elements within their classical receptive field (Hubel & Wiesel 1962). One of the major issues in vision research are the mechanisms which are used by the visual system to integrate these local elements into global salient contours to facilitate robust boundary detection and object recognition. This process has been considered (Sajda & Finkel 1994; Zucker, Dobbins, & Iverson 1989) as being composed of two components: the process of enhancement, where local edge information is combined cooperatively for smooth and salient contours, and the process of segmentation or grouping, where separate contour elements are assigned to different groups.

Recent neural models of contour integration have mainly focused on the enhancement stage of visual processing. Many models of orientation tuning are composed of competitive orientational "On-Center-Off-Surround" interactions within hypercolumns (Ben-Yishai, Lev Bar-Or, & Sompolinsky 1995; Mundel, Dimitrov, & Cowan 1997), that are modulated by experimentally found horizontal or lateral long range interactions (Gilbert 1992; Weliky, Kandler, Fitzpatrick, & Katz 1995) to produce contextual effects. Enhancement through feature linking is addressed by the model of Yen & Finkel (1996, 1998), which, however, requires rule-based interactions and global activity normalizations. Li (1998) has stated a more biologically plausible dynamical model for contour integration in area V1 and discusses properties of the oscillatory correlations in the model which may be useful to facilitate synchronization-based segmentation. As the results of Li (1998) demonstrate, the complex excitatory and inhibitory interactions lead to distorted and less predictable correlations for more complex visual scenes, leaving the question of a robust grouping process mostly unanswered.

In the following the application of the CLM is presented as a new approach to the grouping stage of contour integration. To keep the model simple and to allow for the application to large feature sets of real images, the generation of local orientation within a hypercolumn is not modelled and left to a preprocessing step. The essential WTA then does not operate between local orientation alternatives within a hypercolumn, but between different grouping alternatives of the edge feature with fixed local orientation.

5.1.1 The CLM Contour Grouping Model

Feature Generation

Suppose that in an initial integration step the local edge information has been subsampled and preprocessed. As a result of the preprocessing, it is assumed that within a localized area only a single edge detector is active at an optimally tuned orientation. Therefore a set of idealized edge feature detectors is considered, indexed by $r \in \{1, \dots, N\}$ and characterized by position $\mathbf{p}_r = (p_r^x, p_r^y)$, and local orientation represented by a unit orientation vector $\mathbf{n}_r = (n_r^x, n_r^y)$ with $\mathbf{n}_r^2 = 1$ (see Figure 5.1). The set of N edge features can be directly generated for a synthetic feature setup. For the application to real-world images a particular subsampling algorithm is used to limit the number of features which have to be simulated.

For a real image the set of edge features is generated by subdividing the image into non-overlapping small subrectangles. On each subrectangle 3x3 pixel Sobel x and y operators are applied which sample the intensity gradient information at orthogonal directions. Within each subrectangle an edge feature indexed by r is chosen at the position of maximum squared Sobel response. The unit orientation vector $\mathbf{n}_r$ is chosen orthogonal to the direction of the normalized Sobel x and y components and is thus aligned with the estimated local edge element. The input intensity h_r of feature r is given by the squared Sobel response added by a small positive constant. This slightly raising of the zero level improves the ability to enhance low input contours through lateral effects. The constant was chosen as ten percent of the maximum edge intensity.

Lateral Interactions and Simulation

The CLM contour grouping model (Wersing & Ritter 1999b) is composed of a set of L layers $\alpha \in 1, \dots, L$ where $L - 1$ "figure" layers are provided to respond to salient contour groups and a single "ground" layer is provided to capture the background features. The

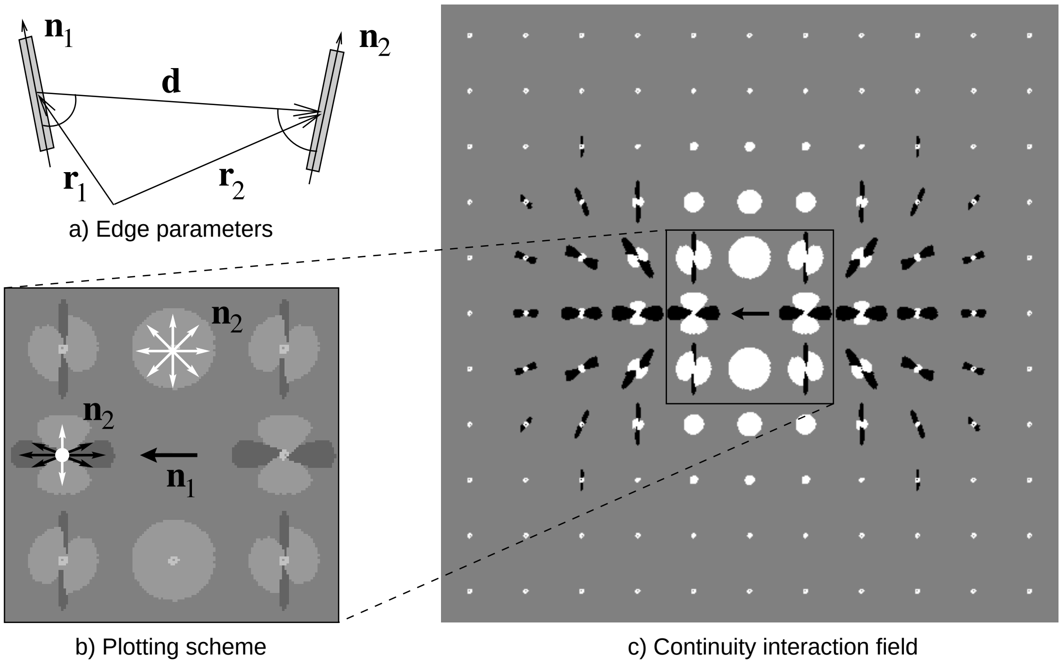


Figure 5.1: Lateral interaction for contour grouping. a) The pairwise interaction depends on the difference vector $\mathbf{d}$ and two unit vectors $\mathbf{n}_1, \mathbf{n}_2$ encoding directed orientation for computational convenience. b) shows a magnification of the plot c) of the interaction field of a single horizontal edge pointing to the left. The plot is generated by computing the interaction of the central directed edge with directed edges of all directions (like a cylindrical plot) at a spatial grid. Black edges share excitatory, white edges share inhibitory interaction with the central edge and length codes for interaction strength. The continuity field depends on position and orientation but is not direction-selective. It supports pairs of edges which are cocircular, i.e. lie tangentially to a common circle.

lateral interactions $\mathsf{f}^{\alpha}_{rr'}$ are given as

$$\mathsf{f}^{1}_{rr'} = m\delta_{rr'}, \qquad \mathsf{f}^{\alpha>1}_{rr'} = f^{\text{cocirc}}_{rr'} - k \tag{5.1}$$

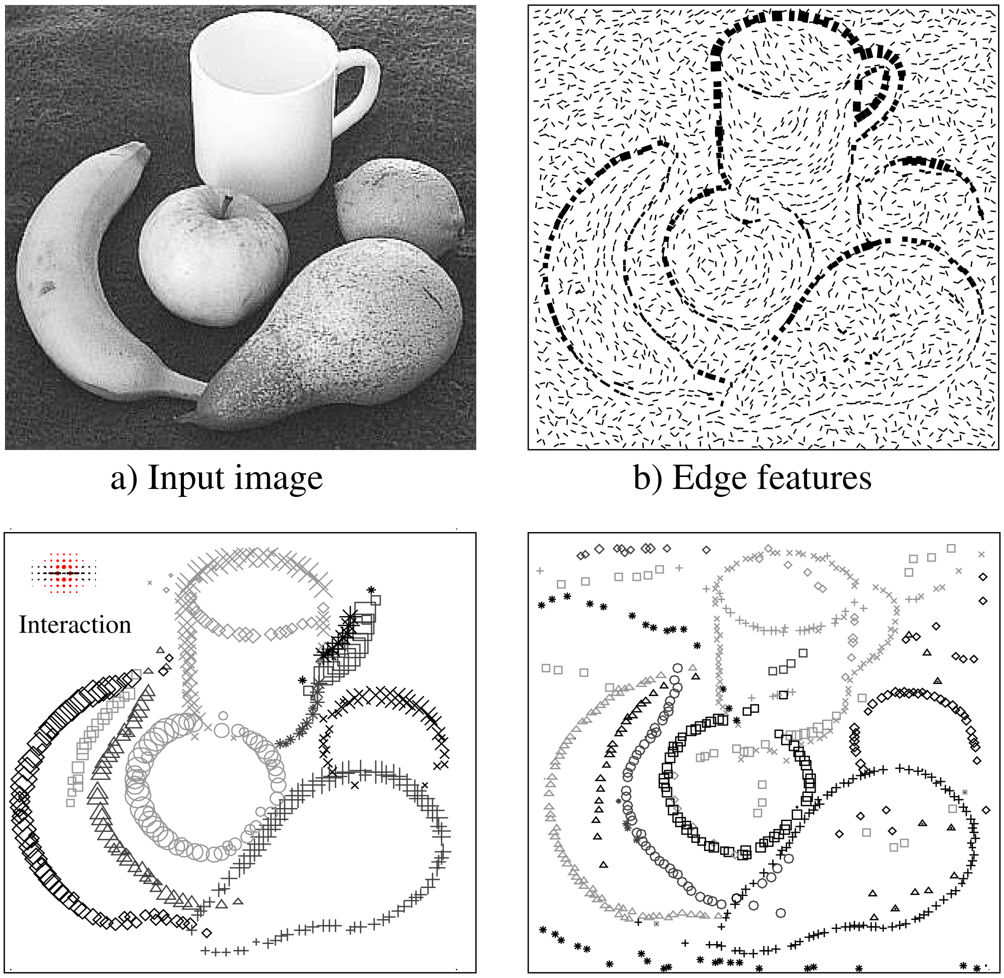
The parameter $m > 0$ defines a self-coupling against which lateral interactions in the figure layers must compete to “pop out” a feature from the ground layer. The parameter k introduces a weak global inhibition to support the separation of remote contour parts. The pairwise lateral interaction $f^{\text{cocirc}}_{rr'}$ between edge features in the figure layers is visualized in Figure 5.1 and formally defined in Appendix A. The excitatory component links edges that are co-circular, i.e. lie tangentially to a common circle, with a tolerance that admits small deviations. This orientational field is superimposed with a Gaussian distance-dependent component. This excitatory interaction field is similar to recent models (Parent & Zucker 1989; Yen & Finkel 1998; Li 1998). A similar interaction pattern between orientation sensitive cells has been observed experimentally (Gilbert, 1992; Weliky et al., 1995). The local inhibitory component is of comparable magnitude as the excitatory field and insensitive to orientation as suggested by experimental findings (DeAngelis et al., 1994). The weak long-range inhibition has no direct biological counterpart in the experimentally observed lateral connection structure, however, Wang & Terman (1997) have speculated about the thalamus providing such a mechanism of global inhibition.

The simulations were carried out with the stochastic update algorithm formulated in Section 4.3.2 with annealing. The vertical interactions and input strength were chosen equally as $J_r = I_r^{\alpha\beta} = J$ for all r, α, β with $J = \mu J_c = \mu \max\left(m, \lambda_{\max}\{\mathsf{f}_{rr'}\}\right)$, where J_c is the critical value implied by the stability margins of Theorem 4 and $\mu \geq 1$.

5.1.2 Contour Grouping on Real Images

The application of the CLM contour grouping model to a real image is shown in Figure 5.2. The preprocessing results in a set of features where object and background textures produce a noisy background. To obtain a robust grouping for the complex scene of interacting edge elements, annealing in T was necessary, where an exponential schedule was used as described in Section 4.3 with $\eta = 0.001$. The grouping result is shown in Fig. 5.2c. By an appropriate choice of the ground layer strength $m = 3.5$, which must approximately match the sum over the cocircular interactions of an edge that is part of a proper contour, the salient contours are detected as single groups. Since noisy edges lack the lateral support of aligned edges, they are then captured by the ground layer and for sufficiently many available layers only those containing salient segments will be active. The lateral interactions cause an amplification of salient low-intensity contours without enhancing noisy fragments. To demonstrate that this amplitude-dependent modulation is crucial in the model, the CLM grouping result is compared to a Potts spin mean field model with equal lateral interaction structure and annealing. Spin models generally lack the ability to represent additional amplitude information. The result in Fig. 5.2d illustrates that this makes them more sensitive to noise in the image, because noisy aligned elements may form spurious groups. Since we were not able to produce significantly better results also by varying parameters m and k it can be concluded that the amplitude-dependent dynamics in the CLM leads in this setting to a better signal-noise separation for complex input patterns.

Figure 5.3 shows the performance of the CLM with the same parameter settings and identical annealing on more complex image data, where for an aerial image the number of layers is not sufficient to allow a single layer for each salient segment. In that case the dynamics tries to find a compromise by combining segments into the same layer such that their mutual inhibition is minimal and thus the overall energy level is minimal. This results in a tendency to prefer a combination of parallel segments, which are separated by a distance larger than the local inhibition field of the cocircular interaction. In principle, it would be possible to suppress multiple segments within a layer by raising the global inhibition. This, however, in turn poses an upper limit on the maximum size of a stable segment, since then the global inhibition may exceed the local cocircular support. The time course of the dynamics during self-inhibitory annealing illustrates the resulting hierarchical group formation process which initially expresses the most salient groups in the available layers.



a) Input image b) Edge features

c) CLM grouping d) Potts spin grouping

Figure 5.2: Grouping of a natural image. An input image a) is preprocessed to generate a set of $\approx$2000 edge features displayed in b). The edge input intensity h_r is visualized as the thickness of the displayed elements. The noise from the background and fruit textures has low amplitude, however, cannot be suppressed by thresholding without losing low contrast contours of the objects. c) shows the grouping result for a CLM architecture with 20 figure layers and one ground layer. The symbols (7 symbols $\times$ black,light and dark grey) represent activity in different layers with symbol size proportional to magnitude. The ground layer is omitted. The scale of the lateral cocircularity interaction is displayed in the upper left corner. The CLM grouping gives 11 segments, which achieve an identification of the most important curve elements in the presence of background noise, where some of the layers remain inactive. The low-intensity edges of the pear and apple are amplified due to the supporting lateral interaction on the salient contours. d) shows the grouping by a Potts spin system with the same interactions. Due to the lack of amplitude information the model has a strong tendency to “hallucinate” curves which cannot be compensated by the ground layer.

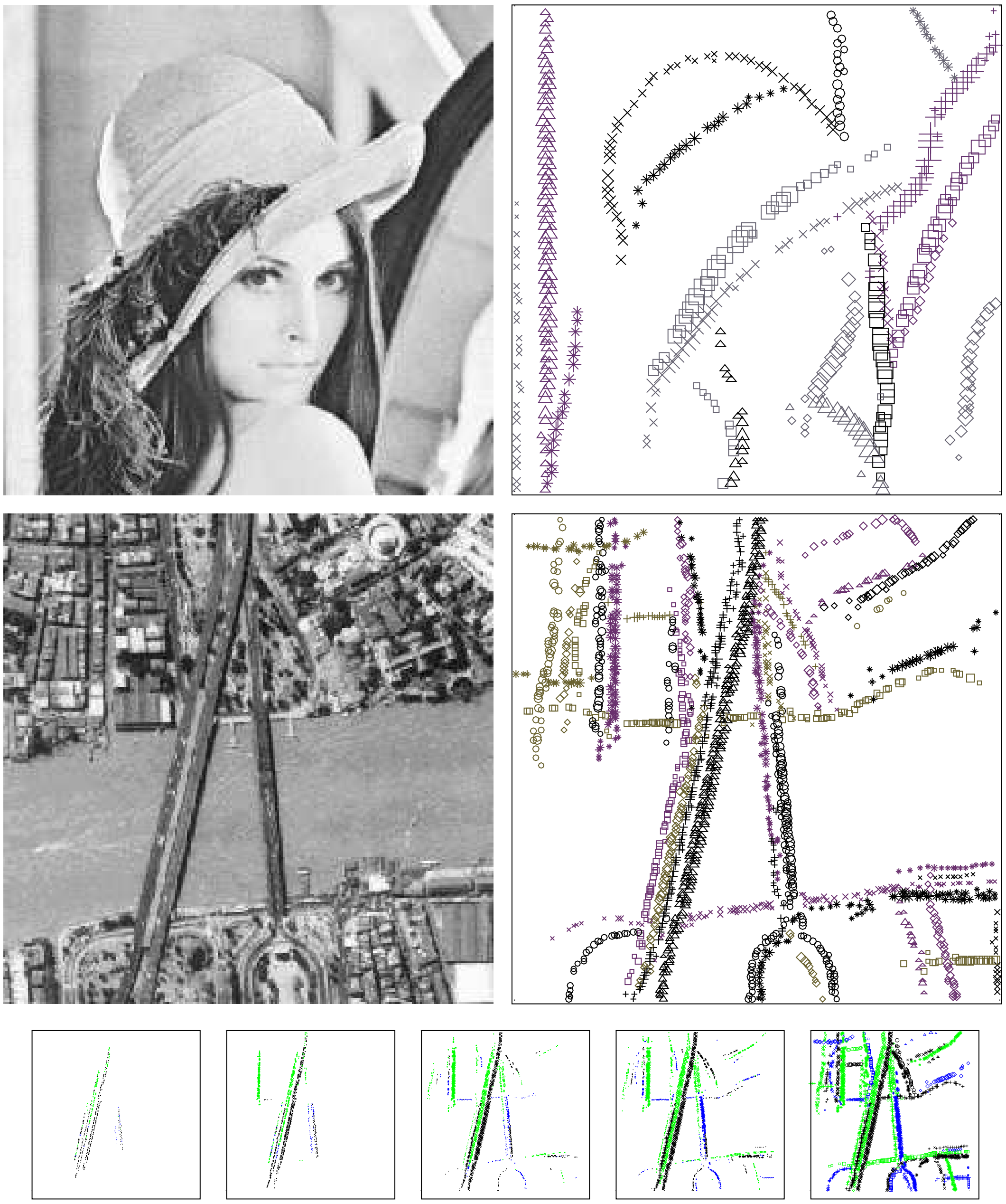


Figure 5.3: Grouping Results. The figure shows the CLM grouping results for two other images with identical parameter settings and 20+ground layers. For the "lena" image (≈ 2000 features) the CLM gives 18 segments which cover the most salient contours. The ground layer strength introduces a minimum size that a segment must have to collect enough lateral support, thus smaller contour segments are not resolved and would require a multi-scale extension of the model. Low intensity edges are enhanced at the brim and top of the hat. The lower aerial image (≈3500 features) is taken from (Yen & Finkel 1996). Due to the presence of more salient contours than layers, some layers actually carry more than a single segment, like e.g. the "black circle" layer, which has three vertical groups in the upper left corner and the center of the image. This illustrates a tendency to combine parallels into the same layer which is most efficient with regard to the energy level of the global grouping. The lower row shows the sequence of emerging groups when lowering the self-inhibition from left to right.

5.1.3 Parameter Dependency and Comparison to Potts Spin Models

In this section it is investigated, how the grouping quality depends on the most important parameters of the CLM contour grouping model. To provide a means of objective evaluation of the grouping performance, two measuring functions are introduced to assess the quality of the final assignment state that is obtained from the CLM grouping algorithm. Suppose the set of features has been synthetically generated such that each feature can be assigned in advance to one of a set of M groups indexed by $l = 1, \ldots, M$. If l_r denotes the group index of feature r then the *label grouping measure* of an assignment state $\hat{\boldsymbol{\alpha}}$ is defined by

$$Q^l = \frac{1}{N^2} \sum_{rr'} q_{rr'}, \tag{5.2}$$

where

$$q_{rr'} = \begin{cases} 1 & \text{if } l_r = l_{r'} \text{ and } \hat{\alpha}(r) = \hat{\alpha}(r'), \\ 1 & \text{if } l_r \neq l_{r'} \text{ and } \hat{\alpha}(r) \neq \hat{\alpha}(r'), \\ 0 & \text{else .} \end{cases} \tag{5.3}$$

The quantity Q^l measures the proportion of correctly grouped pairs of features, if $Q^l = 1$ then all features are grouped in accordance with the a priori known object relations. If grouping on real images is performed, the object labels are generally not available. In that case the lateral contribution of the CLM energy function provides an indirect measure for the grouping quality. The *energy grouping measure* is defined by

$$Q^E = \sum_{\alpha} \sum_{rr'} \mathsf{f}^{\alpha}_{rr'} g(x_{r\alpha}) g(x_{r'\beta}), \tag{5.4}$$

where

$$g(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0. \end{cases} \tag{5.5}$$

The activities are passed through the squashing function $g(.)$ to make the value of Q^E independent of the contextual activity modulation, since here we are only interested in the binary group relations. Note that Q^E is only appropriate for proper final assignment states, where each feature is active in only one layer.

Annealing and Contextual Modulation

The speed of annealing is a very important parameter with respect to the convergence rate of the CLM. It is characterized by the parameter η with $0 < \eta < 1$, which determines the time-constant of the exponential annealing schedule. If the annealing is carried out very slowly with $\eta \ll 1$, then the time to reach a proper assignment state is mostly determined by the speed of the self-inhibitory annealing process. Another important parameter is the vertical coupling strength J, which must be sufficiently large to avoid run-away activity. If J is chosen close to the critical value J^c, then the activity profile is modulated by lateral interactions. The relation is given by $\mu = J/J^c$.

Figure 5.4 shows the combined exploration of the main effects of the parameters η and μ for a synthetic pattern which consists of four objects and a noisy background. The global

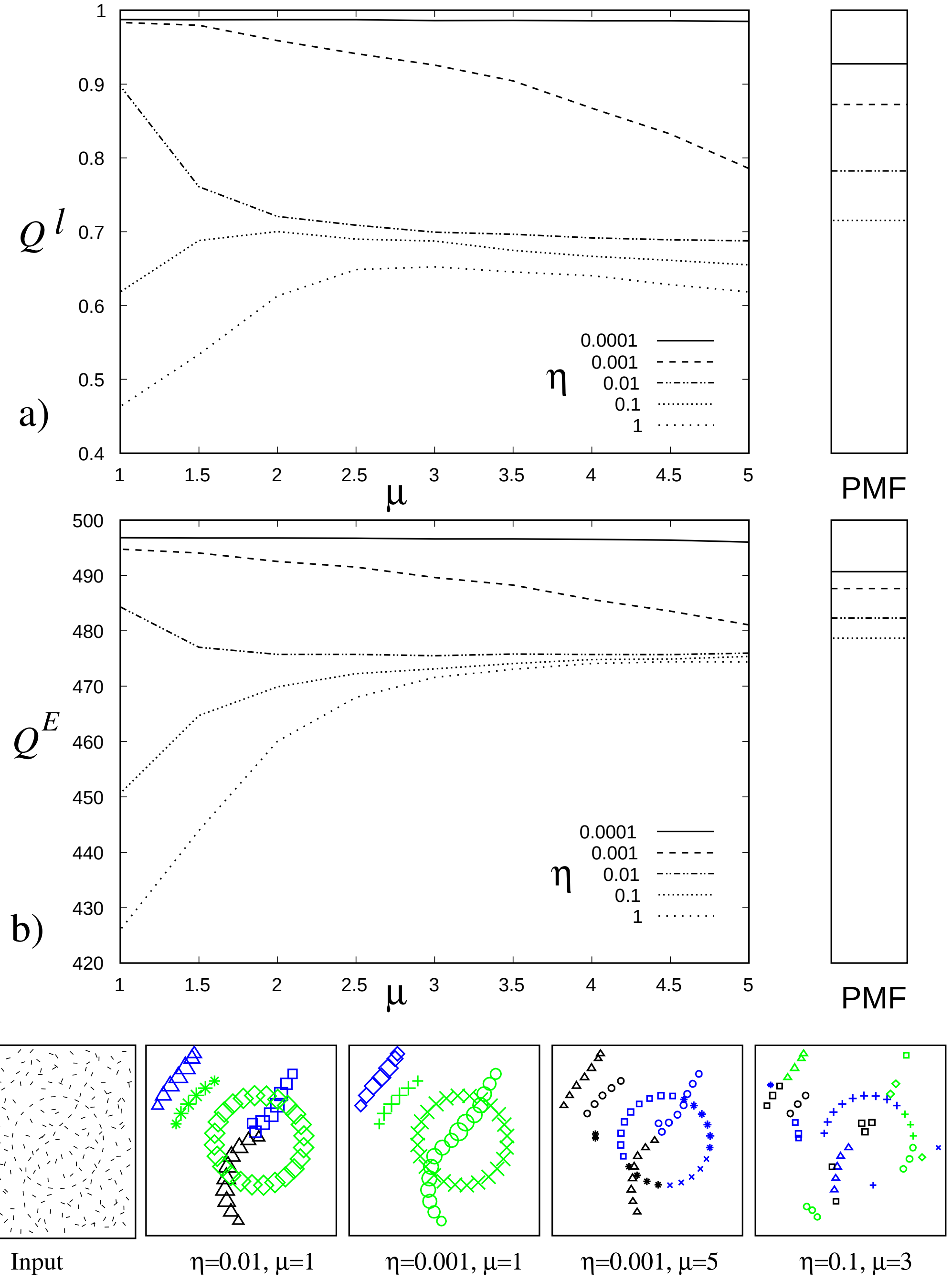


Figure 5.4: Grouping performance dependent on annealing and contextual modulation for a synthetic image averaged over 100 simulations. a) shows the labeling measure Q^l, which gives the proportion of correctly matching pair-relations in the grouping, dependent on the speed of annealing η and the parameter μ. Reasonable grouping is achieved at values above $Q^l > 0.9$. For fast annealing the grouping is largely fragmented. For moderately slow annealing, the quality is increased by operating the system at the stability margin with $\mu = 1$. b) shows the corresponding energy measure Q^E, which varies in accordance with Q^l. Simulation with a Potts mean field (PMF) model achieve lower values than the CLM for $\mu \to 1$. The lower row shows the input image and typical grouping results (ground layer omitted) with corresponding parameters.

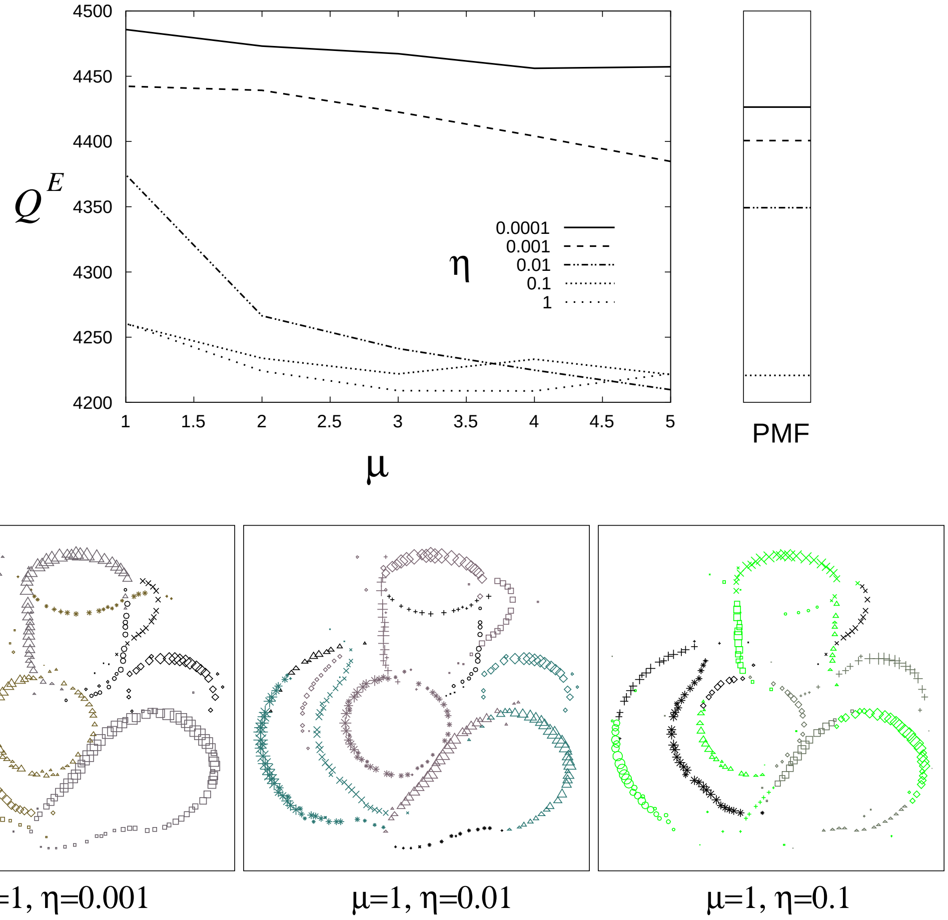


Figure 5.5: Grouping performance dependent on annealing and contextual modulation for a real image averaged over 50 simulations. The quality is increased by operating the CLM at $\mu \to 1$. A robust grouping requires sufficiently slow annealing with $\eta \leq 0.001$. Compared to the synthetic image, however, the much higher number of features leads to more regular groups even for fast annealing. The Potts mean field approach achieves lower grouping quality than the CLM for $\mu \to 1$.

inhibition was set to $k = 0.3/N$ and the ground layer strength was given by $m = 1.5$. For fast annealing the grouping quality measured by Q^l is rather low and even decreases for $\mu \to 1$. This effect, which is caused by very rapid convergence of the dynamics, however, has no practical relevance, since values of $Q^l < 0.7$ represent almost random edge feature arrangements. For sufficiently slow annealing with $\eta \leq 0.01$, the grouping quality increases by letting J approach the critical value J^c. This effect is mainly caused by the destabilization of small fragments. Since larger fragments can develop stronger activity, they tend to attract small fragments into their group for $\mu \to 1$. If very slow annealing is applied, the result is independent of μ, since the chance of entering local minima is close to zero. The energy measure Q^E varies in accordance with Q^l.

For comparison also a simulation of a Potts mean field (PMF) model was carried out with equal lateral interaction scheme and an asynchronous solution algorithm with similar annealing. The results show that the PMF achieves lower grouping quality than the CLM for $\mu \to 1$, but gives comparable results for $\mu \gg 1$ and moderate annealing. The reason is

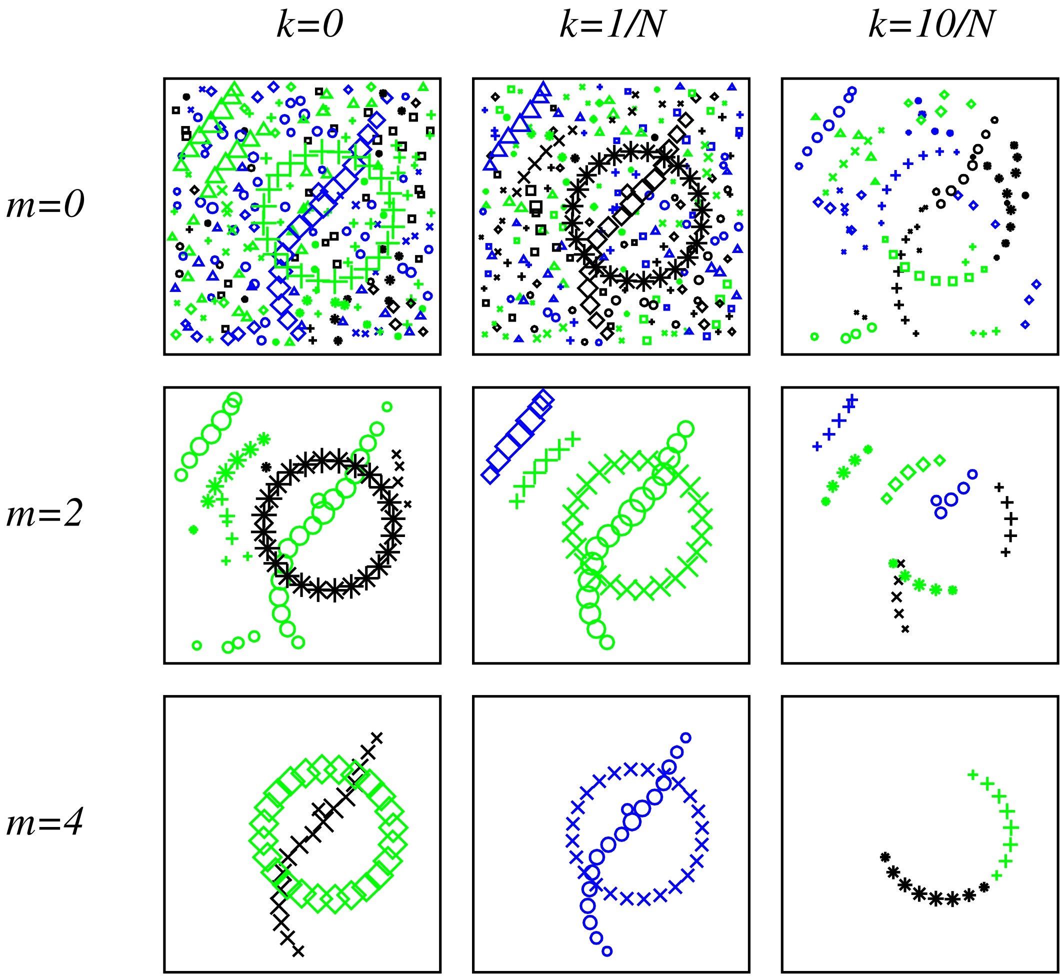


Figure 5.6: Typical grouping results dependent on global inhibition k and ground layer strength m. Raising the ground layer strength m, smaller and less salient groups are attracted to the ground layer. Raising the global inhibition to large values suppresses larger groups. The best results are achieved for a weak but nonzero global inhibition and a value of m which matches the lateral support of features which are part of a salient group. The number of features was given by $N \approx 200$.

that due to the hard summation constraint in the PMF, the contextual activity modulation is absent. For very slow annealing there is also a greater tendency to form spurious groups in the presence of multiple noisy elements.

Figure 5.5 shows the corresponding results for an exploration of η and μ for a large feature set derived from a real image. The other parameters were set to $k = 0.3/N$ and $m = 3.5$. since for this image the feature density is doubled compared to the synthetic image. The dependence on μ is similar, with the difference that the grouping quality consistently increases for $\mu \to 1$ for all annealing speeds. Again the CLM at $\mu \to 1$ achieves higher grouping qualities than a corresponding PMF model with equal annealing.

Ground Layer Strength and Global Inhibition

The background layer strength, determined by the parameter k, induces a self-coupling in the ground layer against which the figure layers must compete by achieving a sufficiently large lateral feedback in a salient group. The global inhibition k results in a competition between the features within the same layer. It can be used to separate remote groups, but may also destabilize single large groups. An exploration of m and k for fixed annealing at $\eta = 0.001$ and $\mu = 1$ is shown in Figure 5.6. By increasing m, smaller segments are more and more attracted to the ground layer, until only the largest and most salient groups remain active. The best results are achieved, if the ground layer strength m approximately matches the sum over excitatory interactions in a proper curve segment. This means that m should be adjusted proportionally to the average feature density after preprocessing, since the sum over lateral interactions is proportional to that factor.

If the global inhibition is zero, then also well-separated groups may be grouped to the same layer. On the contrary if k is very large, then larger groups are unstable. In the intermediate regime, where k is small compared to the typical sum over excitatory lateral interactions, the effect of separation can be robustly achieved, however, for $k \to 0$, slower annealing is required (see Section 4.2.4).

5.1.4 Discussion

The presented CLM approach to contour grouping can achieve reasonable grouping results on complex and noisy real-world images. To achieve good results, it is necessary to operate the model in the regime of strong contextual lateral modulation and to use the mechanism of self-inhibitory annealing. As opposed to other recent models of contour integration which also employed binding mechanisms (Yen & Finkel 1998; Li 1998) and focused on figure-ground segmentation, the CLM converges to a segmentation of the scene into multiple groups, which are clearly separated in the figure layers and a background layer.

A main difference of the CLM compared to spin models of segmentation is the explicit usage of amplitude information in the input. The usage of dynamical WTA circuits results in a context-dependent amplitude modulation which enhances salient groups due to their supporting lateral interactions. Our comparison to a Potts spin model shows that this results in a better noise tolerance for the contour grouping example. This process has been discussed by Li (1998) and Li & Dayan (1999) as selective amplification. We have considered a setup where only the contributions of a set of coarsely sampled features are active, while others are omitted from the model. This decreases the ability to hallucinate contours, since we do not have a full range of feature detectors at all orientations as in the model by Li (1998). This also excludes a direct “filling in” of contours which are not present in the input. To ensure a proper operation of the model in an extended case of features at all orientations, it would be necessary, similar as in (Li 1998), to increase the threshold of the non-saturating linear transfer function from zero to a finite positive value to suppress the activation of neurons with zero afferent input.

5.2 Segmentation of Cell Images from Fluorescence Microscopy

The recent years have seen an increasing demand for automatic processing of medical image data, because the automation of sample preparation and microscopy imaging in biological experiments has led to a an enormous increase of digitized image data. Usually one deals with noisy intensity images, showing objects of same cellular type, but of varying contour. In a collaborate project with Tim W. Nattkemper the application of the CLM feature binding model was investigated for the automated segmentation of lymphocyte cells from fluorescence microscopy (Nattkemper, Wersing, Schubert, & Ritter 2000a). The images were recorded by the biomedical research group at the University of Magdeburg, headed by Walter Schubert and the presented works were done in a close collaboration with the experts of this group.

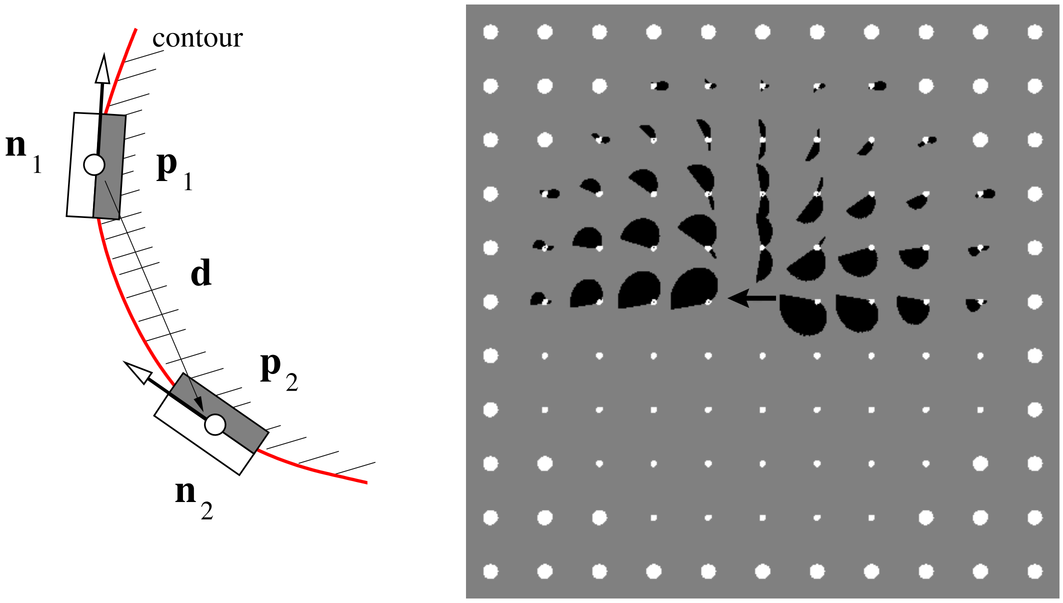
As a basis for later data analysis each image has to be segmented into separately labelled regions, occupied by fluorescent cells, and the background. The cell contours have to be estimated to analyze the fluorescence intensity distribution across the cell body and provide later processing stages with highly relevant information. Evaluation of a large number of these micrographs by human experts is practically impossible, because visual inspection of such noisy intensity images is tiresome. The observer's concentration decreases rapidly, so evaluation of the image data by humans is often not only time consuming but also leads to non-reproducible results.

Segmentation algorithms in related applications are for example based on (i) flooding schemes in gradient mountains (Meyer & Beucher 1990), (ii) wave propagation (Hanahara & Hiyane 1990), (iii) boundary tracing (Galbraith et al., 1991) or (iv) backtracking search through combinations of line segments by calculating a saliency measure (Jacobs 1996). Since the migrating lymphocytes exhibit considerable shape variation, the definition of a common contour model which fits for all cells in the image is infeasible. Furthermore, the images are noisy because of heterogenous signals reflecting different concentrations of cell surface proteins. Because of this heterogenous fluorescence behaviour an intensity binarization, necessary for (iii) and (iv), would be a non-trivial task. Thus, the above approaches seem unsuitable for this scenario.

5.2.1 The CLM Cell Segmentation Approach

The segmentation is based on edge features which are characterized by position $\mathbf{p}_r$ and local intensity gradient orientation $\mathbf{n}_r$ like in the previous Section 5.1 and the feature generation was carried out as described before. A difference is that the input intensity was chosen as $h_r = 1$ for all edges to make the groping less sensitive to intensity fluctuations of the fluorescence imaging method. For this application the lateral interaction is also direction-dependent, taking into account the sign of the local intensity gradient, as opposed to the lateral cocircular interaction outlined in Section 5.1 (see Figure 5.7). This allows to focus the grouping of cocircular edges to the inner convex boundary of the cell and gives better results for the lymphocyte images that were considered.

The interaction function for binding features to cell contour groups is identically defined for all layers, except for the first layer $(\alpha = 1)$, referred to as the "ground layer" which is used to perform simultaneous figure-ground segmentation and grouping. In the remaining layers ($\alpha > 1$) the interaction of two features at positions $\mathbf{p}_1$ and $\mathbf{p}_2$ is computed from their separation $\mathbf{d} = \mathbf{p}_1 - \mathbf{p}_2$ and their unit edge orientation vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ (see Figure 5.7). The complete formal definition of the lateral interaction $f^{\text{cell}}(\mathbf{m}_1, \mathbf{m}_2)$ of two edge features is given in Appendix A. The interaction supports gradient arrangements of two kinds. First, those pairs interact positively which are close enough and which have collinear and equally oriented gradients. Secondly, the more distant pairs of features interact positively, if their gradients are collinear but of opposite orientation, which supports the dynamics in grouping not only convex but *closed* structures together.



a) Interaction scheme b) Convex/continuity interaction field

Figure 5.7: Contour grouping interaction for the grouping of cell boundaries. The interaction supports approximately cocircular edges in a local neighbourhood, but is also dependent on the direction of the gradient as sketched in a), which allows to focus on convex groups of edge elements at the inner boundary of the cell. The interaction field for a single central edge pointing to the left is shown in b) (see Figure 5.1 for explanation of the plotting method). The field mainly supports gradients which "turn right". Outside the local neighbourhood the interaction is weakly inhibitory.

5.2.2 Results

In (Nattkemper et al., 2000a) the CLM contour grouping algorithm was applied to real images of size 64×64 and 128×128. The number of fluorescent lymphocyte cells in the images varied from 1 to 5 or 4 to 20, respectively. All images were processed with the same set of parameters. The CLM consisted of 20 layers plus one ground layer. To compute a one-pixel wide cell contour from the group, standard morphological operators were applied (Sonka, Hlavac, & Boyle 1993). The processing steps for three example images are shown in Figure 5.8. The results show that the cell contours are found including their ring of fluorescence. The grouping performance remains largely unaffected by local distortions in the convexity or variations of the contour in grey value intensity.

The simulations have shown that larger image patches with many cells require slower annealing. Therefore, the processing speed for a given large image can be increased by considering smaller image patches, which are placed at particular focus points. The focus can be generated with an architecture based on an LLM (local linear map) network (Ritter, Martinetz, & Schulten 1992), which is trained by manually determined patches containing salient cells (Nattkemper, Ritter, & Schubert 1999). The combined stages of LLM focus point detection and CLM contour grouping were developed by Tim W. Nattkemper into a hybrid system for high-throughput cell identification (Nattkemper et al., 2000b).

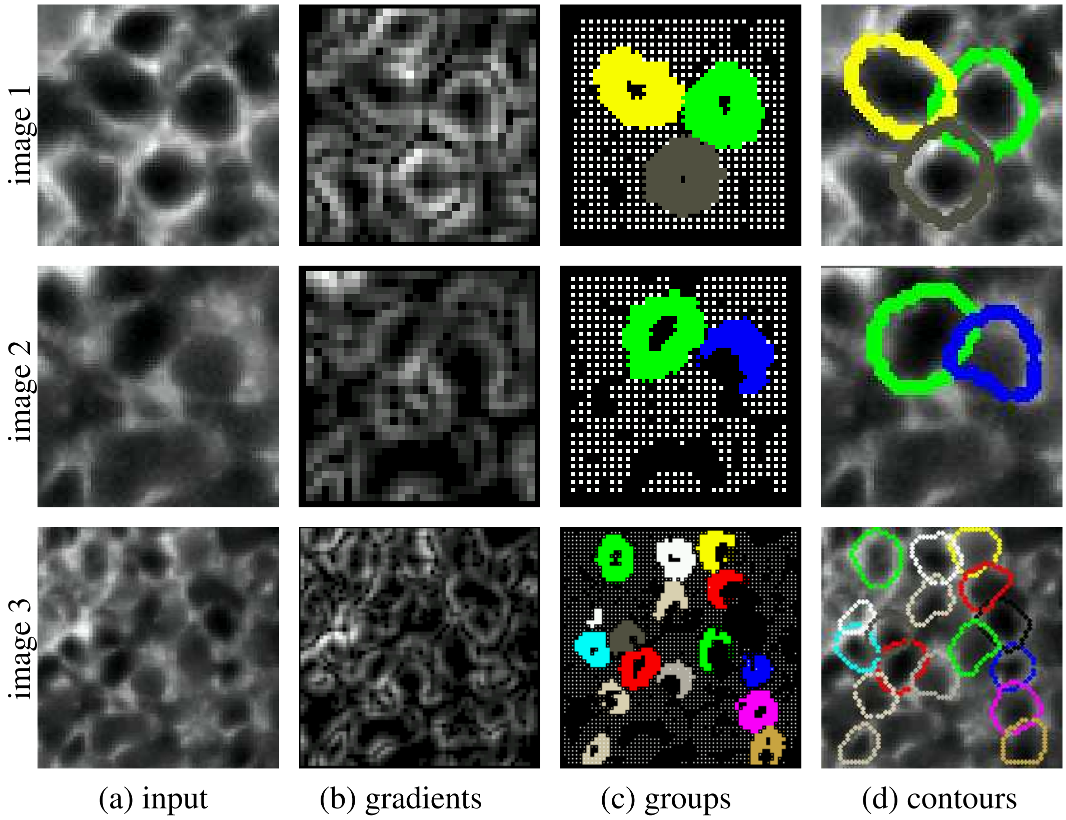


Figure 5.8: CLM contour grouping in fluorescence microscopy images (from Nattkemper et al. 2000a). The input images (a) are subsampled by factor 2 and the x, y-gradients are computed with Sobel x, y operators. Column (b) shows the sum of squared gradient values. In the third column the grouping results can be seen. The small white dots are the features which are grouped into the ground layer. The gaps are omitted features with very small gradients. Groups are visualized by grey boxes with gray values coding for different layers. Even for 20 layers, the number of selected groups is in good accordance with the selection of a biomedical expert (personal communication Walter Schubert). The bottom row shows the application of the model with equal lateral interactions to a larger image (128×128). The contours in (d) were generated from (c) with a standard morphological closing operator.

5.3 Motion Grouping

The estimation of motion in scenes which contain multiple moving objects is an important and difficult problem in computer vision. The main difficulty arises from the so-called aperture problem (see Figure 5.9). The motion information that can be obtained from local optical flow estimation is inherently ambiguous, since it only contains information on the velocity component in the local intensity gradient direction. To solve this problem, most models of motion segmentation assume a two stage scheme, whereby local readings are first computed, and then spatially integrated in a second stage to produce feasible globally consistent velocity estimates.

In this section an approach to planar motion segmentation is suggested based on feature binding in the CLM. The approach is similar to models that fit multiple global motion models to the image data (Wang & Adelson 1994; Black & Anandan 1996; Chang, Tekalp,

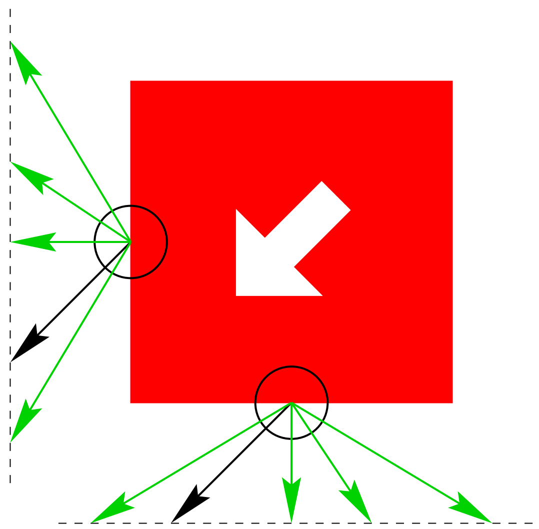


Figure 5.9: The aperture problem. A square is moving into the down left direction and local optical flow is computed at two positions within local windows, depicted by the two circles. The local optical flow constrains only the velocity component in the intensity gradient direction, for example the view seen through the “aperture” of the lower circle would be the same if the square was moving into the down right direction. The grey and black arrows represent consistent velocity vectors for each of the two local optical flow estimations. Only by combining the two local measurements the single (black) velocity vector can be determined, which is consistent with both local optical flow estimations. This illustrates that velocity estimation requires some mechanism of spatial integration.

& Sezan 1997). Therefore it does not aim at providing a conceptually new approach to motion segmentation, but it serves as a demonstration, how the concept can be dynamically realized in the consistent neurodynamical model of the competitive layer model.

5.3.1 CLM Approach to Motion Segmentation

The basic idea is to represent coherent and spatially smooth optical flow field segments in the different layers of the CLM. At each image position $\mathbf{p}_r$, a local flow feature indexed by r is computed from an estimation of the local intensity gradient ∇I_r and the local intensity change $\dot{I}_r$ in two subsequent images. If an object is moving with velocity $\mathbf{v}_\alpha$ over the position $\mathbf{p}_r$ in the image, then the continuity equation

$$\mathbf{v}_\alpha * \nabla I_r + \dot{I}_r = 0 \tag{5.6}$$

is satisfied, which expresses the relation of the local intensity change $\dot{I}_r$ to the intensity gradient component ∇I_r in the velocity vector direction. For a given pair $(\dot{I}_r, \nabla I_r)$ there exists an affine 1-dimensional subspace of velocities $\mathbf{v}$ which satisfy (5.6).

Each layer α is associated with a particular velocity vector $\mathbf{v}_\alpha$, where the two-dimensional space of planar velocities is sampled at a rectangular grid. The activity $x_{r\alpha}$ is interpreted as the certainty of $\mathbf{v}_\alpha$ being an appropriate *globally consistent* velocity estimation at position $\mathbf{p}_r$. To obtain such a global estimate, the grouping dynamics of the CLM is used in the following way: The external input to a neuron in layer α with activity $x_{r\alpha}$ is positive, if the corresponding velocity $\mathbf{v}_\alpha$ is compatible with the local flow feature r. Therefore the input $h_{r\alpha}$ is in this setting also dependent on the layer index and given by

$$h_{r\alpha} = e^{-S(\mathbf{v}_\alpha * \nabla I_r + \dot{I}_r)^2}. \tag{5.7}$$

The strict condition (5.6) is relaxed by a positive sensitivity parameter S, which is necessary due to the finite pixel resolution and noise. Therefore also velocities will receive input which satisfy (5.6) only approximately. The introduction of a layer-dependent input does not affect the stability and convergence properties of the CLM and all previously proved conditions remain valid in this case. The Winner-Take-All dynamics can now be used to select for each feature among the layers, which represent consistent velocities, those which offer the best global interpretation for all features. This is done by using the lateral interaction to enforce spatial regularity of the segmented velocity field. The lateral interaction is given by a local cooperative Gaussian distance-dependent interaction of the form

$$\mathsf{f}_{rr'} = e^{-D(\mathbf{p}_r - \mathbf{p}_{r'})^2}. \tag{5.8}$$

The parameter D controls the range of this spatial cooperative interaction. The dynamics is given by

$$\dot{x}_{r\alpha} = -x_{r\alpha} + \sigma\Big(J\big(h_{r\alpha} - \sum_\beta x_{r\beta}\big) + \sum_{r'} \mathsf{f}_{rr'} x_{r'\alpha} + x_{r\alpha}\Big), \tag{5.9}$$

and minimizes the corresponding energy function

$$E = -J \sum_{r\alpha} h_r x_{r\alpha} + \frac{1}{2} J \sum_r \sum_{\alpha\beta} x_{r\alpha} x_{r\beta} - \frac{1}{2} \sum_\alpha \sum_{rr'} \mathsf{f}_{rr'} x_{r\alpha} x_{r'\alpha}. \tag{5.10}$$

If we assume that J is chosen sufficiently large, such that for a proper assignment state there holds

$$x_{r\hat{\alpha}(r)} \approx h_{r\hat{\alpha}(r)}, \quad x_{r\beta} = 0 \text{ for all } \beta \neq \hat{\alpha}(r), \tag{5.11}$$

then the resulting energy value of an assignment state is given by

$$E \approx -\frac{1}{2} J \sum_r (h_{r\hat{\alpha}(r)})^2 - \frac{1}{2} \sum_r \sum_{r' | \hat{\alpha}(r) = \hat{\alpha}(r')} \mathsf{f}_{rr'} h_{r\hat{\alpha}(r)} h_{r'\hat{\alpha}(r')}. \tag{5.12}$$

The two terms in (5.12) represent the two optimisation goals of this segmentation approach. the first term weighted by the vertical coupling J is minimized by assigning the local flow feature to the velocity layer with strongest input. The second term serves as a spatial regularizer, which enforces spatial smoothness due to the Gaussian distance-dependent lateral interaction $\mathsf{f}_{rr'}$. The choice of J allows to balance the two components. If J is small and close to the critical value J^c, the spatial component will be strong, causing a strong spatial averaging of the obtained motion segments.[1] If J is large, then the input will dominate the grouping assignment to the velocity layers.

5.3.2 Results

A result of the CLM motion grouping approach for a simple moving square stimulus is shown in Figure 5.10. Image positions with vanishing intensity gradient were omitted. The

[1] Strictly speaking, in this case the approximation $x_{r\hat{\alpha}(r)} \approx h_{r\hat{\alpha}(r)}$ is no longer valid, and the value of $x_{r\hat{\alpha}(r)}$ is modulated by lateral effects.

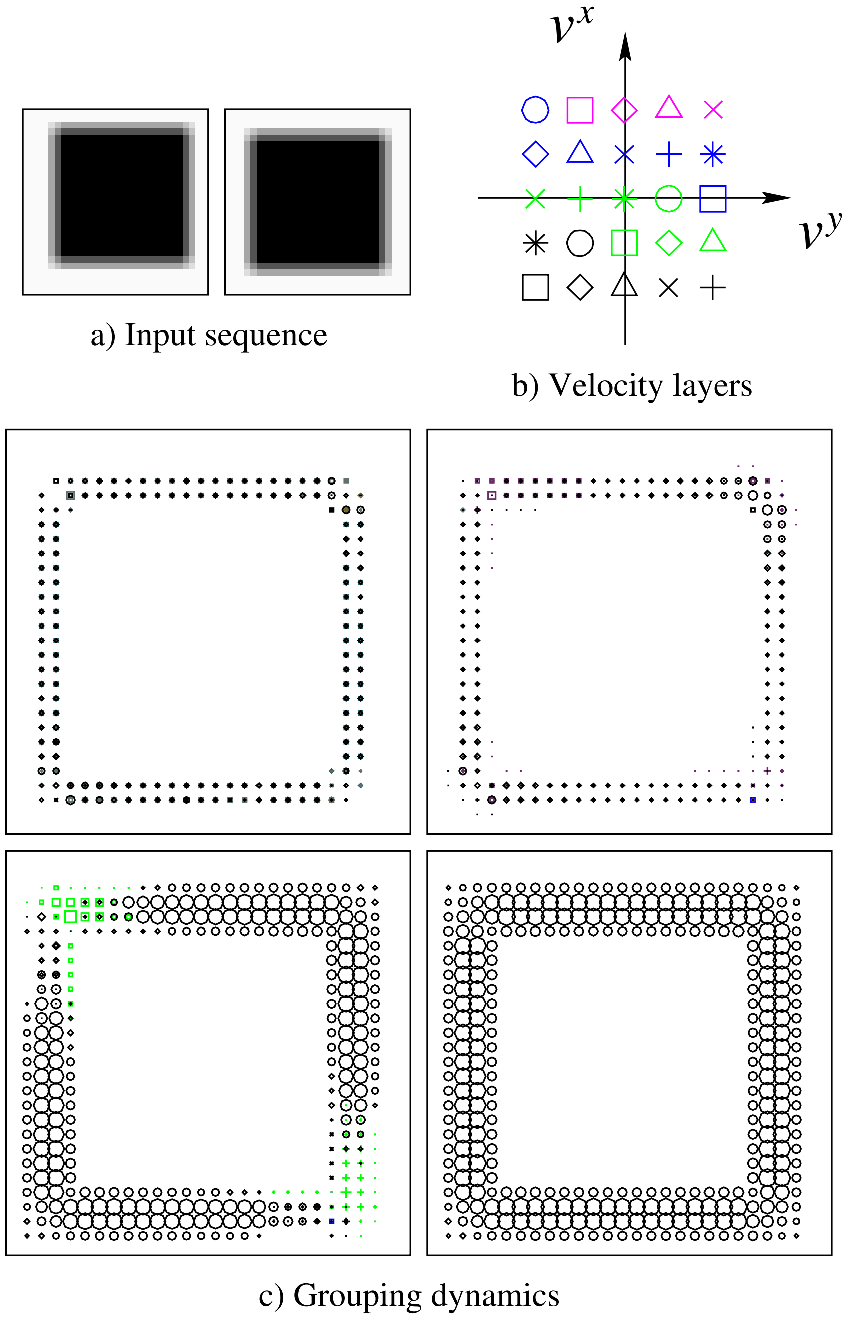


Figure 5.10: Motion Grouping with the CLM. The input sequence a) consists of a square moving to the bottom left direction. In b) the assignment of the 25 symbols representing activity in layers to a particular velocity is shown, where the velocity components in the x and y direction are given by $-2, 1, 0, 1, 2$ pixels per frame. c) shows the time-course of grouping. At the beginning, the sides of the square are active within all layers which are compatible with the local flow feature. For example, for the lower side of the square the layers α are active with $v^y_\alpha = -1$ and $v^x_\alpha \in -2, -1, 0, 1, 2$. The disambiguation between equally driven layers begins at the corners of the square where due to the lateral interactions the velocity layer which is compatible with both v^x and v^y velocity components acquires largest activity. This is propagated along the contour of the square.

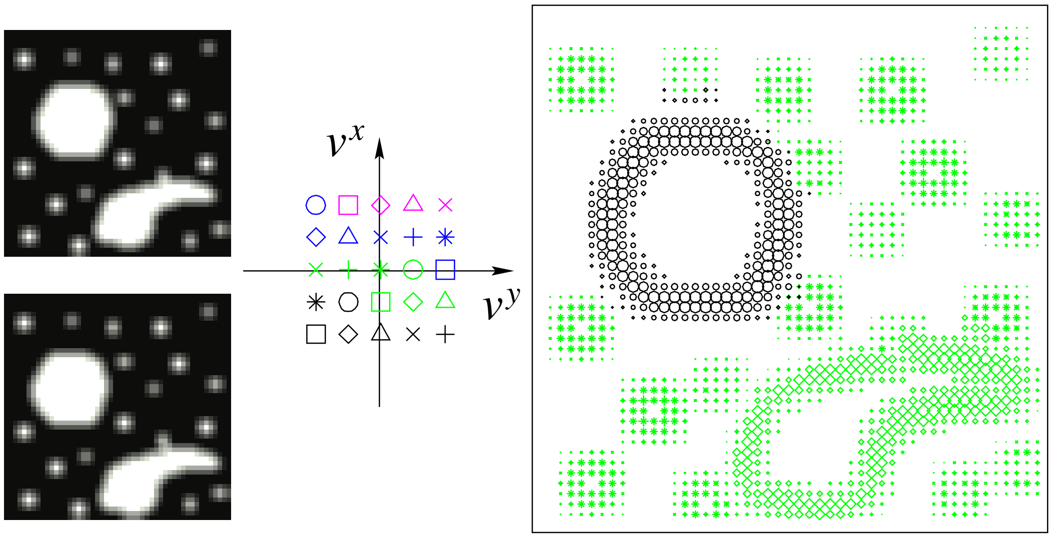


Figure 5.11: Multiple motion segments. For the artificially generated input sequence of two moving objects three groups are formed, of which one represents the background with zero velocity. The "blob" is detected as moving down left, while the "tadpole" is moving down right. Due to the constraint of spatial regularity some background elements are attracted to the moving segments.

velocity selectivity was set to $S = 20$ and the range parameter was chosen as $D = 0.01$ (the spatial dimensions are scaled into a unit square). The velocities were represented by 25 layers, with velocity components $v^x_\alpha \in \{-2, -1, 0, 1, 2\}$ and $v^y_\alpha \in \{-2, -1, 0, 1, 2\}$ pixels per time step. Due to the ambiguity at the sides of the square, after initialization all layers compatible with the vertical or horizontal flow component receive equal input $h_{r\alpha}$. The integration process begins at the corners of the square, where the local lateral cooperation is largest for the velocity layer which is compatible with both x and y components. This larger activity is then propagated along the sides of the square. Figure 5.11 shows the multiple segmentation of two moving objects in front of a fixed background.

The shown results are limited to artificial image data, and the application to real-world image sequences would require more velocity components and more features. Similar to related approaches, the method also suffers from a so-called correspondence problem since for objects with high velocities subsequent frames may not allow to simultaneously evaluate the intensity gradient and the intensity change. This limits the spatiotemporal resolution of this type of integration approach. Nevertheless, the application illustrates that also a variation of the input intensities across layers can be used to influence the grouping behaviour of the CLM.

5.4 Greyscale Segmentation

The task of greyscale segmentation is to detect regions of homogeneous greyscale intensity in an image, which may for example correspond to connected surface elements of a particular object. Classical methods in computer vision are region-based techniques like region-growing and splitting (Adams & Bischof 1994; Zucker 1976), where iterative steps are taken to grow or split pixels into a connected region if all the pixels in the region satisfy some condition. In the recent neural networks literature two feature binding approaches were considered, which were successfully applied to the greyscale segmentation of real images, the LEGION relaxation oscillator model (Wang & Terman 1997) and the ECU spin model (Opara & Wörgötter 1998).

The LEGION model is based on locally excitatory and globally inhibitory coupled relaxation oscillators, where the phases of the oscillators encode the binding of features. Based on lateral interactions, which express the compatibility of features, a lateral potential is introduced, which is used to remove noisy image regions with low lateral support. Since the consistent dynamical oscillator model cannot separate more than about five groups, Wang & Terman (1997) have considered an algorithmic abstraction from the original LEGION model for image segmentation, which captures some of the properties of the dynamical model but can represent more segments and resembles seeded region growing. A further simplification to G-LEGION with even greater similarity to region growing was considered by Shareef, Wang, & Yagel (1999) for the segmentation of intensity images from magnetic resonance imaging (MRI).

The ECU (energy-based cluster update) spin model is a particular update procedure for a system of discrete Potts spin variables. In subsequent iterations clusters are generated according to lateral interactions and the current spin states, and then spins are coherently flipped within clusters with a probability depending on the energy change of the cluster spin flip (see Section (2.3.2)). The method does not perform annealing since it operates at a fixed temperature, for which it converges to a stochastic detailed balance according to the Boltzmann distribution (Opara & Wörgötter 1998). Compared to spin models which use annealing, the method shows rather fast convergence, allowing the application to large spin lattices derived from real images. Opara & Wörgötter have demonstrated the application of the algorithm to greyscale segmentation, but the framework could be also applied to other features and lateral connection schemes.

5.4.1 Approach

The features r which are used for the grouping process are image pixels, which are characterized by greyscale intensity I_r and their position $\mathbf{p}_r$ in the image. Therefore each feature is parametrized by $\mathbf{m}_r = (\mathbf{p}_r, I_r)$. Let us assume that $\mathbf{p}_r$ is given in pixel units. The lateral interaction between two pixel features indexed by r and r' is defined such that regions with similar grey values get positive weights, whereas dissimilar regions get negative weights (Vorbrüggen 1995; Opara & Wörgötter 1998). We define the set

$$\mathcal{N}_r = \{r' | \max(|p^x_r - p^x_{r'}|, |p^y_r - p^y_{r'}|) \leq R\} \tag{5.13}$$

as the set of neighbours of r within (maximum norm) distance R. The lateral interaction is then given by

$$\mathsf{f}_{rr'} = \begin{cases} 1 - \frac{|I_r - I_{r'}|}{\Theta} - k & \text{if} \quad r' \in \mathcal{N}_r, \\ -k & \text{else} \ , \end{cases} \tag{5.14}$$

where k is the strength of a global inhibition and Θ is the intrinsic short-range similarity given by

$$\Theta = \frac{1}{N} \frac{1}{(2R+1)^2 - 1} \sum_r \sum_{r' \in \mathcal{N}_r} |I_r - I_{r'}|, \tag{5.15}$$

where N is the number of pixels in the image. The CLM simulations were carried out with the algorithm outlined in section 4.3.2 with $\mu = 1$, $\eta = 0.0001$, and $k = 0.3/N$. The number of layer was equal to $L = 20$ and a ground layer was used with strength $m = 0$ to reduce the disturbing effect of small less salient fragments.

The simulations of the ECU algorithm were carried out as described in section 2.3.2. A slight modification which was not described in (Opara & Wörgötter 1998), but is important for rapid convergence in the image segmentation application (personal communication R. Opara), was done in the spin update step. The clusters for update are not picked at random, but instead first a random permutation of the cluster indices is computed and then the updates are done sequentially for all clusters in the permutated list. The lateral interactions were equal to the CLM interactions and the number of spin states equal to 20.

In their LEGION model approach to greyscale segmentation Wang & Terman have used a coupling scheme where the lateral interaction weights $w_{rr'}$ between two pixel units were given by

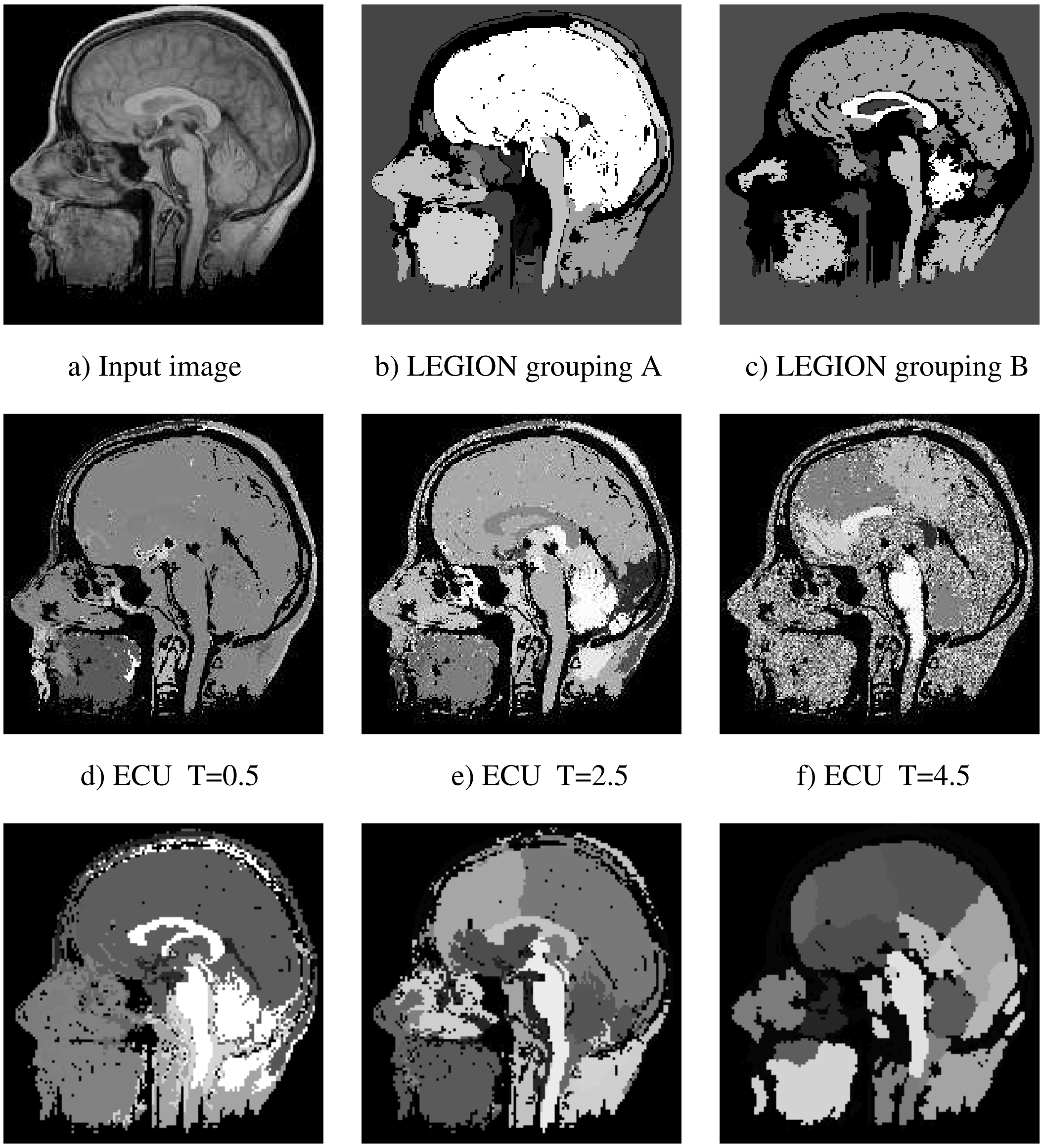
$$w_{rr'} = \frac{I_{\max}}{1 + |I_r - I_{r'}|}. \tag{5.16}$$

where $I_{\max}$ is the maximum intensity value and only connections to the 8 nearest pixel neighbours with $R = 1$ were considered. It should be noted, however, that the dynamical mechanisms responsible for the synchronization-based grouping results are essentially different to the two approaches of CLM and ECU which minimize very similar energy functions.

5.4.2 Results

A comparison of the segmentation results obtained by the CLM to the results of the LEGION oscillator model and the ECU spin model is shown in Figure 5.12. The image shows an MRI cross section of a human head, where the intensity values depend on the characteristics of the different tissues. The results highlight some of the properties of the three different binding approaches.

The LEGION approach emphasizes connectedness and can achieve long-range coherence also for a very short-ranged interactions with $R = 1$. Therefore also very narrow stripes at the top of the head are assigned to a single segment. The parameters of the model can be adjusted to increase the separation between groups which is shown in the second result (taken from Wang & Terman 1997). The figure-ground mechanism can be used to suppress minor fragments, which are shown in dark blue in the figure. A thorough discussion of the LEGION approach to MRI segmentation can be found in (Shareef, Wang, & Yagel 1999).



a) Input image b) LEGION grouping A c) LEGION grouping B

d) ECU T=0.5 e) ECU T=2.5 f) ECU T=4.5

g) CLM R=5 h) CLM R=3 i) CLM R=1

Figure 5.12: Comparison of greyscale segmentation with the LEGION, ECU, and CLM models. The input image a) shows an MRI image of a human head (courtesy of N. Shareef, Ohio State University). b) and c) show the LEGION segmentation for two particular parameter settings (from Wang & Terman 1997). LEGION can enhance connectedness also with small interaction ranges which can be seen for the narrow stripes at the top of the head. d) e) and f) show results obtained with the ECU algorithm at three different temperatures with medium interaction range (R=3). The lower row shows the CLM results, with g) large, h) medium, and i) small spatial interaction range. The CLM results show that long-range coherence requires larger spatial interaction ranges, which, however, reduces the spatial resolution as in g), where cerebellum and spinal cord are merged into a single group.

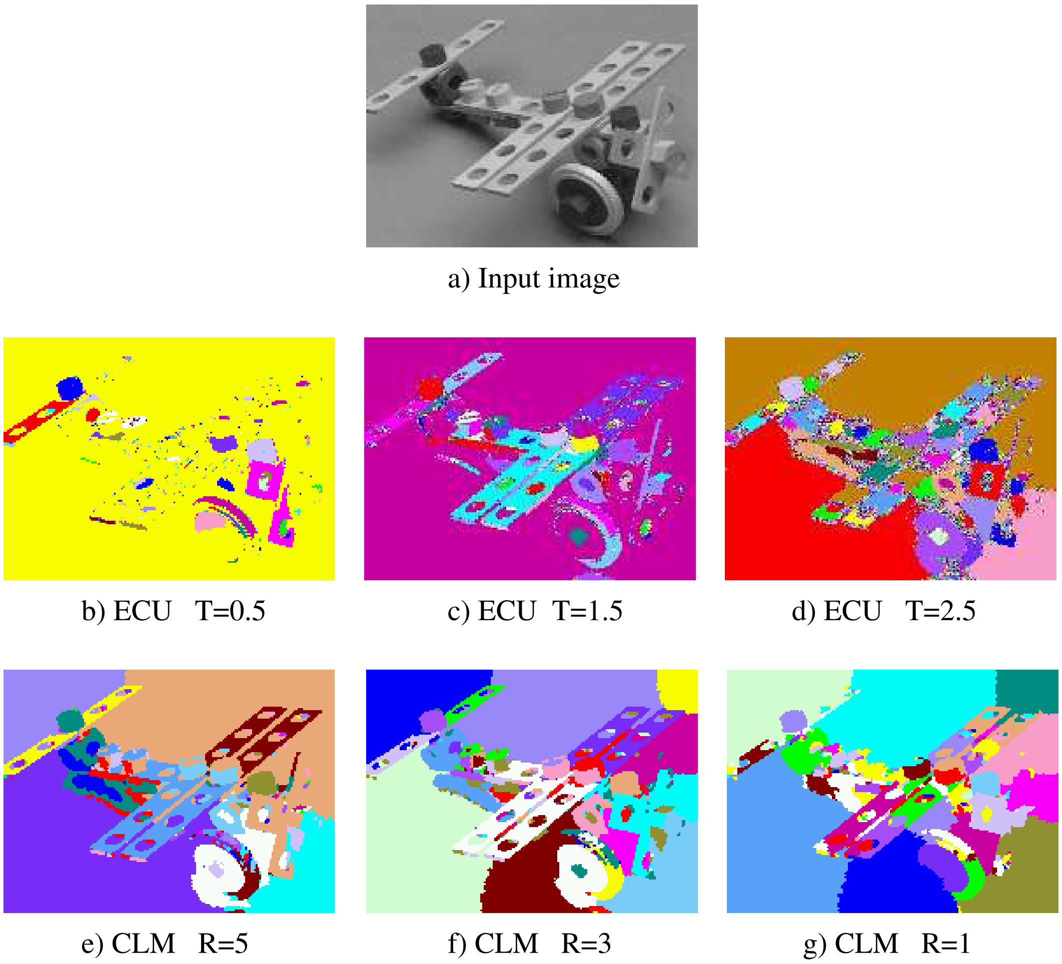


Figure 5.13: More segmentation results for ECU and CLM. The input image a) shows a toy plane from the Baufix domain. The ECU algorithm achieves reasonable grouping within a particular temperature window also for small interaction ranges (R=3). The CLM requires longer interaction ranges for long-range coherence, which results in a tradeoff for spatial resolution.

The ECU algorithm achieves good segmentation results in a particular temperature window. Too high temperatures lead to strong fragmentation, while too low temperatures quickly merge major image regions into a single group. The CLM results all required very slow annealing, since the interaction ranges, which ranged from 1 to 5 pixels, are much smaller compared to the desired group sizes than in the previously discussed contour grouping examples. Increasing the ranges increases long-range coherence, but decreases the possible spatial resolution that can be achieved. For large ranges groups may be connected, which are disconnected on the single pixel level. Another drawback is that the simulation times required are about 20 times longer due to slow annealing than for the ECU algorithm at comparable solution quality. The results given in Figure 5.13 show similar effects for another image.

5.5 Texture Segmentation

Texture segmentation is an important example for preattentive grouping in visual perception. The principles which underlie these segregation phenomena in human texture perception were pioneered in psychophysical studies by Bela Julesz and coworkers (Bergen & Julesz 1983). Julesz proposed the texton theory, which explained the segmentation of visual stimuli in terms of first-order differences between local texture features (Julesz 1981; Julesz 1986). Later it was shown that the texton theory is inconsistent with the visual segregation of particular stimuli and other models were formulated on the basis of spatial frequency and orientation-selective filters (see Ontrup & Ritter 1998 for references).

In his diploma thesis Jörg Ontrup has investigated a model for perceptual grouping of textures which combines an approach of features based on Gabor filters with the grouping model of the CLM. His work has shown, that a model based on an appropriate feature generation stage with a lateral compatibility interaction between texture features can achieve good similarity to human texture discrimination. This section gives a short review of the results.

5.5.1 Approach

The feature generation consists of a combination of linear and nonlinear filtering stages. First, the image is convolved with a set of even-symmetric Gabor filters at 5 orientations and 3 spatial frequencies. Odd-symmetric filters are excluded, because there is evidence that human texture segregation is mainly based on even-symmetric mechanisms. In each filter channel an additional sigmoid nonlinear function is applied, which can be motivated from psychophysical findings. The features are generated by spatial pooling from the generated 15 filter images. A local texture feature r is characterized by a 30-dimensional parameter vector $\mathbf{z}_r = (\boldsymbol{\mu}_r, \boldsymbol{\sigma}_r)$, where

$$\boldsymbol{\mu}_r = (\mu_{11}, \ldots, \mu_{35}), \quad \boldsymbol{\sigma}_r = (\sigma_{11}, \ldots, \sigma_{35}) \tag{5.17}$$

are the mean values and the variances of the filtered images convolved with a localizing Gaussian centered at position $\mathbf{p}_r$ of the feature. The range of the Gaussian is roughly two times the wavelength of the corresponding Gabor filter. The 30-dimensional vector $\mathbf{z}_r$ provides a condensed description of the local image statistics at position $\mathbf{p}_r$. Together with the position information, the features are then defined by $\mathbf{m}_r = (\mathbf{p}_r, \mathbf{z}_r)$.

The lateral interaction is defined as

$$\mathsf{f}_{rr'} = c_{\text{prox}} e^{-|\mathbf{p}_r - \mathbf{p}_{r'}|^2/R^2_{\text{prox}}} + e^{d(\mathbf{z}_r, \mathbf{z}_{r'})/R^2_{\text{text}}} - k, \tag{5.18}$$

where

$$d(\mathbf{z}_r, \mathbf{z}_{r'}) = (\mathbf{z}_r - \mathbf{z}_{r'})^T P(\mathbf{z}'_r - \mathbf{z}_r). \tag{5.19}$$

The parameter c_{prox} controls the weighting of the first term, which implements a cooperative interaction based on spatial proximity with a range characterized by R_{prox}. The texture similarity is computed from a distance function $d(\mathbf{z}_r, \mathbf{z}_{r'})$, where P is a projection matrix onto the first three principal components of the covariance matrix C of all texture vectors $\mathbf{z}_r, r = 1, \ldots, N$ given by

$$C = \frac{1}{N} \sum_r \mathbf{z}_r \mathbf{z}_r^T. \tag{5.20}$$

The projection step effectively reduces the number of dimensions to achieve better noise-robustness, and achieves an automatic selection of directions in the 30-dimensional texture representation, where discrimination based on a distance measure is very effective. The parameter k gives the strength of a global inhibition. For the simulations the procedure described in Section 4.3.1 was used, but no annealing was applied, i.e. T=0. Annealing was not necessary for achieving long-range coherence, because the lateral interactions were not necessarily constrained to be local and the texture component of the interaction could also connect texture features at distant image positions.

5.5.2 Results

The Figure 5.14 shows an overview of the results obtained with the CLM texture segmentation approach. All the results were obtained with a single set of parameters for all images. The CLM texture grouping model performs reasonably on textures from the Brodatz texture album (Brodatz 1966). The application to microtextures from psychophysical experiments shows that the model achieves a similar discrimination capability as human subjects without "oversegmenting" stimuli which humans do not separate preattentively. For a set of stimuli usually illustrating the Gestalt laws of proximity and similarity, the texture-based approach provides very interesting grouping results. The results show how texture provides an abstraction from the usual interpretation of the grouping of these stimuli, which considers the interaction of single objects like circles and dots. Due to the principal component analysis in the texture space, the texture dimensions are automatically selected, which allow optimal separation. This can be interpreted as an implementation of the Gestalt law of Prägnanz.

5.6 Discussion

5.6.1 Binding Properties

The CLM binding model shares some components with other recent binding models. For the examples of contour grouping, motion grouping, and greyscale segmentation that were considered, the lateral interactions are based on a local compatibility measure combined with a weak global inhibition, which is the same for LEGION (Wang & Terman 1997) and the ECU spin update model (Opara & Wörgötter 1998). The LEGION model also uses a mechanism to perform figure-ground segmentation by suppressing features with low lateral support. The main difference lies in the dynamical implementation, which for the CLM is given by a consistent model of neural activity dynamics in a layered system of coupled WTA columns.

In the CLM long-range coherence for local interactions and separation are enforced by a self-inhibitory annealing mechanism, that causes controlled expression of dynamical modes, but also results in a tradeoff for convergence time. The CLM application to texture segmentation (Ontrup & Ritter 1998), has shown that for sufficiently dense and long-ranged feature interactions, the CLM performs well also without annealing. The models of Wang & Terman (1997) and Opara & Wörgötter (1998) employ mechanisms similar to region growing and have been successfully applied to the task of greyscale segmentation of image data, however the complex interactions of overlapping and intersecting curve elements might pose problems for their application to contour grouping. On the contrary, long-range coherence is a problem for the CLM for the greyscale segmentation setting where the local interaction range is very small compared to the individual segment size which may spread

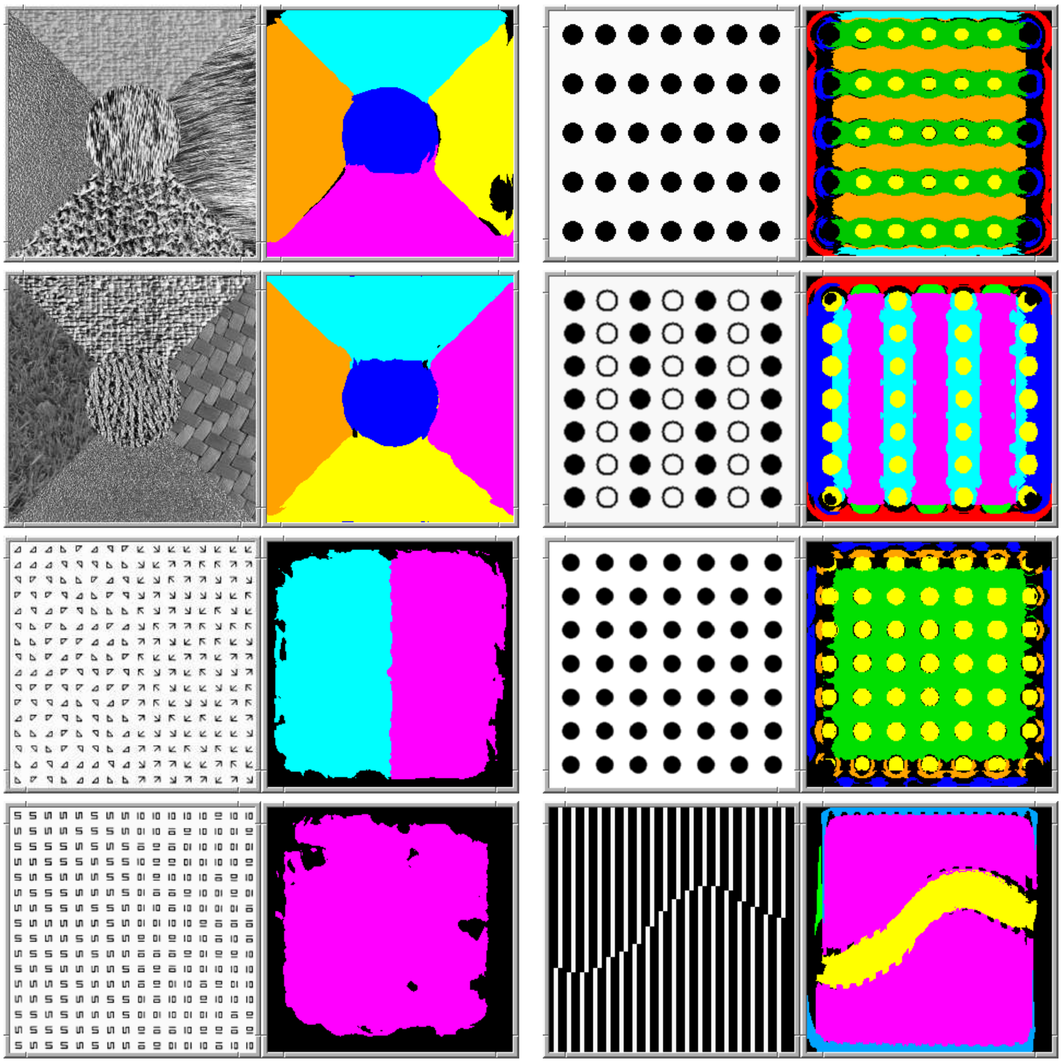


Figure 5.14: Texture Segmentation with the CLM (from Ontrup & Ritter 1998). Different layers are represented by different colors. Black pixels correspond to features, where the activity does not exceed a particular threshold and which are thus not part of a salient group (the background layer mechanism was not yet used). The upper left column shows images derived from the Brodatz texture album. Apart from irregularities in the textures, which were omitted from the grouping, the segments are appropriately detected. The microtextures in the lower left column are examples for stimuli from psychophysical experiments which do or do not segregate preattentively. The CLM segmentation can reproduce the human performance without "oversegmentation". The right column shows a set of stimuli used for illustrating the Gestalt laws of proximity and similarity. The dots are not explicitly represented, instead their spatial layout induces a particular image texture, which can be used by the texture segmentation approach of the CLM to obtain a salient grouping. This automatic process of detecting the appropriate stimulus dimensions in the texture space is due to the principal component analysis and can be interpreted as an implementation of the Gestalt law of Prägnanz.

over the complete image. Therefore the CLM as a binding approach which is most suitable for tasks with a more complex interaction topology, and where lateral interaction ranges are sufficiently large compared to the maximum group size.

5.6.2 Biological Relevance

The key principle of the CLM architecture is the encoding of feature bindings by the assignment to separate populations of laterally interconnected and locally competitive neurons. A biologically realistic interpretation of the CLM architecture may be expressed either in terms of the prominent layered structure of the real visual cortex, or it may be implemented in the rich local connectivity structure of a single neuronal layer itself. Here, the discussion is restricted on the macaque monkey primary visual cortex, area V1, which is one of the most intensively studied regions of the brain.

Although it is known that neurons in the granular and supergranular layers of the primary visual cortex are connected via horizontal connections of considerable lateral spread (Rockland & Lund 1983) there is currently no experimental data available which supports the idea of specific inhibitory (competitive) interactions between disjunct modules of laterally interconnected neurons. However, the CLM model illustrates, that this principle could be easily implemented in a simple neural circuit and therefore searching for related anatomical structures could provide an interesting experimental question. The presence of sufficiently many layers to carry each salient segment in an image is hard to justify in the visual system, albeit it may be useful from a technical applications point of view. Nevertheless, since the suggested self-inhibitory control mechanism together with the figure-ground separation allows for a concentration on the most salient groups, a limited number of layers may not form a severe restriction.

The principle of feature binding within competitive and spatially segregated layers may help to process visual information via different routes through the cortical neuropile. For example, there is evidence of local topographic inhibition between those layers in area V1 which receive the majority of afferent inputs from the LGN (Hubel & Wiesel 1972); the principal input layers $4C\alpha$ and $4C\beta$ are coupled by reciprocal inhibitory interneurons (Lund 1987) (varieties α-3, β-3). Since the α and β sublayers are characterized by different response properties (Blasdel & Fitzpatrick 1984; Hawken & Parker 1984) as well as different spread of lateral connections (Yoshioka, Levitt, & Lund 1994), a putative interpretation of the inhibitory circuitry could be that feature binding in each sublayer is performed by such a competitive interaction. Moreover, the neurons in layer $4C\alpha$ and $4C\beta$ project to different target layers within area V1, which in turn contain key sets of efferent neurons to other cortical areas (Yoshioka, Levitt, & Lund 1994; Yabuta & Callaway 1998). Therefore, a binding of features to different depth of layer 4C may form the basis of the subsequent processing stages.

The layered structure of the CLM involves a certain degree of neural redundancy and could be considered as a “waste” of neural hardware. The fixed hardwired layers might seem less flexible than the synchronization-based approach, which carries additional information into the temporal domain. Models have shown, however, that complex interactions lead to strong limitations with regard to stability and separation of the groupings. We suggest the principle of topological segregation, as used in the CLM, as an additional binding principle which could improve the robustness and stability of the feature binding processes.

Chapter 6

Learning

In this chapter learning methods are presented to determine appropriate lateral interactions for grouping and segmentation from the feature input data. The application of learning algorithms is an attractive alternative to the trial and error method of manually tuning the functional form and parameters of lateral interactions. This manual design method also becomes infeasible for more complex feature spaces, where no intuition on the proper feature interactions is available.

After a review of related learning methods two supervised learning approaches are presented, which are based on the adaptation of the lateral interactions to a set of manually labelled training patterns. The first approach was suggested by Pelillo & Fanelli (1997) for the learning of relaxation labeling compatibilities and can be directly applied to the CLM. The second is a new approach which formulates the learning problem as a convex quadratic consistency optimization problem. To reduce the number of parameters which have to be adapted, the lateral interactions are represented as a weighted superposition of basis functions, for which the weight coefficients are learned. The robustness and generalization capabilities of the two learning approaches are compared for an artificial data set. Finally, the application of the quadratic consistency optimization approach to the segmentation of fluorescence cell images is investigated for different sets of basis functions.

6.1 Learning Approaches to Segmentation and Binding

Let us shortly revise the lateral interaction designs for the perceptual grouping applications of the CLM which were presented in Chapter 5. For the grouping principles of continuity and motion all the parameters determining the lateral interactions of features were chosen manually to achieve best performance for a certain image domain. For the example of greyscale similarity and texture, however, already some of the parameters were determined from the image data. For greyscale similarity this was accomplished by basing the compatibility measure between two pixels on the average local intensity difference within the image which is to be segmented. In the CLM texture segmentation model of Ontrup & Ritter (1998) the compatibility function between two high-dimensional Gabor jet features was a priori chosen to be distance-dependent. The distance in the Gabor feature space, however, was computed by projecting the features onto the first three principal components of the resulting set of Gabor jet feature vectors obtained from the image to be segmented. By this procedure the distance computation is restricted to the directions where feature parameters show largest variation for the single image. As the results of Ontrup & Ritter demonstrate, this image data-driven adaptation results in a segmentation performance over a wide range of image types which is in good accordance with human texture perception.

The two considered methods for adapting the lateral interactions for greyscale intensity and texture are examples for *unsupervised* methods in the machine learning terminology, where information is derived directly from the image data. The distinction between unsupervised and supervised learning can be summarized as follows:

- In **unsupervised learning** all the available information is given by an ensemble of input patterns. The task is to adapt the network parameters from the higher-order dependencies in the input data, to enable the network to perform an unsupervised categorization or segmentation of its input.

- For **supervised learning** there is for each input pattern a corresponding correct output pattern available. In that case the task of learning amounts to the adaptation of the network parameters such that the actual network output matches the given correct target output. For the segmentation or binding problem this means that for a given input pattern, a labeling of the input features according to their groups must be given.

Unsupervised learning has the advantage of not requiring an explicit teacher. Nevertheless, its application faces difficulties in high-dimensional complex feature domains, where many parameters must be estimated. Since it is usually based on statistical estimates of the factorial feature distribution, many input patterns are necessary for a robust parameter estimation. Supervised learning on the other hand can succeed with much less training data. A critical problem, however, is to achieve good generalization. The network is said to generalize well, if it not only exhibits good performance on the training data but also on data, which was not part of the training procedure. To obtain good generalization it is necessary to adjust the number of free parameters for a particular application. If the number of parameters is too small, the representational capacity of the network is too low to capture the interesting structures in the data. If, on the other hand, the number of parameters is too large, the phenomenon of overfitting is encountered. The parameters are fully adapted to the possibly noisy training data, but the generalization to new input patterns is low, since too much noise has entered the structural representation.

There exist a wide range of learning approaches to image segmentation problems (see Pal & Pal 1993 for a review). Of direct relevance to the feature binding methodology based on lateral interactions, as considered in this thesis, are approaches which are based on adapting pairwise feature similarities. Examples for other widely applied methods which are beyond the scope of this thesis are pixel classification based on trained multi-layer perceptrons, vector quantization algorithms and self-organizing maps (Littmann & Ritter 1997). In the following supervised and unsupervised methods for adapting pairwise similarities are revised.

6.1.1 Supervised Learning Methods

Recurrent Adaptation for Relaxation Labeling

Pelillo & Refice (1994) suggested a supervised learning method for the compatibility coefficients of relaxation labeling algorithms. They assume that for each training pattern i a labeling is given in the form of an assignment vector $\mathbf{y}^i$ with components $y^i_{r\alpha}$. The vector $\mathbf{y}^i$ is to be interpreted as the desired activity vector $x_{r\alpha} = y^i_{r\alpha}$ for all r, α to which the relaxation algorithm should converge and is therefore of the form of a proper assignment with

$$y^i_{r\hat{\alpha}(r)} = 1, \qquad y^i_{r\beta} = 0 \quad \text{for all} \quad r, \beta \neq \hat{\alpha}(r). \tag{6.1}$$

The approach is to minimize the difference between the target assignment vector $\mathbf{y}^i$ and the assignment vector $\mathbf{x}(K)$ obtained after a fixed number of K iterations of the relaxation algorithm. The relaxation algorithm (see Section 2.3.1) maps the assignment state $\mathbf{x}(t)$ into a new state $\mathbf{x}(t+1)$ dependent on the compatibilities $\mathsf{f}^{\alpha\beta}_{rr'}$,

$$\mathbf{x}(t+1) = F(\mathbf{x}(t), \{\mathsf{f}^{\alpha\beta}_{rr'}\}), \tag{6.2}$$

where F denotes a nonlinear update function. For a given set of training examples with corresponding assignment vectors $\mathbf{y}^i$, the summed quadratic distance between the output $\mathbf{x}(K)$ after K iterations and the training vector $\mathbf{y}^i$ is given by

$$E^Q = \sum_i \sum_{r\alpha} \left(x_{r\alpha}(K) - y^i_{r\alpha}\right)^2. \tag{6.3}$$

The cost function E^Q is minimized by gradient descent in the interaction parameters $\mathsf{f}^{\alpha\beta}_{rr'}$. Due to the repeated application of the nonlinear relaxation algorithm, the evaluation of the partial derivatives, however, involves complex recursive computations. Since E^Q is therefore highly nonlinear in the optimization variables $\mathsf{f}^{\alpha\beta}_{rr'}$, the method suffers from local minima. The performance can be increased by applying genetic algorithms, as was shown in an application to part-of-speech disambiguation (Pelillo, Abbatista, & Maffione 1995).

Attractor Learning for Relaxation Labeling

An alternative approach for learning of relaxation labeling (RL) compatibility coefficients was suggested by Pelillo & Fanelli (1997). The method is based on adapting the compatibility weights such that the training patterns are consistent in the relaxation labeling sense (see Section 2.3.1). Consistency of the labeling also means that it is a local minimum of the RL energy function and thus an attractor of a relaxation algorithm which minimizes this energy. To satisfy the notion of consistency for all training patterns, for each training pattern $\mathbf{y}^i$ the lateral feedback $F_{r\alpha}(\mathbf{y}_i)$ in a column must be largest for the label $\hat{\alpha}(r)$, which gives the conditions

$$F_{r\beta}(\mathbf{y}^i) < F_{r\hat{\alpha}}(\mathbf{y}^i) \quad \text{for all} \quad r, i, \beta \neq \hat{\alpha}(r), \tag{6.4}$$

which results in a set of linear inequalities given by

$$\sum_{r'\gamma} \mathsf{f}^{\gamma\beta}_{rr'} y^i_{r'\gamma} < \sum_{r'\gamma} \mathsf{f}^{\gamma\hat{\alpha}}_{rr'} y^i_{r'\gamma} \quad \text{for all} \quad i, r, \beta \neq \hat{\alpha}(r). \tag{6.5}$$

Pelillo and Fanelli suggested an iterative solution algorithm to solve these inequalities which was earlier proposed by Eremin (1962). They presented an application, where the relaxation labeling dynamics was used as an associative memory for binary pattern recall. The algorithm will be discussed in detail in the next section, where a deeper analysis reveals that the method is not applicable if the training patterns are inconsistent.

6.1.2 Unsupervised Methods

Peleg & Rosenfeld (1978) proposed to use the correlation between features as a measure for determining the compatibility coefficients for relaxation labeling, since the correlation coefficient matches the intuitive notion of compatibility: It is positive if two features are frequently co-occurring, negative if they exclude each other and zero if they are indifferent. Alternatively also mutual information between features can be used to determine their compatibility. A problem of this approach is that the different frequencies of features require normalizations of the coefficients and that there is no clear statistical interpretation of the resulting energy function available.

Hofmann, Puzicha, & Buhmann (1998) considered unsupervised texture segmentation by a pairwise clustering approach on feature vectors derived from Gabor filter banks at different frequencies and orientations. In their model the pairwise feature compatibilities are determined by a divergence measure of the local feature distributions which is shown to achieve good segmentation results for a range of image types. A lot of other studies on unsupervised texture segmentation have used Markov random fields by estimating the local intensity distributions of images and then performing segmentation based on the corresponding probability model (see Li 1995 for related references).

6.2 Learning of CLM Lateral Interactions

As was already discussed in Section 4.4.3, the competitive layer model has a strong conceptual similarity to relaxation labeling. For the same lateral interaction pattern both models aim at minimizing analogous quadratic energy functions. A difference arises due to the capabilities of contextual modulation in the CLM, but can be suppressed by choosing the vertical interaction components sufficiently strong. Under these conditions, a consistent assignment vector of relaxation labeling is also (approximately) a local minimum of the CLM energy function and thus an attractor of its dynamics. This section presents supervised learning methods for CLM lateral interactions which are based on this principle of attractor shaping through consistency.

6.2.1 Formulation of the Learning Problem

The overall task of the learning algorithm is to adapt the lateral interactions, given by the interaction coefficients $\mathsf{f}^{\alpha\beta}_{rr'}$, such that the CLM architecture performs appropriate segmentation on the labeled training data and also generalizes to the case of new test data. In the following only lateral interactions without cross-layer terms are considered, where $\mathsf{f}^{\alpha\beta}_{rr'} = \delta_{\alpha\beta}\mathsf{f}^{\alpha}_{rr'}$. The generalization to the arbitrary case is straightforward.

In the perceptual grouping applications considered so far, the lateral interactions $\mathsf{f}^{\alpha}_{rr'}$ were only computed for the actual set of N^i features $\mathbf{m}_r$, indexed by $r = 1, \ldots, N^i$ which were obtained from the preprocessing of a particular image i. On the contrary, here we want to obtain an appropriate lateral interaction for *all possible* features by learning. It is therefore assumed that the feature vectors are discretized to a finite set, such that all possible N features can be uniquely indexed by $s = 1, \ldots, N$. In the following the indices s, s' will be used if a summation is carried out over all possible features, and the indices r, r' are used if the summation is done for a particular pattern containing only a subset of the possible features. An example would be the image pixels of a 10×10 image, characterized by position $p^x_s \in \{1, 2, \ldots, 10\}$, $p^y_s \in \{1, 2, \ldots, 10\}$, and intensity $I_s \in \{0, 1, \ldots, 255\}$. The

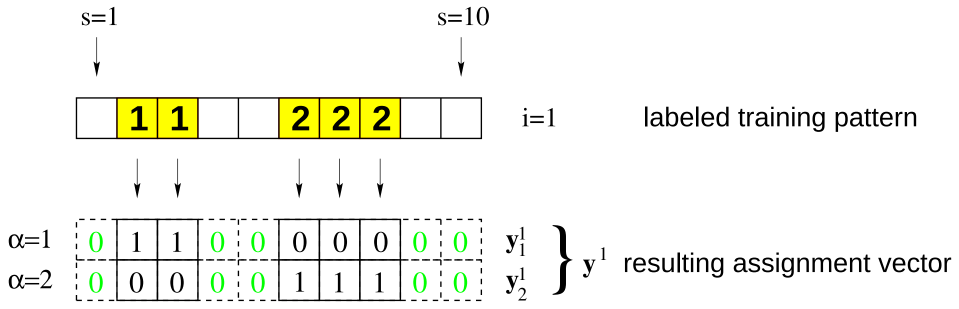


Figure 6.1: Generation of an assignment vector from a labeled training pattern. The training pattern contains only a labeled subset of all possible features.

features are characterized by $\mathbf{m}_s = (p_s^x, p_s^y, I_s)$ with a total of $N = 10 \cdot 10 \cdot 255$ possible features. The learning task is then to determine optimal values for the L $N \times N$ matrices[1] $\mathsf{f}_{ss'}^{\alpha}, \alpha = 1, \ldots, L$. Later it will be shown how to reduce this large parameter space by introducing weighted basis interactions. A single labeled 10×10 image, which could serve as a training pattern, only contains a subset of 100 features, since only one intensity value is realized at each pixel position. In the following a formal definition of the training data is given.

Representation of the Training Data

The training data consists of a set of M labeled training patterns $\mathcal{P}^i$, $i = 1, \ldots, M$, where each pattern $\mathcal{P}^i$ consists of a subset $\mathcal{R}^i = \{r_1^i, \ldots, r_{N^i}^i\}$ of N^i different features with their corresponding labels $\hat{\alpha}$. Thus $\mathcal{P}^i$ is of the form

$$\mathcal{P}^i = \{(r_1^i, \hat{\alpha}^i(r_1^i)), \ldots, (r_{N^i}^i, \hat{\alpha}^i(r_{N^i}^i))\}. \tag{6.6}$$

For each labeled training pattern an assignment vector $\mathbf{y}^i$ is constructed (see Figure 6.1) by choosing

$$y_{r\hat{\alpha}^i(r)}^i = 1, \quad y_{r\beta}^i = 0 \qquad \text{for all} \quad r \in \mathcal{R}^i,\ \beta \neq \hat{\alpha}^i(r) \tag{6.7}$$

for the labelled columns. The columns for features which are not contained in the training pattern are filled with zeroes according to

$$y_{s\alpha}^i = 0 \quad \text{ for all } s \notin \mathcal{R}^i, \alpha. \tag{6.8}$$

Note that it was assumed that the CLM input value is given by $h_r = 1$, however, if also other input values are to be considered then simply choose $y_{r\hat{\alpha}^i(r)}^i = h_r$ in (6.7). The assignment vectors $\mathbf{y}^i, i = 1, \ldots, M$ form the basis of the learning approach since they represent the target activity distribution, which we want to obtain after iterating the CLM with appropriately adjusted lateral interactions. In the following the abbreviation $\hat{\alpha}$ for $\hat{\alpha}^i(r)$ is used to keep the notation readable.

[1]Due to symmetry, there are only $N(N+1)/2$ independent parameters in each matrix.

Learning Goal

The goal of the learning process is to make the training patterns consistent, which is expressed by the inequalities

$$\sum_{r'} \mathsf{f}^{\beta}_{rr'} y^{i}_{r'\beta} < \sum_{r'} \mathsf{f}^{\hat{\alpha}}_{rr'} y^{i}_{r'\hat{\alpha}} \quad \text{for all} \quad i, r \in \mathcal{R}^i, \beta \neq \hat{\alpha}. \tag{6.9}$$

These $(L-1)\sum_i N^i$ inequalities define the learning problem that we want to solve in the following. Let us develop a more compact notation. We can rewrite (6.9) as

$$\sum_{\gamma s s'} Y^{ri\beta}_{\gamma s s'} \mathsf{f}^{\gamma}_{ss'} < 0 \quad \text{for all} \quad i, r \in \mathcal{R}^i, \beta \neq \hat{\alpha}, \tag{6.10}$$

where

$$Y^{ir\beta}_{\gamma s s'} = \delta_{rs}\big(\delta_{\beta\gamma} y^{i}_{s'\beta} - \delta_{\hat{\alpha}\gamma} y^{i}_{s'\hat{\alpha}}\big). \tag{6.11}$$

The form of the inequalities can be simplified by introducing multiindices $\mathbf{k}$ and $\boldsymbol{\nu}$ which correspond to

$$\boldsymbol{\nu} \leftrightarrow (\gamma, s, s'), \qquad \mathbf{k} \leftrightarrow (i, r, \beta) \quad \text{and} \tag{6.12}$$

$$\mathsf{f}_{\boldsymbol{\nu}} \leftrightarrow \mathsf{f}^{\gamma}_{ss'}, \qquad Y^{\mathbf{k}}_{\boldsymbol{\nu}} \leftrightarrow Y^{ir\beta}_{\gamma s s'} \tag{6.13}$$

The index $\mathbf{k}$ runs over all $(L-1)\sum_i N^i$ consistency relations defined for the labeled columns of the assignment vectors. The vectors $\mathbf{Y}^{\mathbf{k}}$ with components $Y^{\mathbf{k}}_{\boldsymbol{\nu}}$ are called *consistency* vectors and represent the consistency constraints for the lateral interaction. The index $\boldsymbol{\nu}$ runs over all entries in the lateral interactions matrices. The vector

$$\mathbf{f} = (\mathsf{f}^{1}_{11}, \ldots, \mathsf{f}^{1}_{NN}, \ldots, \mathsf{f}^{L}_{11}, \ldots, \mathsf{f}^{L}_{NN}) \tag{6.14}$$

with LN^2 components contains the corresponding matrix entries. The inequalities (6.9) can then be written in the form

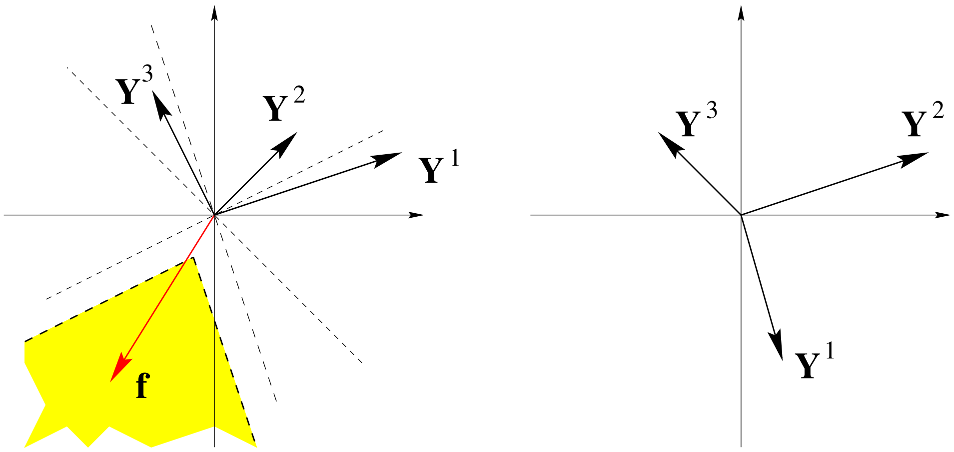
$$\sum_{\boldsymbol{\nu}} Y^{\mathbf{k}}_{\boldsymbol{\nu}} \mathsf{f}_{\boldsymbol{\nu}} < 0 \quad \text{for all} \quad \mathbf{k}. \tag{6.15}$$

This illustrates the nature of the learning problem. The problem is to find a weight vector $\mathbf{f}$ which leads to a lateral interaction matrix, such that the consistency vectors lie in the opposite half space of the weight state space (see Figure 6.2).

Since the conditions (6.15) determine the attractivity of the training patterns, it is customary to introduce a positive margin $\kappa > 0$ to achieve greater robustness. This gives the target inequalities

$$\sum_{\boldsymbol{\nu}} Y^{\mathbf{k}}_{\boldsymbol{\nu}} \mathsf{f}_{\boldsymbol{\nu}} + \kappa < 0 \quad \text{for all} \quad \mathbf{k}, \tag{6.16}$$

which we want to solve in $\mathbf{f}$ for given training data. If the system of inequalities admits a solution for $\mathbf{f}$ it is called *compatible*. If there is no $\mathbf{f}$ satisfying all constraints, the system is called *incompatible*.



a) compatible case b) incompatible case

Figure 6.2: Geometrical sketch of the learning problem. To achieve consistency, the lateral interaction parameter vector $\mathbf{f}$ must have a negative scalar product with all consistency vectors $\mathbf{Y}^i$. The grey shaded area in a) shows the feasible region for $\mathbf{f}$, if an additional positive margin is required. The inequalities are incompatible if there exists an n-dimensional subspace of the weight space which is spanned by more than n consistency vectors.

6.2.2 Superposition of Basis Interactions

If the feature space is large, the number of parameters in the complete interaction matrix $\mathsf{f}^{\gamma}_{ss'}$ is much too large to be robustly estimated from a limited number of training examples. To achieve generalization from the training data it is necessary to reduce the number of parameters which have to adapted during learning. This is also useful to incorporate a priori knowledge into the interaction. An example is to choose basis functions which incorporate invariances such as translation and rotation invariance, or which satisfy the constraint that the interaction is equal in all layers. A simple but powerful approach is to choose a set of K fixed basis interactions g^j of the form

$$g^{j\gamma}_{ss'}, \qquad j = 1, \ldots, K \tag{6.17}$$

from which the interaction $\mathsf{f}^{\gamma}_{ss'}$ is obtained by a linear superposition

$$\mathsf{f}^{\gamma}_{ss'} = \sum_j c_j g^{j\gamma}_{ss'} = \sum_j c_j g^j_{\boldsymbol{\nu}} \tag{6.18}$$

with weight coefficients $c_j, j = 1, \ldots, K$. Now the learning problem of solving the inequalities (6.16) can be recast in the new free parameters c_j. After inserting (6.18) into (6.16) we obtain the transformed problem

$$\sum_{\boldsymbol{\nu}} Y^{\mathbf{k}}_{\boldsymbol{\nu}} \sum_j c_j g^j_{\boldsymbol{\nu}} + \kappa = \sum_j c_j Z^{\mathbf{k}}_j + \kappa < 0 \quad \text{for all} \quad \mathbf{k}, \tag{6.19}$$

where $Z^{\mathbf{k}}_j = \sum_{\boldsymbol{\nu}} Y^{\mathbf{k}}_{\boldsymbol{\nu}} g^j_{\boldsymbol{\nu}}$ is the component of the consistency vector $\mathbf{Y}^{\mathbf{k}}$ in the basis interaction g^j. The basis interactions can thus be used to reduce the dimensionality of the learning problem. In principle it would also be possible to consider a more general nonlinear dependency on a set of parameters; this would then, however, require the much more involved solution of a system of nonlinear inequalities. To avoid any redundancy the basis interactions should be linearly independent. Although the functions are denoted "basis" functions they need neither be orthogonal nor span the whole space of interactions $\mathbf{f} \in R^{LN^2}$. A particular set of basis functions $g^j, j = 1, \ldots, K$ is denoted as an *interaction template* in the following.

6.2.3 The Learning Algorithm of Pelillo and Fanelli

The task of solving the set of inequalities in $\mathbf{f}$ for the set of consistency vectors $\mathbf{Y}^{\mathbf{k}}$ in the compatible case is equivalent to the problem of training a linear perceptron on a linear separable binary classification task. Therefore in the compatible case the perceptron learning rule (Rosenblatt 1962) could be used to obtain a solution in a finite number of steps (Minsky & Papert 1969). Pelillo & Fanelli (1997) suggested to use an iterative solution algorithm for linear inequalities which was given by Eremin (1962). The algorithm has the property that it converges to a feasible solution if the inequalities are compatible and to the solution of a particular approximation of the system if the inequalities are incompatible. Although Pelillo & Fanelli claim that this allows to obtain reasonable results also for incompatible patterns, the following discussion shows that the algorithm is not applicable in the incompatible case.

Given the target inequalities,

$$\sum_{\boldsymbol{\nu}} Y_{\boldsymbol{\nu}}^{\mathbf{k}} \mathrm{f}_{\boldsymbol{\nu}} + \kappa < 0, \qquad \text{for all} \quad \mathbf{k}, \tag{6.20}$$

Eremin's algorithm works as follows: Determine the inequality with index $\mathbf{k}_{\text{max}}$ which is maximally violated:

$$D = \max_{\mathbf{k}} \Big(\sum_{\boldsymbol{\nu}} Y_{\boldsymbol{\nu}}^{\mathbf{k}} \mathrm{f}_{\boldsymbol{\nu}} + \kappa \Big), \qquad \mathbf{k}_{\text{max}} = \arg\max_{\mathbf{k}} \Big(\sum_{\boldsymbol{\nu}} Y_{\boldsymbol{\nu}}^{\mathbf{k}} \mathrm{f}_{\boldsymbol{\nu}} + \kappa \Big). \tag{6.21}$$

If $D < 0$ then $\mathbf{f}$ provides a feasible solution to the inequalities. If $D > 0$ then $\mathbf{f}$ is changed into the direction of the maximally violated constraint $\mathbf{k}_{\text{max}}$:

$$f_{\boldsymbol{\nu}}(m+1) = f_{\boldsymbol{\nu}}(m) - \lambda_m D Y_{\boldsymbol{\nu}}^{\mathbf{k}_{\text{max}}}, \tag{6.22}$$

where λ_m is a sequence for which $\lim_{m \to \infty} \lambda_m = 0$ and $\lim_{m \to \infty} \sum_m \lambda_m = \infty$. As Eremin proved, this algorithm converges to a solution of the system (6.20) if the inequalities are compatible. If they are incompatible, it converges to a solution of the Chebyshev approximation of the system. The Chebyshev approximation of the system (6.20) is defined as the system

$$\sum_{\boldsymbol{\nu}} Y_{\boldsymbol{\nu}}^{\mathbf{k}} \mathrm{f}_{\boldsymbol{\nu}} + \kappa \leq \epsilon_0, \tag{6.23}$$

where ϵ_0 is called the *defect* of the system (6.20) and defined as

$$\epsilon_0 = \min_{\mathbf{f}} d(D), \quad \text{where} \quad d(x) = \max(0, x). \tag{6.24}$$

The minimum over $d(D)$ is carried out over all $\mathbf{f} \in R^{LN^2}$. Pelillo and Fanelli claimed that if the input patterns are incompatible, a solution of the Chebyshev approximation provides an

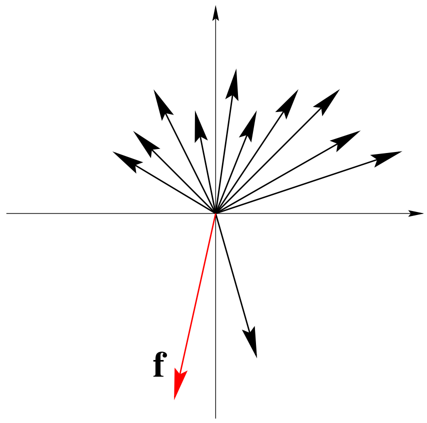


Figure 6.3: Generic case for the resulting consistency vectors. In a typical real world data set the majority of training vectors contain the relevant information, while single spurious states may be present due to noise. A robust learning algorithm for $\mathbf{f}$ must interpolate the result by gradually ignoring less relevant training patterns.

approximative solution to the original problem of determining the interaction coefficients. In that case, however, the defect is equal to $\epsilon_0 = \kappa$ and thus the iteration converges to a solution of the system

$$\sum_{\nu} Y_{\nu}^{\mathbf{k}} \mathsf{f}_{\nu} \leq 0, \qquad \text{for all} \quad \mathbf{k}, \tag{6.25}$$

which only allows the trivial solution $\mathbf{f} = 0$ due to the inconsistency of the training patterns. This is a major drawback of the algorithm, because a single spurious training example is sufficient to cause a breakdown of the learning method. The problem occurs especially if the number of training patterns and thus the number of consistency relations is large compared to the number of free parameters in the lateral interactions. This will be demonstrated by a learning example in the later applications section.

6.2.4 A Robust Quadratic Consistency Optimization Approach

The algorithm outlined in the previous section suffers from a certain lack of robustness, since it cannot deal with incompatible training patterns. The generic case in any real world application is illustrated in Figure 6.3. The majority of training vectors contains the relevant information, while single spurious vectors may be present due to noise or other disturbing factors. These vectors should only cause small distortions without corrupting the whole learning result.

An alternative approach to adapt the interactions such that the consistency equations are satisfied is given by considering the following convex cost function

$$E^{\mathrm{QCO}} = \sum_{\mathbf{k}} \Big(\sum_{\nu} Y_{\nu}^{\mathbf{k}} \mathsf{f}_{\nu} + \kappa \Big)^2. \tag{6.26}$$

A similar minimization approach was suggested by Perfetti (1995) for the imprinting of attractors for the Brain-State-in-a-Box (BSB) model, and a recent study of Park, Cho, & Park (1999) has shown that the approach is competitive with other more sophisticated methods for designing BSB associative memories.

For a fixed positive margin $\kappa > 0$, the cost function (6.26) is minimized by making the inner products of the weight vector and the consistency vectors negative. The global minimum with $E^{\mathrm{QCO}} = 0$ is attained if the inner products are all equal to $-\kappa$, which can be interpreted such that all consistency inequalities are fulfilled in an equal manner. Although this additional regularizing constraint is hard to justify on theoretical grounds, the later application shows that it works quite well for the application examples considered. One important aspect is that the approach is more robust since the minimization allows a violation of the consistency constraint for single indices $\mathbf{k}$, if this is necessary to achieve consistency for the majority of other patterns.

If we insert the expansion of $\mathbf{f}$ in the basis of g^j function templates we obtain according to (6.18)

$$E^{\mathrm{QCO}} = \sum_{\mathbf{k}} \Big(\sum_j c_j Z_j^{\mathbf{k}} + \kappa \Big)^2, \tag{6.27}$$

which results in a K-dimensional convex quadratic minimisation problem in the c_j parameters. The coefficients $Z_j^{\mathbf{k}}$, which give the components of the training patterns in the basis interactions are given by

$$Z_j^{\mathbf{k}} = \sum_{\boldsymbol{\nu}} Y_{\boldsymbol{\nu}}^{\mathbf{k}} g_{\boldsymbol{\nu}}^j = \sum_{r'} y_{r'\beta}^i g_{rr'}^{j\beta} - \sum_{r'} y_{r'\hat{\alpha}}^i g_{rr'}^{j\hat{\alpha}}. \tag{6.28}$$

The quadratic optimisation problem is then given by minimizing

$$E^{\mathrm{QCO}} = \sum_{jl} A_{jl} c_j c_l + \sum_j B_j c_j + \kappa^2, \tag{6.29}$$

where

$$A_{jl} = \sum_{\mathbf{k}} Z_j^{\mathbf{k}} Z_l^{\mathbf{k}}, \qquad B_j = 2\kappa \sum_{\mathbf{k}} Z_j^{\mathbf{k}}. \tag{6.30}$$

If the coefficients c_j are unconstrained, then the minimum of (6.29) can be obtained by solving the linear system of equations $\partial E/\partial c_j = 2\sum_l A_{jl} c_l + B_j = 0$ for all j. The application results presented in the following sections have shown, however, that for the templates which were considered, the grouping performance is better if additional interval constraints $|c_j| \leq 1$ for all j are imposed.[2] The minimization of the convex function (6.29) under interval constraints can then be done by the projection gradient descent method proposed by Rosen (1960) or other convex minimization algorithms. We observed that the main effect of this modification is to keep single coefficients from obtaining large values which disturbs the group formation process. The overall distribution of excitatory and inhibitory interactions, however, remains approximately the same.

6.2.5 Additional Heuristic Learning Conditions

The two presented learning approaches are based on adapting the lateral interactions such that the training patterns are local minima and thus attractors of the dynamics. The perceptual grouping applications of the CLM have shown, however, that there usually exist a lot

[2] Park, Cho, & Park (1999) considered additional eigenvalue constraints in the related problem of Brain-State-in-a-Box attractor shaping.

of other spurious local minima which correspond to fragmented groupings. The structure of the grouping or segmentation problem does not allow to suppress these local minima completely. An alternative approach is to ensure that the spurious minima have higher energy values and then use annealing to reduce their chance of occurrence. This means that the learning process should not only turn the training patterns into local but also global minima.

The assignment vectors derived from the training patterns contain the proper label assignments. We can use them additionally to obtain negative training examples which should acquire higher energy values than the proper labelings or groupings. In that way additional heuristic learning conditions can be motivated.

Local Condition

Let us first consider the local energy change if the label $\hat{\alpha}(r)$ of a feature r is changed to the new label β. We only consider the grouping part of the energy which is given by

$$E^g = -\frac{1}{2}\sum_{\alpha r r'} \mathsf{f}^{\alpha}_{rr'} y^i_{r\alpha} y^i_{r'\alpha}. \tag{6.31}$$

If we assume that the interaction is equal in all layers with $\mathsf{f}^{\alpha}_{rr'} = \mathsf{f}_{rr'}$ for all α we can compute the energy change as

$$\Delta E^g = y^i_{r\hat{\alpha}}\Big(- \mathsf{f}_{rr} y^i_{r\hat{\alpha}} + \sum_{r'} \mathsf{f}_{rr'} y^i_{r'\hat{\alpha}} - \sum_{r'} \mathsf{f}_{rr'} y^i_{r'\beta}\Big), \tag{6.32}$$

If we now require that the energy change $\Delta E^g > 0$ for this spurious rearrangement and divide by $y^i_{r\hat{\alpha}} > 0$ we reobtain the consistency relation

$$\mathsf{f}_{rr} y^i_{r\hat{\alpha}} - \sum_{r'} \mathsf{f}_{rr'} y^i_{r'\hat{\alpha}} + \sum_{r'} \mathsf{f}_{rr'} y^i_{r'\beta} < 0 \tag{6.33}$$

where the margin is given by $\kappa = \mathsf{f}_{rr} y^i_{r\hat{\alpha}}$. We can interpret this as a minimum value for the margin parameter κ in order to make the grouping stable against a columnar label exchange.

Cluster Condition

Instead of flipping a single label, we can also consider the combined flip of the whole cluster of features with label γ to the label β, which is already populated by another feature cluster. Since the two clusters should not be merged, the energy change of this reconfiguration should be positive. Again we only consider the case of equal interactions. Since then all intra-cluster contributions cancel out we obtain for the total change

$$\Delta E^g = - \sum_{r|\hat{\alpha}(r)=\gamma} y^i_{r\hat{\alpha}}\Big(\sum_{r'} \mathsf{f}_{rr'} y^i_{r'\beta}\Big). \tag{6.34}$$

Since we want to derive conditions which are comparable to the local consistency conditions, we only consider the local energy change, which covers the contribution of the feature under consideration. A single column contributes in 6.34

$$\Delta E^g_r = -y^i_{r\hat{\alpha}}\Big(\sum_{r'} \mathsf{f}_{rr'} y^i_{r'\beta}\Big). \tag{6.35}$$

If we require that $\Delta E_s^g > 0$ with the same margin κ as the consistency relations and divide by $y^i_{r\hat{\alpha}}$ we obtain a set of additional target learning inequalities

$$\sum_{r'} \mathsf{f}_{rr'} y^i_{r'\beta} + \kappa < 0, \quad \text{for all} \quad \mathbf{k}, \tag{6.36}$$

where those cases for k are excluded, where there is no feature in layer β. Equations (6.36) are denoted as the additional heuristic *cluster conditions*. The main effect of these additional conditions is to induce a stronger "repulsion" of different groups, since (6.36) requires to strengthen the inhibitory connections with $\mathsf{f}_{rr'} < 0$.

If we insert the expansion of the basis function we obtain the new coefficients for the quadratic consistency optimisation problem

$$A_{jl} = \sum_{\mathbf{k}} Z^{\mathbf{k}}_j Z^{\mathbf{k}}_j + \sum_{\mathbf{k}} C^{\mathbf{k}}_j C^{\mathbf{k}}_j, \qquad B_j = 2\kappa \sum_{\mathbf{k}} (Z^{\mathbf{k}}_j + C^{\mathbf{k}}_j), \tag{6.37}$$

where $C^{\mathbf{k}}_j = \sum_{r'} y^i_{r'\beta} g^{i\beta}_{rr'}$. Note that this choice automatically excludes the case where there is no feature in layer β.

Background Layer

A background layer introduces a particular deviation from the connection pattern in the figure layers, since there is only a self-coupling interaction. The strength of the background layer self-coupling m provides an interesting target for our learning approach.

Let us assume that the background layer has label $\alpha = 1$. If a feature r in the training pattern carries the background label $\hat{\alpha}(r) = 1$, then the consistency equations read

$$-m y^i_{r1} + \sum_{r'} \mathsf{f}^{\beta}_{rr'} y^i_{r'\beta} + \kappa < 0, \qquad \text{for all} \quad \beta \neq 1, \tag{6.38}$$

and can be effectively used to constrain the relations between the self-coupling m and the other figure-layer interactions $\mathsf{f}^{\beta}_{rr'}$. If, however, the consistency is compared between a feature with label $\hat{\alpha}(r) > 1$ and the background layer, the consistency equations are given by

$$-\sum_{r'} \mathsf{f}^{\beta}_{rr'} y^i_{r'\beta} + \kappa < 0, \tag{6.39}$$

since there is no activity in the background layer available for the corresponding assignment vector. The consequence is that m cannot be appropriately adjusted. As a simple heuristics we suggest instead of (6.39) to use the modified inequality

$$m y^i_{r\hat{\alpha}} - \sum_{r'} \mathsf{f}^{\beta}_{rr'} y^i_{r'\beta} + \kappa < 0, \tag{6.40}$$

which expresses the condition that the lateral support in the figure layer must be larger than if the same feature was active in the background layer.

If $\alpha = 1$ is the background layer index, then the corresponding modification of the optimisation coefficients $Z^{\mathbf{k}}_j$ due to the change made in (6.40) is given by

$$Z'^{\mathbf{k}}_j = \begin{cases} Z^{\mathbf{k}}_j + m y^i_{r\hat{\alpha}} & \text{if} \quad \hat{\alpha}(r) > 1 \text{ and } \beta = 1, \\ Z^{\mathbf{k}}_j & \text{else,} \end{cases} \qquad \text{for all} \quad \mathbf{k}. \tag{6.41}$$

If additionally the cluster condition is applied, it should be restricted to the figure layers, since it has no reasonable interpretation in the background layer. This implies the modification

$$C'^{\mathbf{k}}_{j} = \begin{cases} 0 & \text{if} \quad \hat{\alpha}(r) = 1 \text{ or } \beta = 1, \\ C^{\mathbf{k}}_{j} & \text{else,} \end{cases} \qquad \text{for all} \quad \mathbf{k}. \tag{6.42}$$

Free Layers

An important advantage of the CLM compared to other pairwise clustering approaches is that the number of layers does not determine the number of groups. If sufficiently many layers are present, some layers may remain empty if fewer salient clusters are present in the input. This, however, is only possible if the lateral interactions are appropriately adjusted. For example for a completely negative lateral interaction the global minimum for sufficiently many features is always given by occupying all layers or labels. To allow the learning method to generalize to the situation that free layers are present, it is therefore necessary to extend the assignment vectors by at least one additional layer for which no label is assigned in the training data.

6.3 Application to an Artificial Data Set

In this section the two different learning approaches of Pelillo & Fanelli using Eremin's algorithm and quadratic consistency optimization are compared on an artificial data set. The focus is to investigate the robustness and the generalization capabilities of the learning methods.

6.3.1 Setup

The Data Set

Each pattern consists of a 1-dimensional lattice of 10 features which are fully characterized by their position $s = 1, \ldots, 10$ on the lattice. Each of the pattern contains two labelled clusters of up to 4 neighbouring features which are separated by at least one free space (see Figure 6.4). Since only translation-invariant interactions will be considered with an inherent generalization over translation, patterns which differ only by a common shift are omitted.

The task of the learning algorithm is to find a lateral interaction which is capable of grouping the unlabeled input patterns such that connected clusters are assigned to different layers. Although only labels $\alpha = 1$ and $\alpha = 2$ are used, it is assumed that a third layer is present which remains unoccupied, hence $L = 3$. This is necessary to let the network learn to leave free layers if the number clusters is smaller than the number of layers (see Section 6.2.5).

Lateral Interactions

Since a grouping application is considered, it is assumed that the lateral interaction does not depend on the layer index, $\mathsf{f}^{\gamma}_{ss'} = \mathsf{f}_{ss'}$ for all $s, s'\gamma$. In the most general case the symmetric 10×10 matrix $\mathsf{f}^{\gamma}_{ss'}$ has 55 free parameters which would have to be adapted from the training

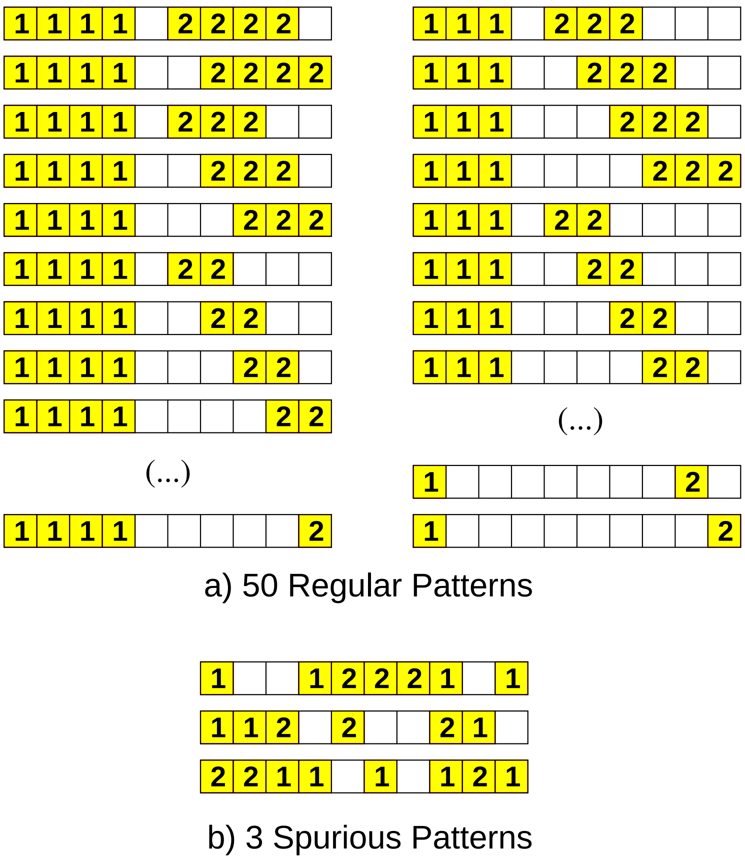


Figure 6.4: Synthetic training patterns for the example of 1-dimensional clustering. a) shows the 50 regular patterns which consider all combinations of two clusters up to 4 units long. b) shows three randomly generated spurious patterns which were used to evaluate the robustness of the learning algorithms.

data. On the contrary, if we assume that the task is translation invariant, we can consider a set of basis interactions which are distance-dependent. We choose the following set

$$g_{rr'}^{j\alpha} = \begin{cases} 1 & \text{if} \quad |r-r'|+1 = j \\ 0 & \text{else} \end{cases}, \qquad j = 1, \ldots, 10 \tag{6.43}$$

of 10 basis functions. Choosing $\mathsf{f}_\nu = \sum c_j g_\nu^j$ implies that the coefficient c_j gives the interaction strength of two features at distance $|r - r'| = j - 1$.

Learning and Evaluation Procedures

Learning is done with and without additional cluster conditions to compare the effect on the grouping performance. In a first step the components $Z_j^{\mathbf{k}}$ (and $C_j^{\mathbf{k}}$ if necessary) of the training consistency vectors in the basis interactions are computed for the current complete training set. The resulting consistency inequalities are solved in the basis interactions coefficients c_j with either of the learning algorithms. Eremin's algorithm was implemented as explained in Section 6.2.3. The constrained minimization for the quadratic consistency optimization approach was done by using the gradient descent algorithm of Rosen (1960), which shows sufficiently fast convergence for this application. Each simulation run consists of the following steps:

1. Choose a particular margin value κ which is a free parameter of both learning methods
2. Apply the learning algorithm based on the consistency relations in the training set
3. Draw a test pattern from the pattern ensemble and use it as input for a CLM with the learned lateral interactions
4. Evaluate the obtained grouping result

Both learning algorithms were stopped after a maximum of 100000 learning steps, where convergence has usually been achieved [3]. The annealing for the CLM dynamics in step 3 was set at $\eta = 0.001$ and $J = 2J_c$ to keep the contextual modulations small. The number of layers was given by $L = 3$.

6.3.2 Robustness

The first experimental setup was to evaluate the learning performance on the set of 50 regular 2-cluster patterns (see Figure 6.4). Of particular interest in view of any real-world application is the robustness of the methods, which can be investigated by adding spurious labeled patterns to the regular training set. For each choice of margin parameter κ, 1000 simulations were performed, and the proportion of correct grouping results was evaluated. We compared both Eremin's algorithm and the quadratic consistency optimization with and without the additional cluster conditions. The basis interactions coefficients were constrained by $|c_j| \leq 1$ for all j. The results are shown in Figure 6.5.

Eremin's method performs well on the undisturbed training set in an interval around the margin value $\kappa \approx 1$ with decaying performance on either side of the margin interval. The addition of a single spurious training pattern causes an overall breakdown of the grouping performance. After adding two spurious pattern the Eremin iteration slowly converges towards the zero vector, which is due to the resulting inconsistency of the training vectors. The application of the cluster conditions showed no significant difference and is therefore omitted.

The quadratic consistency approach requires a margin parameter $\kappa > 1$ to learn, which can be also seen by considering the cost function E^Q. If $\kappa \leq 1$ then the E^Q attains its minimum at $E^Q = 0$ for the diagonal interactions with $\mathsf{f}_{rr'} = \kappa\delta_{rr'}$. This result is also obtained from the minimization algorithm. For $\kappa > 1$ the algorithm produces good results with a slight decay for larger margin values. If spurious patterns are added, there is only a mild degradation of the results, since the spurious patterns give only a marginal contribution in the quadratic cost function. By using the additional cluster conditions the overall performance is improved. Due to the changed coefficients there is also an effective region for $\kappa < 1$ with moderately increased performance.

6.3.3 Generalization

To investigate the generalization capabilities of the approaches, the training set was reduced to 50%, 25%,and 10% of the 50 different regular patterns while the test set still contained the complete set. The results are displayed in Figure 6.6. Eremin's algorithm shows a strong

[3] except for the incompatible cases where Eremin's algorithm slowly converges to zero

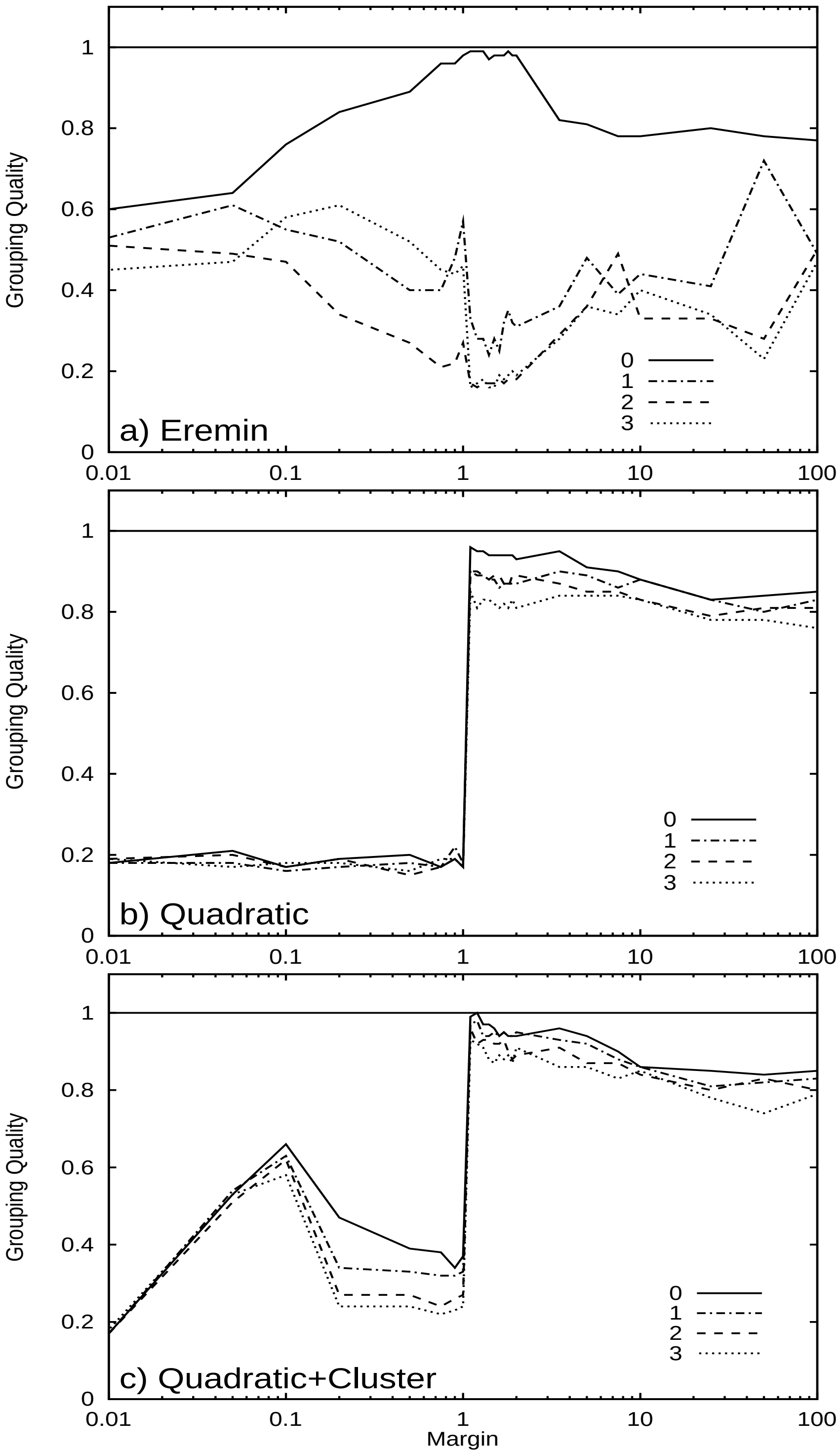


Figure 6.5: Performance of Eremin's algorithm and quadratic consistency optimization. The plots show the proportion of correct grouping results averaged over 1000 simulations dependent on the margin κ. The curves give the performance if 0 to 3 additional spurious patterns are added to the regular training set. Eremin's algorithm shown in a) cannot deal with spurious patterns. The quadratic minimization approach shown in b) shows good performance for $\kappa > 1$ and is robust under addition of spurious patterns. The use of the cluster conditions causes further improvements.

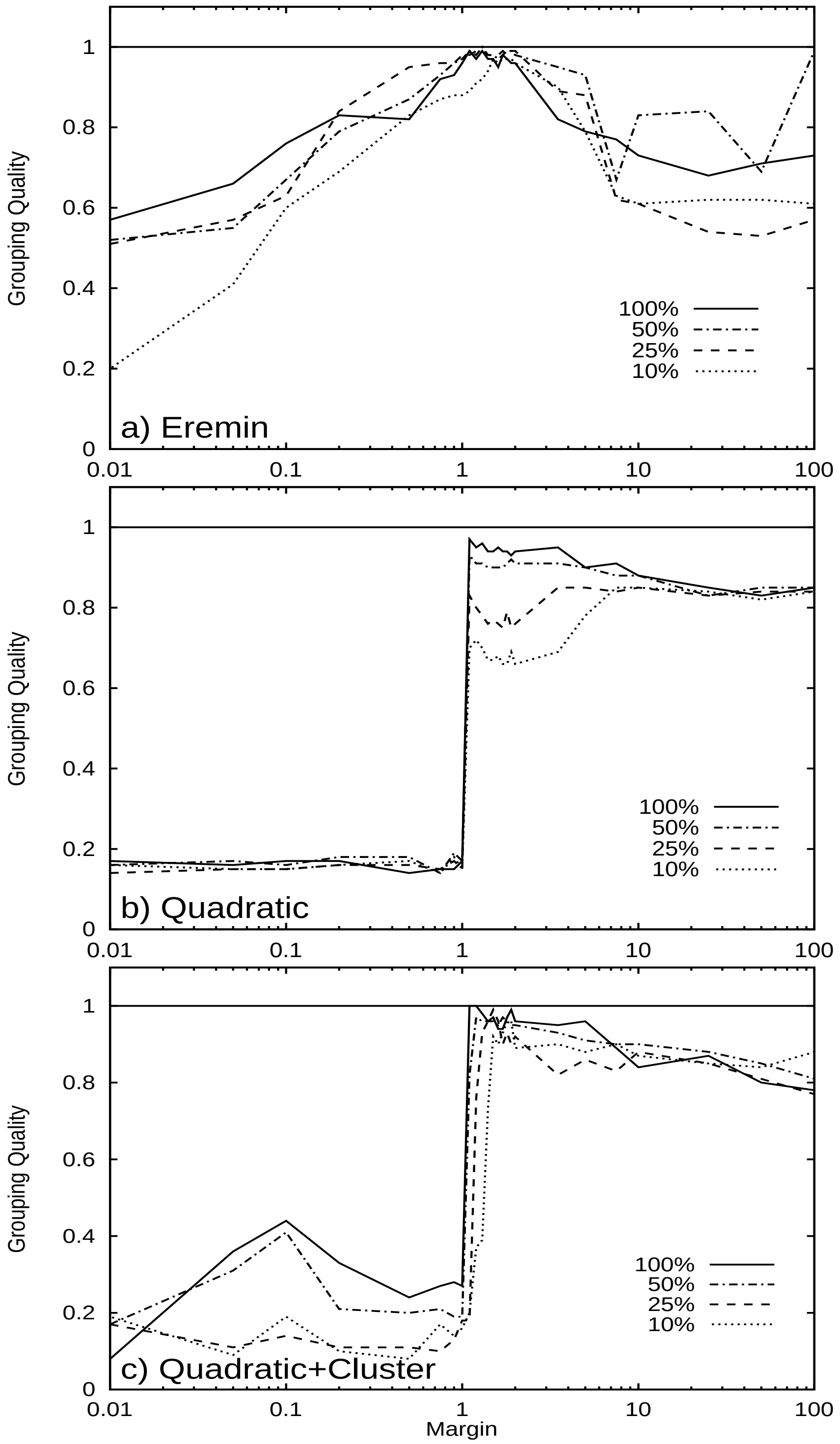


Figure 6.6: Generalization of Eremin's algorithm and quadratic consistency minimization. The curves show the performance if $100\%, 50\%, 25\%, 10\%$ of the 50 regular patterns is used for training. As is shown in a), Eremin's algorithm has a strong generalization capability, and needs only very few training examples in its optimal margin regime. Quadratic consistency optimization in b) shows a degradation for few training examples, which can be compensated by larger margin values. The additional cluster conditions in c) cause a strong improvement over a wide margin range compared to b).

generalization capability with only little degradation for less training data in its optimal margin regime around $\kappa \approx 1$. The strong constraints to fulfil all consistency relations cause a strong restriction on the weight parameter space which leads to good generalization in this setting.

The quadratic consistency optimization algorithm shows a stronger degradation for fewer training examples, however the effect can be compensated to some extent by larger margin values. Applying the additional cluster constraints lead to a much better performance, since the additional conditions impose stronger constraints on the lateral interactions which in turn requires less training data.

6.4 Application to Cell Segmentation

In Section 5.2 it was shown how a manually tuned lateral interaction can be used to perform segmentation of fluorescence cell images with the CLM. The application of a supervised learning algorithm provides an interesting alternative to that heuristic design process and is presented in the following.

6.4.1 Setup

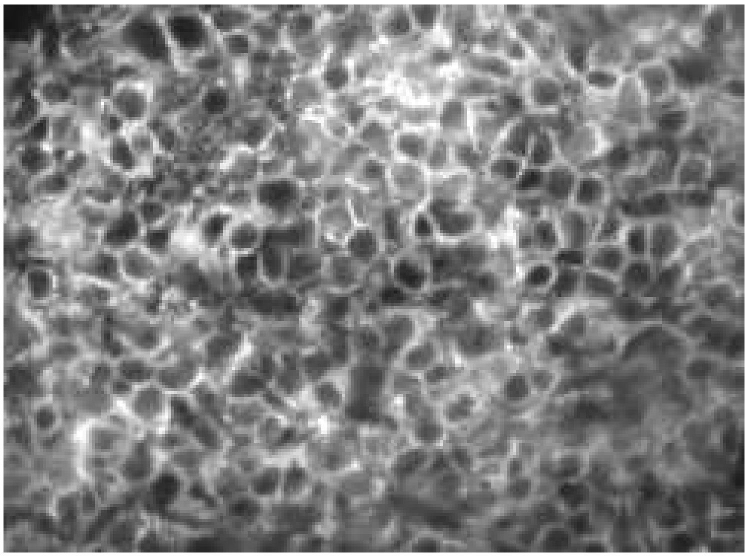
Training Data

Figure 6.7 shows a fluorescence microscopy image from a tissue section containing lymphocyte cells (see Section 5.2). In the bottom row smaller image patches taken from the larger image are displayed, where individual cell regions were manually labeled to obtain training data for the learning process. For each of the image patches, a training vector consists of a list of labelled edge features parametrized by $\mathbf{m}_r = (\mathbf{p}_r, \mathbf{n}_r)$, where $\mathbf{p}_r$ is the position in the image and $\mathbf{n}_r$ is a unit local edge orientation vector computed from the intensity gradient as described in Section 5.1.1. For a 40×40 pixel image this amounts to a set of 40^2 labeled edge features.

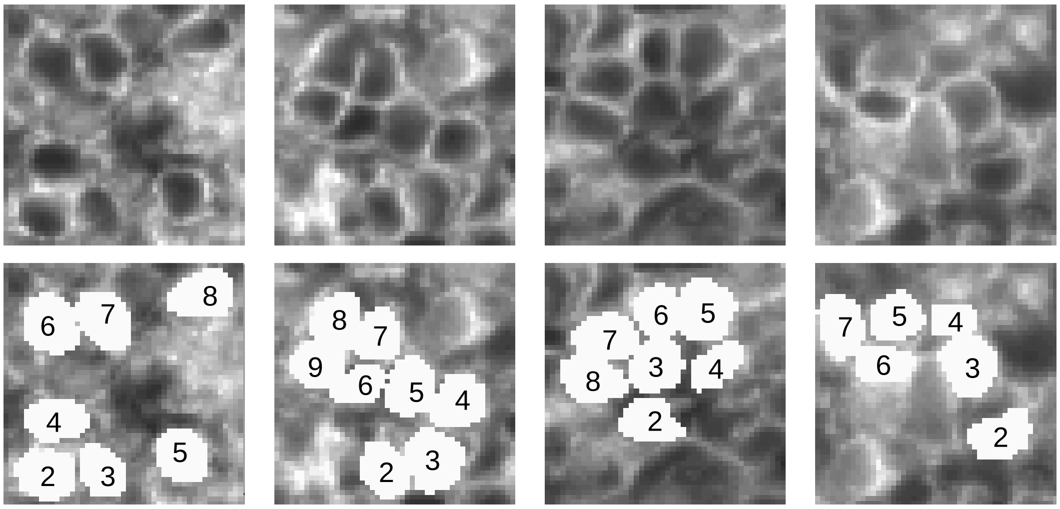
Since the figure-ground mechanism is crucial for the cell segmentation application, features which are not labelled as part of a cell obtain the corresponding background label, which is assumed to be given by $\alpha = 1$. It is also assumed that each training pattern contains one additional free layer, to enable the learning algorithm to generalize over the number of layers.

Interaction Template 1

The first interaction template is constructed by taking the cell interaction of Nattkemper et al. (2000a) and breaking it into three components, which make up the interaction basis templates. The three components are given by the background layer self-coupling g^1, the local orientation-selective field g^2, and the non-local orientation-insensitive global inhibition g^3. The two components g^2 and g^3 are visualized in Figure 6.8, for an explicit definition see Appendix A. Therefore the parameter c_1 corresponds to the background layer strength m, c_2 is the strength of the local angular and distance-dependent field and c_3 corresponds to the strength of the global inhibition k. The task of learning is to adapt the relation between the three parameters.



a) Fluorescence micrograph



b) Manually labelled training patterns

Figure 6.7: Labelled training patterns from a fluorescence micrograph. The labeling was done by hand, unlabeled features obtain the background label $\alpha = 1$.

Interaction Template 2

The second set of basis functions is more flexible, since the angular local interaction is not fixed a priori as in the previous example. Instead, there is a parameter for each particular combination of pairwise orientations of edge elements. This requires first a discretization of the space of orientations. Since the interaction should also be direction-dependent to fully exploit the gradient information, it is useful to turn the unit-vector representation into an representation with an angular variable $\theta \in [0, 2\pi[$. To make the interaction rotation-invariant the angles are taken relative to the difference vector $\mathbf{d} = \mathbf{p}_2 - \mathbf{p}_1$.

The basis interactions g^1 and g^2 are the same as for the previous interaction 1 and give the background coupling and the global inhibition, respectively. The angles θ_1 and θ_2 are

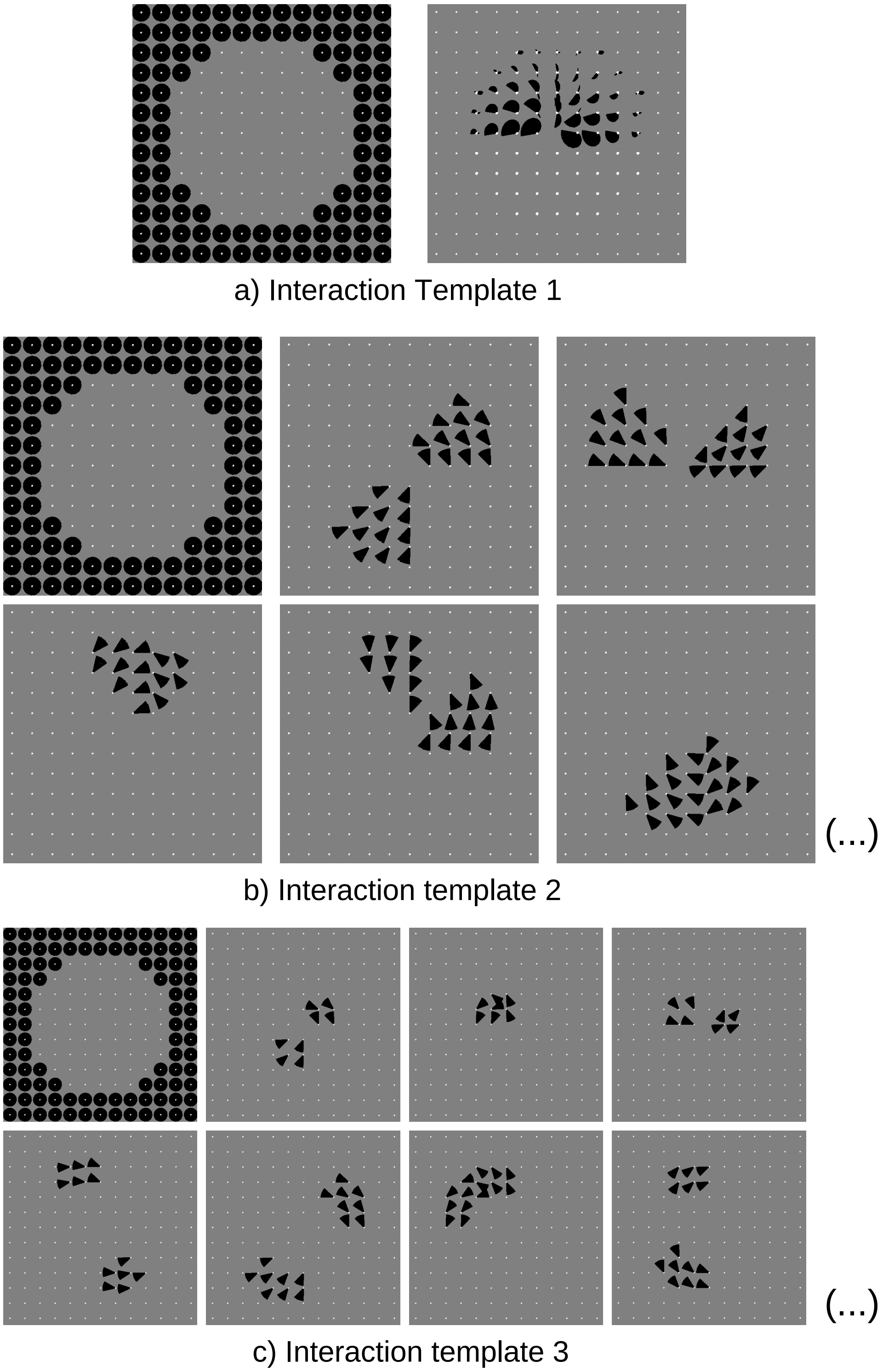


Figure 6.8: Lateral interaction templates. The plots show the interaction fields of a single edge at the center for each component of a basis template set. In b) only 6 out of 38 and in c) only 8 out of 74 components are shown.

discretized by partitioning the interval $[0, 2\pi[$ into 8 subintervals. For each combination of the two edge orientations there is an interaction template generated (see Figure 6.8). Since the interaction must be symmetric under feature exchange, this does not result in $8^2 = 64$ different combinations, but only 36 independent templates. The precise definition is given in Appendix A. Apart form the discretization, the interaction represents the most arbitrary angular-dependent interaction within the local neighbourhood.

Interaction Template 3

The third lateral interaction is identical to the second one regarding the background layer, global inhibition and angular component. The only difference is a further distance-dependent subdivision within the local neighbourhood which is further separated into two partitions. This offers a greater flexibility, since the angular interaction can be more locally differentiated. Since this gives a second set of 36 angular components, the total number of parameters is 74.

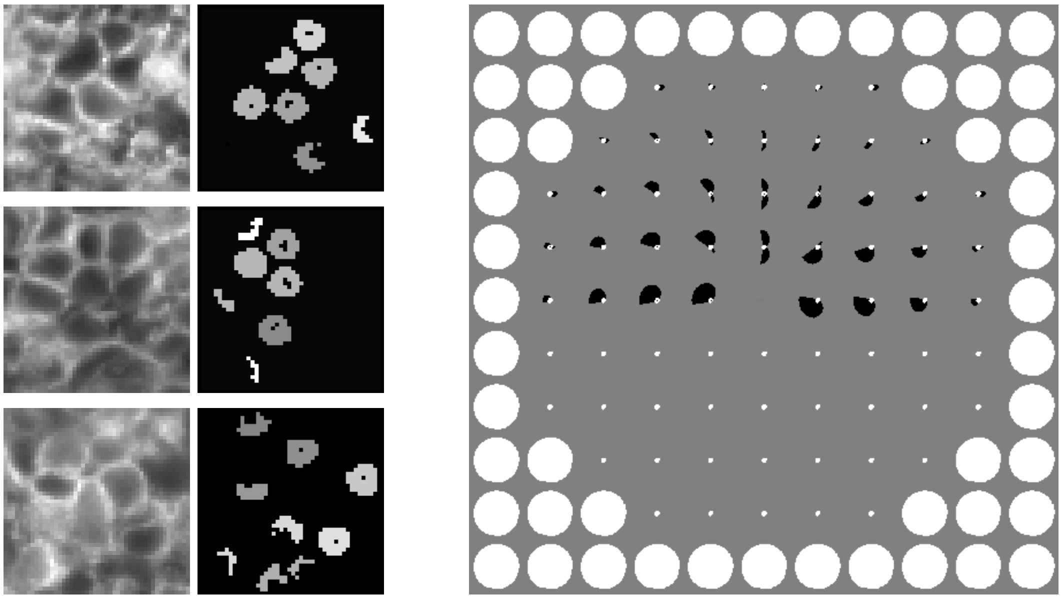
6.4.2 Results

The performance of the learning approach was investigated by choosing one of the manually labelled patterns as a training pattern, which had to contain at least 6 cells to provide sufficient data for the training procedure. After training was completed, a different image patch was selected as a test pattern and the CLM grouping was performed with the lateral interaction learned before.

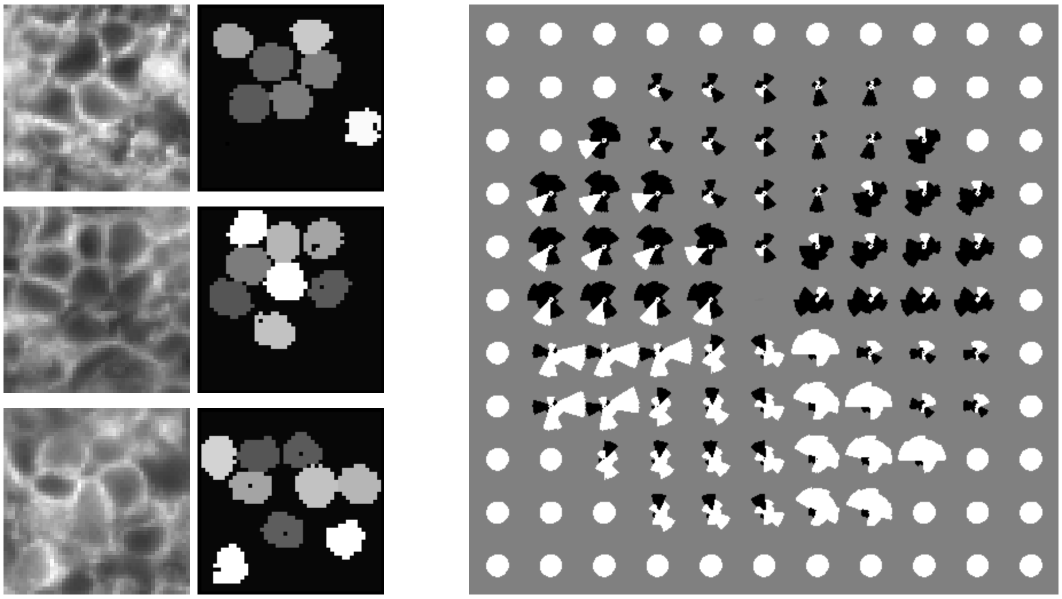
Eremin's algorithm is not applicable in this case, since for all the interaction templates and training patterns considered the algorithm converged to the zero vector. This shows that for this example of real-world data the incompatible case is generic. The incompatibilities may enter either through the noise inherent to the imaging technique, the considerable form variation of the cells, or simply by an inappropriate manual labeling of the training patterns. All these factors, however, cannot be avoided, and a robust learning approach must perform a reasonable averaging over all the training data to extract the relevant information.

The quadratic consistency optimization was performed as described in the previous section with the background and free layer modifications, and with or without the additional cluster conditions. The results have shown that the application of the cluster condition requires a slight modification. Its main effect is to cause a greater "repulsion" of groups by increasing the overall inhibition level. It is necessary to introduce a weighting factor to control the rise of the inhibition. The effect of the cluster condition in the convex quadratic cost function can be reduced by weighting the components $C_{\nu}^{\mathbf{k}}$, which capture the contribution of the cluster conditions. An appropriate choice was found to be given by ${C'}_{\nu}^{\mathbf{k}} = 0.1 C_{\nu}^{\mathbf{k}}$. This introduces an additional free parameter, which affects the overall inhibitory level. Another important target of the learning method is the adaptation of the background layer strength m, which is for all three templates given by the coefficient c_1. Since m must be chosen in relation to the sum over lateral interactions of a feature in a salient group, its magnitude is restricted by $|c_1| \leq K$, while $|c_j| \leq 1$ for all $j > 1$. This ensures that c_1 can be chosen large enough compared to the more extended interactions in the figure layers. Annealing was performed with $\eta = 0.001$, and the vertical coupling was set to $J = 2J_c$ with $J_c = \max(m, \lambda_{\max}\{\mathsf{f}_{rr'}\})$.

The results obtained with the quadratic consistency optimization approach are shown in Figures 6.9 and 6.10. For all the interactions displayed, the margin was given by $\kappa = 10$.

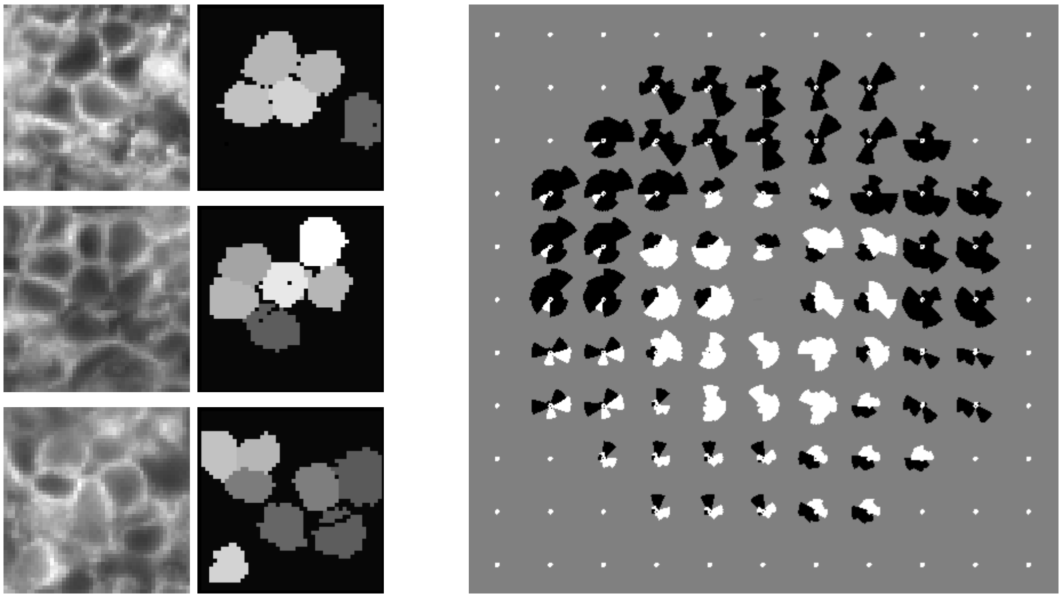


b) Results and interaction field of template 1

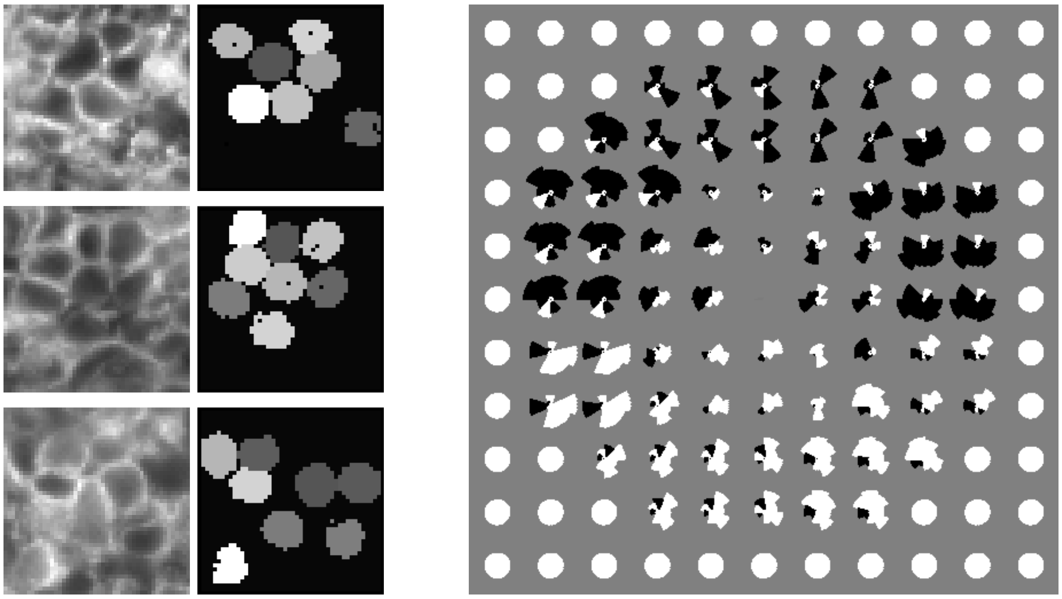


b) Results and interaction field for template 2 with cluster conditions

Figure 6.9: Grouping results and learned interactions for templates 1 and 2. a) shows the resulting superimposed interaction field after learning with template 1 without cluster condition. Since the global inhibition is too strong, only small and partially segmented groups are formed. With cluster condition the inhibition is even stronger (not shown). b) shows the result for template 2 with cluster condition. The interaction also supports gradients which “turn right”, but is more broadly tuned. In the local neighbourhood the raised inhibition is used to suppress non-convex gradient feature arrangements.



a) Results and interaction field for template 3



b) Results and interaction field for template 3 with cluster conditions

Figure 6.10: Grouping results and interaction field for template 3. In a) the resulting superimposed interaction field is shown after learning with template 3 without cluster condition. The local inhibition in the direct neighbourhood is large while the global inhibition is very small. The outer "ring" of the local field is rather insensitive to orientation, which causes the formation of more extended cell segments. Small cells are not resolved due to short-ranged local inhibition. In b) the additional cluster condition was applied which raises the level of global inhibition. The grouping results are similar to the results of template 2.

The grouping results were not very sensitive to κ in a range of $5 < \kappa < 100$.

For template 1 the result of the optimization algorithm is essentially different from the interaction as it was proposed by Nattkemper et al. (2000a). The level of inhibition is much larger compared to the lateral angular part, with an even higher level if the cluster condition is applied. The results show that this induces a tendency to form only small groups where some cells are only partially segmented.

For template 2 the resulting interaction field has a similar overall orientation preference of supporting gradients which "turn right". However, the local inhibition is more pronounced in certain directions and there exist other additional excitatory interactions. The grouping results show a good segmentation performance where most of the salient cells are detected as single groups. There are some spurious groups where a dark image region forms an additional group. For the shown result the cluster condition was applied. Without the conditions the results are similar to template 3 with cluster conditions which is discussed in the following.

For template 3 with cluster conditions the grouping results are very similar to the results obtained with template 2 and the angular distribution of excitatory and inhibitory interactions is analogous. Due to the additional distance-dependent subdivision, the inner interactions are weaker than the outer angular component. Without cluster condition, the local inhibition is enhanced, while the global inhibition is strongly reduced. The grouping results shows that this leads to the formation of spatially more extended groups, while smaller cells are not segmented due to stronger short-range inhibition.

6.5 Discussion

The learning algorithm suggested by Pelillo & Fanelli is only applicable to cases where the training patterns lead to consistent target inequalities. The presence of spurious patterns causes a breakdown of the learning method which makes its application to raw real-world data infeasible. Nevertheless, it exhibits an interesting generalization capability on sufficiently "clean" data. Therefore a possible heuristics to save its performance could be an adequate preprocessing of the training data which excludes inconsistent data outliers.

The proposed quadratic consistency optimization algorithm performs well also under the addition of spurious data. The reason is the approximate solution approach to the target inequalities which also allows to violate the consistency relations for single patterns. The additional cluster constraint can be used to improve its performance with regard to the grouping quality and the generalization from small training sets. The generalization is better, since more information is obtained from the training data by implicitly constructing "negative" examples from the data.

For the application to cell segmentation, the two more general lateral interaction templates achieved a good grouping performance. The template with components constructed from the interaction of Nattkemper et al., however, could not be properly adapted from the training data. In that case the solution of the convex optimisation problem does not coincide with the manually optimised interaction coefficients, which shows some limitations of the approach. On the contrary, for the two other templates alternative interactions were found which cause a similar and appropriate grouping behaviour applied to the cell images. The resolution of the segmentation process could be slightly enhanced by using the weighted cluster conditions. This is, however, paid by the introduction of another free parameter which must be manually adjusted.

The results show that appropriate lateral interactions can be obtained from the quadratic consistency optimization approach. The only a priori conditions which were used for the template design were the properties of locality, symmetry, and translation and rotation invariance. Although still some parameters must be manually adjusted, their number is much smaller than for the complete design of a lateral interaction from scratch.

It remains an interesting question if some of the heuristic conditions could be dropped if a more flexible optimization problem was considered. The presented quadratic consistency optimization method is based on choosing equal margins for all consistency inequalities. There exist, however, other approaches to large margin classification, like support vector machines (Vapnik 1995), where more sophisticated methods were suggested for appropriate margin determination. The application of similar methods to the supervised learning of CLM interactions provides an interesting field for future work.

Chapter 7

Backtracking Deterministic Annealing

This chapter continues the investigation of methods for adapting the interaction weights of the competitive layer model neural architecture by learning processes. The focus is to optimize the lateral interaction structure for the CLM in order to improve its capabilities for solving complex constraint satisfaction problems. The basis of the presented "backtracking deterministic annealing" method (Wersing & Ritter 1999a) is a type of learning algorithm, which is running online, as opposed to the learning approach of the last chapter, where the weights were determined directly from the training data.

First the solution strategies are analyzed which are implemented by deterministic annealing in the analog neural representation of Potts-Mean-Field (PMF) and penalty-function-based competitive layer model (CLM) neural networks. The analysis reveals strong analogies between the two approaches. The key idea of the extension of these plain deterministic annealing approaches is motivated by classical backtracking algorithms. It is shown that their ability for iterative local pruning of the search space can be implemented within the framework of deterministic annealing by introducing local temperature parameters which are "reheated" when locally unresolved conflicts occur. To achieve the pruning of the search space, reheating is accompanied by a modification of the constraint-implementing weight matrix to reduce the chance of reentering the same local configuration. The weight changes provide a learning mechanism that facilitates the generation of a solution for subsequent runs. The benefits of the resulting "backtracking deterministic annealing" algorithm (BDA) are demonstrated by applying it to a pentomino tiling problem. It is shown that the method reliably finds perfect solutions to the task, while the plain deterministic annealing approach for both PMF and CLM is unable to solve the task in a comparable or even considerably larger number of iterations.

7.1 Introduction

Combinatorial optimisation problems have a long tradition in the neural network field dating back to the seminal work of Hopfield and Tank (Hopfield & Tank 1985). While the early approaches suffered from problems of spurious and local minima, the later developed deterministic annealing (see Rose 1998 for a recent review) or Potts-Mean-Field annealing networks (Peterson & Soderberg 1989) provide techniques which are competitive with algorithms like simulated annealing with respect to computation time and solution quality.

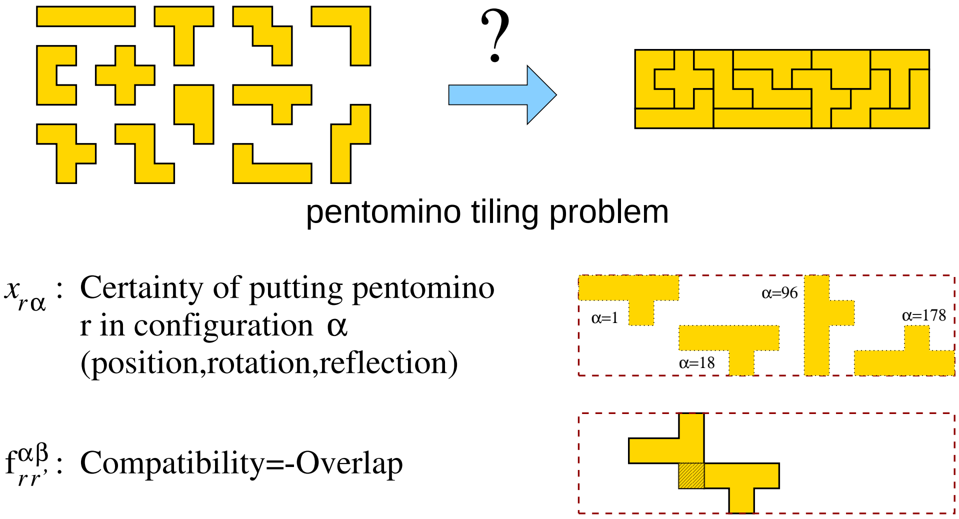


Figure 7.1: Pentomino tiling as a constraint satisfaction problem. It is a challenging puzzle to arrange the twelve pentominoes such that they tile a rectangle (4×15 in this case). The variables $x_{r\alpha}$ represent the certainty of placing a pentomino r in a particular position, labelled by α. If the compatibilities are chosen as the negative overlap of two parts, then the global minimum of the resulting quadratic cost function is a feasible labeling solution.

However, the complexity of the constraints can pose problems requiring the development of special techniques. An example is the matching (or "two-way") assignment constraint[1] in the Travelling Salesman problem (Kosowsky & Yuille 1994). Polyomino tiling problems (Golomb 1994) provide much more complex interlocking constraints that challenge existing deterministic annealing approaches even further, since they may require complex reconfigurations when local constraints become violated in later stages of the annealing phase. Constraint satisfaction also plays a major role in perceptual processes, since the interpretation of the sensory input is done under the physical constraints of illumination, shading, occlusion, transparency, and various other factors. A simple example is the three-dimensional interpretation of polygonal line drawings which was considered by Waltz (1975) as one of the first applications of iterative constraint satisfaction solution algorithms.

Backtracking has been very successfully applied to constraint satisfaction problems (Bitner & Reingold 1975). This suggests that it might be possible to find an algorithm that can combine some of the benefits of backtracking with the superb capabilities of deterministic annealing algorithms to explore complex search spaces. In the following, an approach is suggested that extends deterministic annealing by introducing a local temperature control scheme which is accompanied by modifications of the constraint-implementing weight matrix to reduce the chance of reentering the same local configuration.

Let us assume that the problem is stated as the minimisation of a quadratic energy, as considered in the previous chapters,

$$E = -\frac{1}{2} \sum_{rr'\alpha\beta} \mathrm{f}_{rr'}^{\alpha\beta} x_{r\alpha} x_{r'\beta} \to \min. \tag{7.1}$$

[1]assuring a Hamiltonian (closed) path

subject to the weighted labeling constraint

$$\sum_{\alpha} x_{r\alpha} = 1, \tag{7.2}$$

and $x_{r\alpha} \in \{0, 1\}$. $x_{r\alpha}$ denotes the certainty of assigning a label $\alpha \in \{1, \ldots, L\}$ to a feature r among L alternatives and $\mathsf{f}_{rr'}^{\alpha\beta}$ encodes the mutual compatibility of assigning label α to feature r and label β to feature r'.

An example for a constraint satisfaction problem and a possible choice of the compatibilities is shown in Figure 7.1. The variables $x_{r\alpha}$ represent the certainty of placing a pentomino r in a particular position, labelled by α. The compatibility $\mathsf{f}_{rr'}^{\alpha\beta}$ of assigning pentomino r to position α and pentomino r' to position β is given by the negative overlap of the two parts. Upon this choice, the energy of a configuration attains its global minimum if no overlaps exist and thus the binary final assignment vector $\mathbf{x}$ characterizes a feasible solution. An important observation is that the energy value of the global minimum is equal to zero and independent of the value of the constraint implementing weights. This implies that their magnitude is somehow arbitrary, however, greatly influences the structure and number of the *local* minima.

7.2 Dynamical Solution Strategies of Deterministic Annealing Algorithms

This section provides an analysis of the solution strategies of common deterministic annealing algorithms, applied to combinatorial optimisation. It gives a comparison between Potts mean field models and the competitive layer model, which reveals strong analogies with regard to the search strategy that is performed by the two approaches.

The Potts-Mean-Field neural network approach to the solution of this problem (see Section 2.2.2) is given by minimising the mean-field free energy F_{PMF} with

$$F_{\text{PMF}} = E + T \sum_{r\alpha} x_{r\alpha} \log x_{r\alpha} \tag{7.3}$$

subject to (7.2). The actual solutions are obtained by iterating the following dynamics (parallel, serially, or asynchronously) together with a gradual lowering of the temperature T

$$\dot{x}_{r\alpha} = -x_{r\alpha} + \frac{\exp(-TE_{r\alpha})}{\sum_{\beta} \exp(-TE_{r\beta})}, \tag{7.4}$$

where

$$E_{r\alpha} = -\frac{\partial E}{\partial x_{r\alpha}} = \sum_{r'\beta} \mathsf{f}_{rr'}^{\alpha\beta} x_{r'\beta} \tag{7.5}$$

is the negative local gradient of the original energy. For fixed T the dynamics (7.4) always converges to a local minimum of the free energy (7.3). For high T the free energy is dominated by the convex entropy term with unique minimum at $x_{r\alpha} = 1/L$. Lowering T, the free energy gains more and more of its original structure until for $T \to 0$ the original energy E is recovered.

The CLM approach to solving the minimisation problem (7.1) under constraints (7.2) (see Section 4.1) is to minimize the following different energy function of the form

$$F_{\text{CLM}} = E + J\sum_r \Big(1-\sum_\alpha x_{r\alpha}\Big)^2 + T\sum_{r\alpha} x_{r\alpha}^2 \tag{7.6}$$

subject to $x_{r\alpha} \geq 0$. The weighted labeling constraint (7.2) is enforced by an additional penalty function dependent on J and the logarithmic entropic constraint term is replaced by a quadratic diagonal term. The quadratic choice in (7.6) leads to a piecewise linear gradient descent dynamics of the form

$$\dot{x}_{r\alpha} = J\Big(1-\sum_\beta x_{r\beta}\Big) + E_{r\alpha} - T\,x_{r\alpha} \ \ , \tag{7.7}$$

subject to the simple non-linearity $x_{r\alpha} \geq 0$.

According to Theorem 7 (see Section 4.4.2) the dynamics (7.7) converges towards an unambiguous labeling. Due to the "soft" enforcing of the weighted labeling constraint, however, the final output $x_{r\hat{\alpha}(r)}$ is modulated by an amount given by $E_{r\hat{\alpha}(r)y}/J$. While this can be rather usefully exploited in certain vision labeling tasks (Wersing & Ritter 1999b), here we choose J sufficiently large such that the deviations are small and the final output state can be simply rectified by $x'_{r\alpha} = \theta(x_{r\alpha})$ where $\theta(0) = 0$ and $\theta(x) = 1$ for $x > 0$.

The following linear analysis reveals that the two deterministic annealing algorithms for PMF and CLM exhibit a similar structure at the dynamical level though using different means of enforcing constraints at the energy function level.

In the limit of large T both energies (7.3) and (7.6) share a common global minimum $x_{r\alpha} = 1/L$ which is then the only attracting fixed point of both dynamics (7.4) and (7.7). Linearising the vector field defining the PMF dynamics in (7.4) about this fixed point gives

$$\dot{x}_{r\alpha} = \frac{1}{L} - x_{r\alpha} + \frac{1}{LT}\Big(E_{r\alpha} - \frac{1}{L}\sum_\beta E_{r\beta}\Big). \tag{7.8}$$

To compare the two linear vector fields in (7.7) and (7.8), a projection onto two orthogonal subspaces is helpful where we consider the DC subspace with $\dot{x}^{\text{DC}}_{r\alpha} = 1/L\sum_\beta \dot{x}_{r\beta}$ and the AC subspace with $\dot{x}^{\text{AC}}_{r\alpha} = \dot{x}_{r\alpha} - \dot{x}^{\text{DC}}_{r\alpha}$ respectively. While the AC subspace captures components which cause differences between labels α in a column r, the DC subspace carries the dynamics of the summed activity in a column r (see also the linear analysis in Section 4.2.3). To compare the PMF and CLM vector fields in the AC and DC subspaces the following variables are helpful, which can be considered as deviations of the annealing gradient directions from the original gradients $E_{r\alpha}$.

$$\begin{aligned} E^{\text{PMF}}_{r\alpha} &= E_{r\alpha}/LT - x_{r\alpha}, \\ E^{\text{CLM}}_{r\alpha} &= E_{r\alpha} - Tx_{r\alpha}. \end{aligned} \tag{7.9}$$

After computing the projections of the vector fields into AC and DC components and inserting definitions (7.9) one obtains for PMF:

$$\dot{x}^{\text{AC}}_{r\alpha} = E^{\text{PMF}}_{r\alpha} - \frac{1}{L}\sum_\beta E^{\text{PMF}}_{r\beta}, \tag{7.10}$$

$$\dot{x}^{\text{DC}}_{r\alpha} = \frac{1}{L} - \frac{1}{L}\sum_\beta x_{r\beta}. \tag{7.11}$$

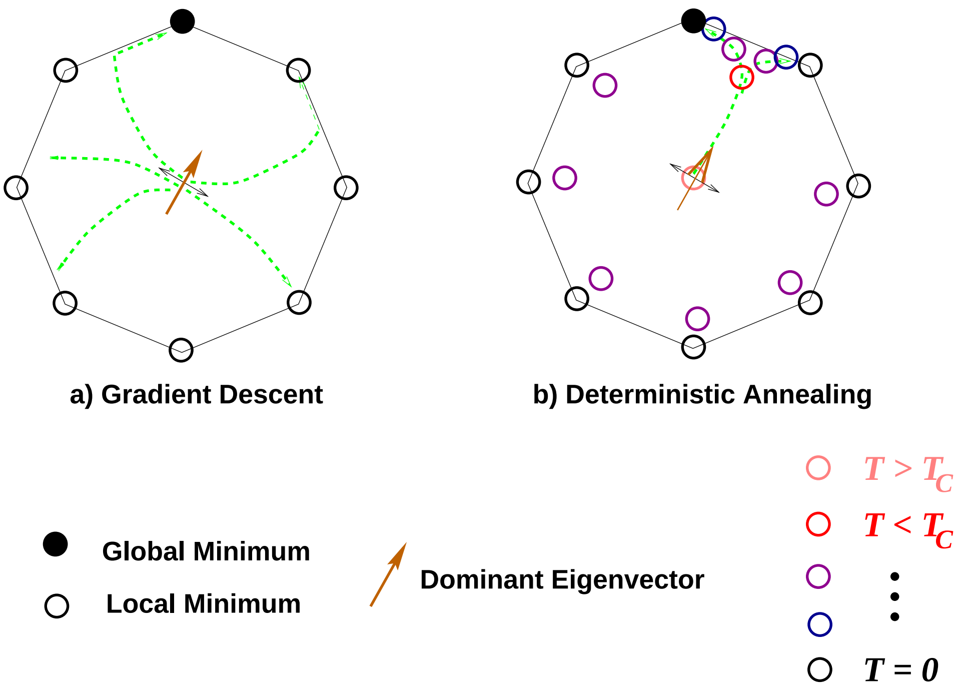


Figure 7.2: Sketch of system trajectories for a direct gradient descent algorithm and deterministic annealing. For the type of minimisation problems considered here, the local and global minima reside at the corners of the state space. a) For direct gradient descent most directions in search space are equally attractive and there is only a small bias towards the dominant eigenvector direction. Consequently the chance of reaching local suboptimal minima is high. b) By performing deterministic annealing, all directions except for the dominant eigenvector direction are suppressed. If the global minimum resides in this direction, then the probability of reaching the global minimum is enlarged. The dominant eigenvector represents global contextual information for the problem.

For the CLM we obtain:

$$\dot{x}_{r\alpha}^{\mathrm{AC}} = E_{r\alpha}^{\mathrm{CLM}} - \frac{1}{L}\sum_{\beta} E_{r\beta}^{\mathrm{CLM}}, \tag{7.12}$$

$$\dot{x}_{r\alpha}^{\mathrm{DC}} = J\Big(1 - \sum_{\beta} x_{r\beta}\Big) + \frac{1}{L}\sum_{\beta} E_{r\beta}^{\mathrm{CLM}}. \tag{7.13}$$

This shows that in the DC space both linear dynamics (for sufficiently large J) are driven towards the affine constraint surface (7.2). In the AC-subspace both dynamics have a very similar structure, differing mainly in the way the temperature T enters the dynamics in (7.9): For the PMF the effective matrix driving the AC dynamics is divided by LT with a constant negative identity matrix, while for the CLM the original gradient is kept and only the diagonal is shifted towards negative eigenvalues. The critical temperatures at which the fixed point gets unstable are given by the values of T for which the largest eigenvalue of the matrices defining the linear vector fields (7.7) and (7.8) is equal to zero. The resulting values are then given by

$$T_c^{\mathrm{PMF}} = \lambda_{\max}\{\mathsf{f}_{rr'}^{\alpha\beta} - 1/L \sum_{\beta} \mathsf{f}_{rr'}^{\alpha\beta}\}/L, \tag{7.14}$$

$$T_c^{\mathrm{CLM}} = \lambda_{\max}\{\mathsf{f}_{rr'}^{\alpha\beta} - J\delta_{rr'}\}. \tag{7.15}$$

If T falls below the critical T_c only the dominant eigenmodes in the AC subspace are expressed by the dynamics while other eigenmodes are suppressed due to the negative eigenvalue shift for both PMF and CLM. This is a central mechanism of the deterministic annealing approach as has also been discussed for clustering applications (Rose 1998). Therefore the success of the deterministic annealing heavily depends on the property, that the maximum eigenvalue directions are directions within the search space with a higher probability of a global minimum occurring (see Figure 7.2). We may consider this as the extraction of some global properties encoded implicitly into the compatibility weights of the optimisation problem.

For lower T the dynamics gets more complex and there are phases of linear growth within restricted configuration subsets at bifurcations and phases of tracking the local equilibria of the free energy which move continuously with respect to T towards the boundary. Due to the softmax nonlinearity the PMF performs a more smooth sliding of equilibria to the minima of E in the $T \to 0$ limit, while the "hard' nonlinearity introduced by the condition $x_{r\alpha} \geq 0$ causes a sharpened exploration of the "corners" of the state space. As is discussed in Section 5.1 this causes a more deterministic behaviour of the CLM algorithm for slow annealing.

7.3 Combining Backtracking and Deterministic Annealing

Backtracking algorithms (Bitner & Reingold 1975) solve constraint satisfaction problems by iterative reduction of the combinatorial search tree whenever local constraint conflicts are detected. The effective pruning of the search space leads to a speed gain which makes them applicable to a wide range of constraint satisfaction problems. While the local search for feasible solutions is exhaustive, however, global properties which follow from the structure of the problem, i. e. implicitly encoded into the weight matrix of the energy (7.1), are not directly exploited and can only be taken into account by hand by choosing an appropriate tree-search strategy. On the other hand, the success of deterministic annealing neural networks is based on extracting global information by dynamically exploring maximum eigenvalue directions of the search space. They can also be simulated on parallel and analog hardware, offering scalable speed gains. This motivates the formulation of a hybrid approach which combines the advantages of the two approaches.

We suggest the implementation of a backtracking deterministic annealing (BDA) network in the following way: The starting point is one of the deterministic annealing approaches like PMF or CLM. The global temperature T is replaced by individual temperatures T_r for each feature r to allow for a locally controlled rearrangement of labels. In the beginning, annealing is performed for all T_r synchronously, however, when $T_r < T^{(1)}$, where $T^{(1)} \ll T_c$ is a typical temperature causing a pronounced decision for one of the labels, say (r, α), it is checked whether a local constraint conflict of (r, α) with another assigned (r', β) has occurred. If this is the case, the weight $\mathsf{f}_{rr'}^{\alpha\beta}$ is changed to $\mathsf{f}_{rr'}^{\alpha\beta} - \Delta$, causing an energy penalty for this subconfiguration. The local temperatures T_r and $T_{r'}$ are raised to an intermediate value $T^{(2)}$ which must be sufficiently high to restore a superimposed

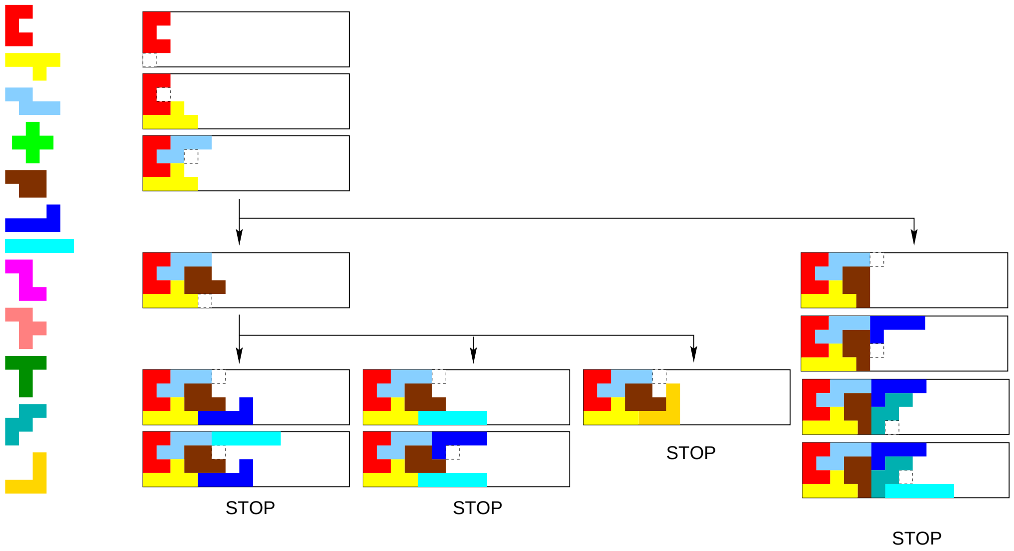


Figure 7.3: Backtracking search tree. A backtracking algorithm consists of a defined prescription of going through the space of all possible configurations in a tree-like like manner. Whenever a subconfiguration is encountered for which the tiling cannot be completed with the remaining pentominoes, the algorithm "backtracks" to the last position, where an alternative choice was possible. In the picture the rule is to place the uppermost pentomino from the array at the left into the most left and most upper free position in the rectangle.

state for features r and r'. The chance of reentering the previously encountered erroneous state is now diminished due to the weight change. Performing these steps synchronously at multiple sites leads to a search procedure, which can effectively escape from local minima. The system "learns" to optimise its own weights for obtaining a solution to the optimisation problem. The explicit definition of the algorithm is given in Appendix B.

Quite generally, constraint satisfaction problems like tiling problems are modelled by inhibitory interactions of conflicting alternatives. In that case the energy level of a proper solution forming a global minimum does not depend on the constraint-implementing weights and by changing these weight components only local minima are affected. The task of the BDA approach is not to optimise these weights with respect to the original problem alone but with respect to the problem *and* the search strategy implemented by the annealing dynamics.

7.4 Application to Pentomino Tilings

The twelve pentominoes (Golomb 1994) are all polygons (modulo rotation and reflection) which are composed of 5 connected squares (see Fig. 7.1). They are known to tile $6 \times 10, 5\times 12, 4\times 15,$ and 3×20 rectangles. These rectangular tilings provide difficult constraint satisfaction problems which can be formulated as quadratic minimisation problems (7.1) and which fall into the class of NP-complete packing problems (Yamamoto, Ninomiya, & Asai 1998). Let r denote the rth pentomino, and let $\alpha \in \{1 \dots L_r\}$ denote the L_r possible ways of putting pentomino r onto the given rectangular field. Then we define $\mathsf{f}_{rr'}^{\alpha\beta}$ as the negative number of overlaps of pentomino r at position α with pentomino r' at position β. To allow for a rejection of single "holes", we choose $\mathsf{f}_{rr'}^{\alpha\beta} = -1$ also if a pair of pentominoes

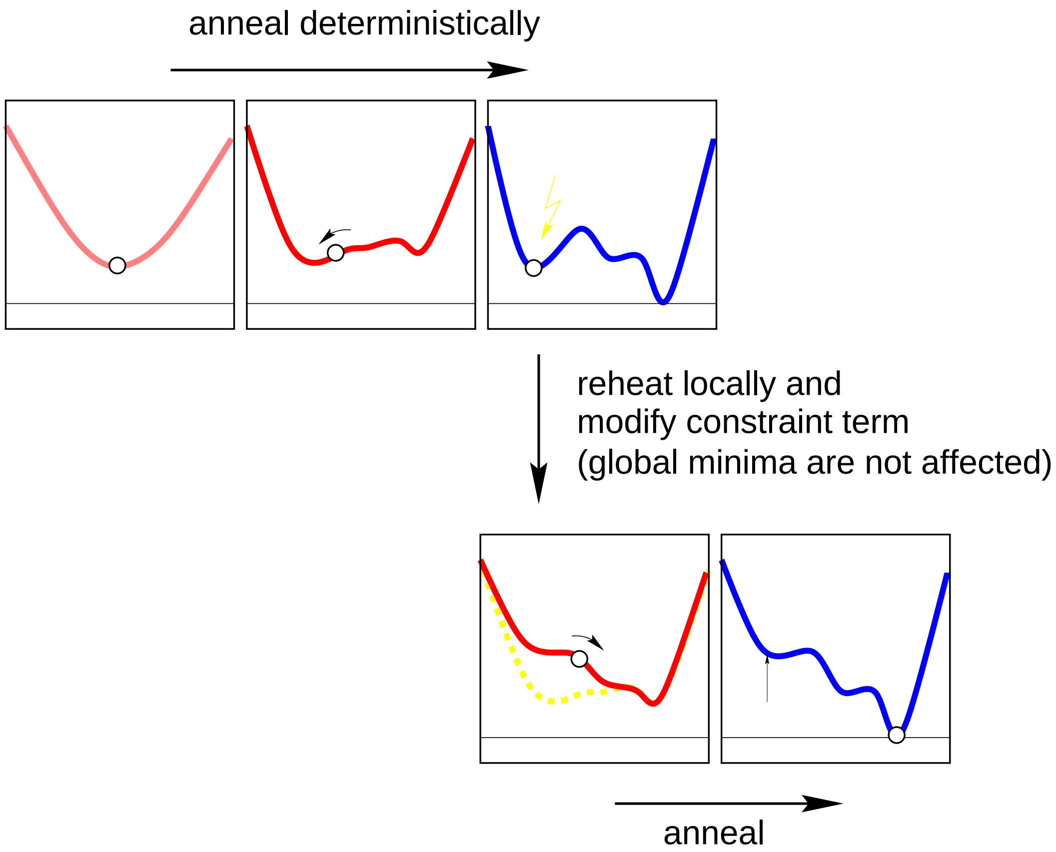


Figure 7.4: Sketch of backtracking deterministic annealing. After annealing, which corresponds to a convex gradual deformation of the energy function, a local constraint conflict is obtained by entering a local minimum. Then the constraint term is modified to “raise” the local minimum and a local reheating is applied. In the next step the local minimum will be avoided and the search continues. The procedure does not affect the global minima, since their position does not depend on the constraint-implementing weights.

surrounds an isolated square. The task is now to find an unambiguous labeling, which forms a feasible solution and is one of the possibly many global minima with zero overlap and thus $E = 0$. To fulfil the conditions of Theorem 1 for the CLM we choose $\mathsf{f}_{rr}^{\alpha\alpha} = 1$ and take $J = 0.01 \cdot \max_{r\alpha} \sum_{r'\beta} |\mathsf{f}_{rr'}^{\alpha\beta}|$ which is sufficient to keep the deviation from the “hard” constraint (7.2) in the CLM small enough.

Pentomino tiling problems can nowadays be solved on standard workstations by plain backtracking algorithms within minutes. The main interest for other solution heuristics lies in the exponential scaling of the problem size for larger problems encountered e.g. in placement and packing problems in VLSI design, where mean-field-annealing has been shown to be competitive with other approaches (C. Aykanat & Haritaoglu 1998). Earlier work on solving smaller tiling problems with neural networks was based on a modified McCulloch-Pitts neuron model performing a dynamical search among discrete configurations of tilings (Takefuji & Lee 1990). Extending the algorithm by a superimposed analog representation of multiple configurations with gradual sharpening, similar to annealing, improves its performance (Yamamoto, Ninomiya, & Asai 1998).

We consider the 4×15 rectangle, for which there exist about 3×10^{24} possible alternatives

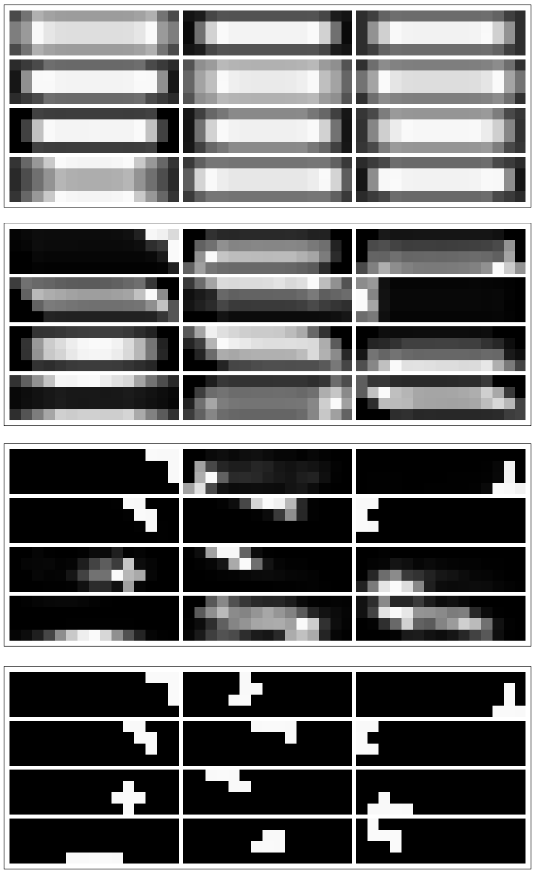


Figure 7.5: Graduated assignment on a 4×15 rectangle during annealing at 4 temperatures using only plain deterministic annealing. Each small field shows the superposition of all the possible placements α of a single pentomino r weighted by the activity $x_{r\alpha}$ of the corresponding label. The final unambiguous output labeling has 2 overlaps.

of placing the twelve pentominoes if arbitrary overlaps are allowed. It is known that there exist 368 different solutions (modulo reflection and rotation). This problem is much more complex than the tiling problem considered in (Takefuji & Lee 1990) where "only" 1.3×10^{14} combinations existed.

Figure 7.5 shows a time-course of annealing for both CLM or PMF. Starting from an undecided state, the assignment is gradually sharpened, until a solution (in this case sub-optimal) is obtained. The intermediate stages reveal, how the interaction of all the superpositions already causes a reasonable placement of parts like the long bar-shaped and the v-shaped pentomino. These constraints are realized by the maximum eigenvalue directions which are first explored due to the analysis in section 7.2. In later stages, however, certain placements which require non-local corrections can not be revised. Figure 7.6 shows histograms of the typical numbers of overlaps generated by CLM and PMF for different annealing speeds. Our PMF simulations were done by asynchronously solving the fixed point equation (7.4). For fast annealing CLM and PMF show similar performance with larger variation than for the slow annealing. With slow annealing the number of different output states is significantly reduced indicating more deterministic system trajectories. This effect is more pronounced for the CLM. Although slower annealing achieves on average

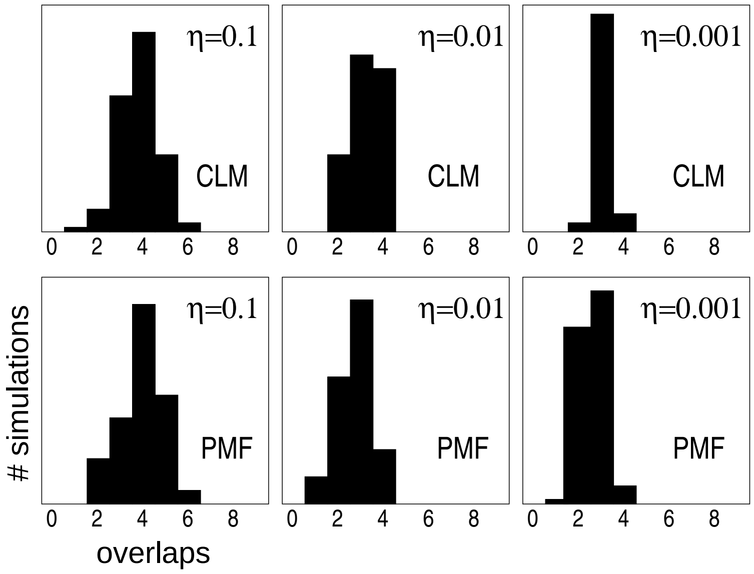


Figure 7.6: Performance of plain DA: Overlaps of output labelings from PMF and CLM at different annealing speeds for 1000 runs ($\eta = 0.9$ and $\eta = 0.99$) and 100 runs ($\eta = 0.999$) (bar heights normalised).

better results for PMF, the effect saturates. In fact, even for 10^5 runs at $\eta = 0.1$ (50 sweeps per run) and 10^4 runs at $\eta = 0.01$ (500 sweeps per run) we found no solution with zero overlaps for both CLM and PMF. Note also that since annealing at $\eta = 0.01$ takes ten times more sweeps, the actual probability of obtaining a 1-overlap solution in ten $\eta = 0.1$ runs is roughly the same, indicating no advantage of slower annealing in that case.

We then performed simulations with the suggested BDA method by choosing $T^{(1)} = T_c/1000$, $T^{(2)} = T_c/10$ and modifying the weight matrix at each detected conflict by $\Delta = 1$ at a threshold of $\theta = 0.8$. Annealing was done rather fast at $\eta = 0.1$. The algorithm terminated on average after 230000 sweeps with a proper zero overlap solution (with a range from 3000 to 800000 sweeps). Although this is a considerable number of sweeps, the plain deterministic annealing did not produce any zero overlap solution for altogether 5×10^6 sweeps. Throughout the iterations there were major rearrangements of single parts, however, we observed that whenever a partially valid solution is found the rearranging is temporarily restricted to the rest. To verify that the weight-changing operation is essential for the solution process we also simulated with $\Delta = 0$ where we obtained no zero overlap solution in 20 runs with 10^6 sweeps each. Another interesting observation is that, if the dynamics is restarted with an adapted weight matrix from a previous solution, the BDA terminates 10-100 times faster. This shows that somehow part of the solution has been learned by the backtracking search process which leads to an optimised weight matrix for the deterministic annealing algorithm. Figure 7.7 also shows an interesting generalisation phenomenon which can be observed. Although translation invariance was not explicitly imposed on the compatibility matrix, the network may, if restarted form the learned compatibilities, converge to another tiling which is composed of similar subconfigurations. This can be interpreted such that the network has learned an association between part compounds and is able to generalize over position.

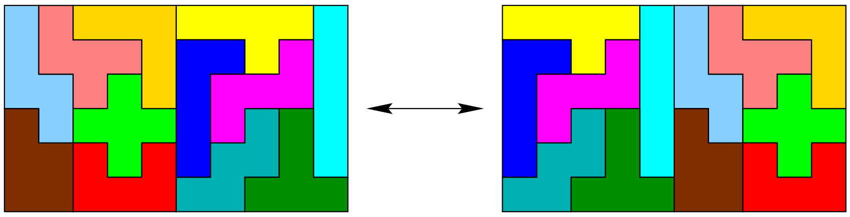


Figure 7.7: Generalization for a symmetric subconfiguration. A network which delivers the left tiling after learning has strong tendency, if reinitialized with learned weights, to also converge to the right configuration which is composed of the same subrectangles.

7.5 Discussion

The presented extension to deterministic annealing for constraint satisfaction problems enables the analog mean-field network to perform a search strategy in the solution space which is similar to classical backtracking algorithms. All operations within the BDA (backtracking deterministic annealing) are of local nature, involving only locally connected neurons which modify their connections due to encountered local constraint conflicts. This offers the opportunity to implement the BDA dynamics in a parallel and analog way.

It is a generic property of the formulation of constraint satisfaction problems as minimisation of quadratic energy functions that the desired global minima which satisfy all the constraints do not depend on the weight components which enforce the constraints. Therefore there is some arbitrariness involved in choosing these weights. The optimal choice depends not only on the energy function but also on the algorithm used to minimise this energy and is usually optimised by hand for a special application problem. The presented BDA network tries to automatically optimise or learn these weights during its exploration of the search space.

For the example of pentomino tilings, the BDA can obtain proper global solutions which satisfy the complex multiple constraints of the tiling problem. Although the exploration takes a considerable number of iteration sweeps and is certainly slower than plain backtracking search algorithms which were optimized for the pentomino tiling problem, our preliminary results indicate that after the optimisation a proper tiling result can be much faster obtained. We therefore consider the BDA a promising new hybrid approach which is especially relevant to problems with complex constraints where deterministic annealing techniques have been successfully applied.

Chapter 8

Conclusions and Outlook

In this chapter the main results of this thesis are summarized and discussed with regard to their relevance and possible future research directions.

One important focus of this work was the implementation of a consistent neurodynamical model for feature binding in complex visual scenes. In the recent years a number of synchronization-based neural binding models were developed which were also successfully applied to real-world data. The complexity of the grouping tasks, however, has required to use stronger algorithmic abstractions from the original dynamical models. The new spatial binding model of the CLM uses linear threshold neurons, which are a standard model for rate-coded activity dynamics in neural networks. Much effort was taken in this thesis to show by mathematical analysis, that both grouping and labeling tasks can be accomplished by this model approach. The particular form of piecewise linear transfer function facilitated a linear analysis of the dynamics which allowed a thorough discussion of the grouping properties related to the lateral coupling schemes. The grouping performance was largely enhanced by applying new results on the stability of general linear threshold networks, which were also established. These results allow to operate the CLM grouping model in a mode, where maximal contextual modulation is possible, which increases the overall grouping quality and decreases the chance of entering spurious grouping states. The suggested self-inhibitory annealing method can be used to improve the segmentation performance, especially for complex real-world scenes, and is fully incorporated into the linear threshold type dynamical model. The gain in grouping quality, however, is paid by a tradeoff in convergence speed.

The presented perceptual grouping applications highlight the strengths and weaknesses of the competitive layer model approach compared to other neural binding models. The CLM grouping approach is general enough to be adaptable to different visual feature domains, such as orientation, motion, intensity, and texture. Its possibility of combined figure-ground segmentation and grouping is very effective in noisy and cluttered scenes. The best results are achieved, if the lateral interaction ranges are not too small compared to the expected group sizes. A problem is to establish long-range coherence with very short-ranged interactions like for the example of greyscale segmentation, where the segment size is much larger than the local interaction ranges. Although the proposed annealing method solves this problem for intermediate interaction ranges, the approach becomes infeasible for very long ranges due to slow convergence. Other binding models, which perform operations similar to region growing seem more suitable for tasks of this type. On the contrary the CLM can obtain robust segmentations of multiple contours in complex real-world images which has yet to be demonstrated by synchronization-based binding models. Another aspect is that from a cognitive modeling point of view, a perfect segmentation performance may also not be adequate as a model of human perception. At least for the example of texture grouping, the imperfectness of the CLM texture grouping model seems to nicely meet the corresponding human texture discrimination capabilities.

The conceptual approach to segmentation which was explored in this study is based on two components. First, suitable image features must be chosen, which carry meaningful local image information, and second, an appropriate lateral interaction must be determined which expresses the mutual compatibility of the features. From a statistical point of view this can be interpreted as basing the segmentation on the second order statistics of the derived features in the data. Due to the recurrent nature of the binding process, however, also higher-order relations can be captured in the grouping model. Since all the interactions that were considered for perceptual grouping in this thesis were translation and rotation-invariant, the grouping approach can be interpreted as being based on searching for groups with high self-similarity.

The determination of appropriate lateral interactions is the crucial factor in neural binding models with pairwise feature coupling schemes. Although a number of perceptual grouping laws were successfully modelled by manually designed interactions, the formulation of a feasible learning approach seems to be an even more important step towards a successful application of the binding methodology in real-world image processing problems. For the example of cell segmentation, the results show, that the complex design process can be to a large extent replaced by supervised learning from a set of training examples. To achieve good generalization, the approach uses a decomposition principle of basis function function templates, which keeps the number of parameters to be adapted during learning in a moderate range.

Another learning method was investigated for the solution of constraint satisfaction problems, which also play an important role in many perceptual processes. The proposed backtracking deterministic annealing algorithm can be interpreted as a learning method to adapt the constraint-implementing weights in the lateral and interlateral interaction connection structure of the competitive layer model. The learning process is running online and implements a search strategy which is similar to classical backtracking algorithms. Due to the promising results for the tiling problems that were considered, the application of the algorithm to other more realistic large-scale problems would be an interesting topic for future work.

The principle of spatial feature binding provides an interesting candidate mechanism for binding in the brain, if we consider the wide range of perceptual grouping phenomena that were modelled by using the competitive layer model. Not much experimental evidence is available for the principle of cortical processing in competitive layers, as was reviewed in Section 5.6.2. Nevertheless, since not much is known about the highly complex neural circuitry in the visual cortex, it remains an interesting hypothesis from a functional modeling point of view. Much of the criticism on the feature binding concept has addressed the plausibility of the temporal correlation hypothesis on physiological grounds of the proposed synchronization mechanisms. The presented spatial binding approach by topographic segregation within layers offers an alternative principle to represent binding which could assist or complement the temporal scheme. While most of the experimental research on binding has centered around the temporal correlation hypothesis, the search for anatomical structures in the brain related to spatial feature binding may provide a new interesting experimental question.

The results presented in this thesis suggest a number of interesting future research directions. One possibility is to refine the modeling approach by a combination of the principle of spatial feature binding with the mechanisms of temporal correlation. The process of spatial segregation of neural activity in different groups could help to increase the robustness and coherence of neurons engaged in synchronous oscillation. On the other hand the studies on relaxation oscillators have shown, that they support rapid long-range coherence which is difficult to achieve for the competitive layer model.

With regard to the application of Gestalt-based perceptual grouping, the view could be broadened from the implementation of single grouping laws towards a combined dynamical selection among multiple grouping cues. In the Gestalt terminology this is called the principle of *Prägnanz* — of all the different grouping laws those will dominate the perception of a visual scene, which allow the simplest interpretation of its components. Of course the main question is the precise meaning of the term “simple interpretation” which can be considered as a major research field in the area of computer vision and machine learning. A possible approach for the competitive layer model could be to implement a similar competitive dynamics between different grouping laws and visual feature domains. Nevertheless, to achieve an adequate comparison between the different grouping modalities would pose a challenging task.

Perhaps the most important issue for future research is the further investigation of learning methods for binding. In the presented application examples only features of low complexity were used, such as local edge orientation and position. It seems very promising to extend these simple features by other local image characteristics, like the Gabor feature vectors considered by Ontrup & Ritter (1998) for texture segmentation. The determination of appropriate lateral interactions in such complex feature spaces should be the premier application domain of automated learning algorithms, where a manual design process is no longer feasible.

Appendix A

Angular Lateral Interactions

A.1 Continuity

The co-circular interaction of two edges, characterized by parameter vectors $\mathbf{m}_1 = (\mathbf{p}_1, \mathbf{n}_1)$, $\mathbf{m}_2 = (\mathbf{p}_2, \mathbf{n}_2)$, at positions $\mathbf{p}_1 = (p_1^x, p_1^y)$ and $\mathbf{p}_2 = (p_2^x, p_2^y)$ with a difference vector $\mathbf{d} = \mathbf{p}_1 - \mathbf{p}_2$, distance $d = ||\mathbf{d}||$ and $\hat{\mathbf{d}} = \mathbf{d}/d$, and unit orientation vectors $\mathbf{n}_1 = (n_1^x, n_1^y)$, $\mathbf{n}_2 = (n_2^x, n_2^y)$ is given by

$$f^{\text{cocirc}}(\mathbf{m}_1, \mathbf{m}_2) = \theta(a_1 a_2 q)\left(e^{-d^2/R^2 - C^2 S}\right) - I e^{-2d^2/R^2}, \tag{A.1}$$

where $a_1 = n_1^x \hat{d}^y - n_1^y \hat{d}^x,\ a_2 = n_2^x \hat{d}^y - n_2^y \hat{d}^x,\ q = \mathbf{n}_1 \cdot \mathbf{n}_2$ and

$$\theta(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases} \tag{A.2}$$

is necessary to exclude skewed symmetric edges. The parameter R controls the spatial range, which is smaller for the inhibitory component. The degree of co-circularity is given by

$$C = |\mathbf{n}_1 \cdot \hat{\mathbf{d}}| - |\mathbf{n}_2 \cdot \hat{\mathbf{d}}| \tag{A.3}$$

which is equal to zero if both edges lie tangentially to a common circle. The parameter $S > 0$ controls the sharpness of the co-circularity constraint and $I > 0$ controls the strength of the local inhibition. For all the simulations in this thesis, if not explicitly mentioned, the spatial dimensions were scaled into a unit square, and the parameters were $R = 0.1$, $S = 300$, $I = 0.5$.

A.2 Convex Continuity for Cell Segmentation

The convex/continuous lateral interaction for cell segmentation for two edges with parameters $\mathbf{m}_1 = (\mathbf{p}_1, \mathbf{n}_1), \mathbf{m}_2 = (\mathbf{p}_2, \mathbf{n}_2)$ (see above) at positions $\mathbf{p}_1$ and $\mathbf{p}_2$ is computed from their separation $\mathbf{d} = \mathbf{p}_1 - \mathbf{p}_2$ and their unit edge orientation vectors $\mathbf{n}_1 = (n_1^x, n_1^y)$, $\mathbf{n}_2 = (n_2^x, n_2^y)$. The normal vectors pointing into the intensity gradient direction are given by $\bar{\mathbf{n}}_1 = (-n_1^y, n_1^x)$ and $\bar{\mathbf{n}}_2 = (-n_2^y, n_2^x)$. The lateral interaction is then chosen as

$$f^{\text{cell}}(\mathbf{m}_1, \mathbf{m}_2) = \begin{cases} \xi\left(\left(\frac{d\pi}{2R}\right)^2 + \psi(\mathbf{n}_1, \mathbf{n}_2)^2\right) & \text{if } d < R,\ \bar{\mathbf{n}}_1\hat{\mathbf{d}} < s \text{ and } \bar{\mathbf{n}}_2\hat{\mathbf{d}} < -s \\ -I_l & \text{if } d < R,\ \bar{\mathbf{n}}_1\hat{\mathbf{d}} \geq s \text{ or } \bar{\mathbf{n}}_2\hat{\mathbf{d}} \geq -s \\ -I_g & \text{if } d \geq R \end{cases}$$

The function ξ is given by

$$\xi(x) = \begin{cases} cos(x) & \text{if } cos(x) > S \\ 0 & \text{if } cos(x) \leq S \end{cases} \tag{A.4}$$

and restricts the positive interactions, where $S = -0.7$. The parameter R is a rough estimate of the expected cell diameter. The function $\psi(\mathbf{n}_1, \mathbf{n}_2) = \frac{\pi}{4}(1 - \mathbf{n}_1 \cdot \mathbf{n}_2)$ measures the difference of the feature's gradient orientations. The parameter s controls the strength of the convexity constraint and $I_g > 0$ and $I_l > 0$ give the global and local inhibition strength, respectively. The lateral interaction parameters were $s = 0.2, R = 9, I_l = 0.9, I_g = 0.5$.

A.3 Interaction Templates for Learning

A.3.1 Interaction 1

The first interaction template is constructed by taking the convex/continuity interaction for cell segmentation and breaking it into three components, which make up the interaction basis templates. The components are given by the background layer interaction g^1, the local orientation-selective field g^2, and the non-local orientation-insensitive global inhibitory field g^3. This amounts to the following choice of three basis functions:

$$g^1_{\boldsymbol{\nu}} = g^{1\alpha}_{rr'} = \delta_{\alpha 1}\delta_{rr'} \tag{A.5}$$

$$g^2_{\boldsymbol{\nu}} = g^{2\alpha}_{rr'} = \begin{cases} f^{\text{cell}}(\mathbf{m}_r, \mathbf{m}'_r) & \text{if} \quad \alpha > 1 \text{ and } d < R \\ 0 & \text{else} \end{cases}$$

$$g^3_{\boldsymbol{\nu}} = g^{3\alpha}_{rr'} = \begin{cases} 1 & \text{if} \quad \alpha > 1 \text{ and } d > R \\ 0 & \text{else} \end{cases}$$

The parameters of f^{cell} are identical to those described in the previous section.

A.3.2 Interaction 2

The second interaction is a general translation and rotation invariant symmetric function in the feature parameters. The interaction is distance-dependent only in the sense that for a feature distance larger than R the global inhibition is active, while for $d < R$ the interaction depends on the relative orientation of the features. For two features with parameter vectors $\mathbf{m}_1 = (\mathbf{p}_1, \mathbf{n}_1)$ and $\mathbf{m}_2 = (\mathbf{p}_2, \mathbf{n}_2)$ the angles $\theta_1(\mathbf{d}, \mathbf{n}_1)$ and $\theta_2(\mathbf{d}, \mathbf{n}_2)$ relative to the difference vector $\mathbf{d} = \mathbf{p}_2 - \mathbf{p}_1$ are computed by taking the angle starting from the difference vector into the mathematically positive direction (see Figure A.1). We assume that $\theta \in$

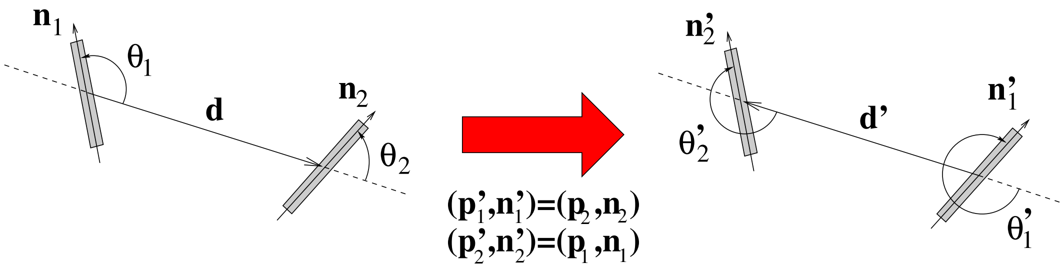


Figure A.1: Relative orientations and symmetry under feature exchange.

$[0, 2\pi[$. These relative angles are invariant under common rotation of the feature pair. The interactions are given by

$$
\begin{aligned}
g^1_{\boldsymbol{\nu}} &= g^{1\alpha}_{rr'} = \delta_{\alpha 1}\delta_{rr'} \\
g^2_{\boldsymbol{\nu}} &= g^{2\alpha}_{rr'} = \begin{cases} 1 & \text{if} \quad \alpha > 1 \text{ and } d > R \\ 0 & \text{else} \end{cases} \\
g^{j\geq 3}_{\boldsymbol{\nu}} &= g^{j\alpha}_{rr'} = \begin{cases} 1 & \text{if} \quad \alpha > 1, d < R \text{ and } j = 2 + q(\theta_1, \theta_2) \\ 0 & \text{else} \end{cases}
\end{aligned}
\tag{A.6}
$$

The function $q(\theta_1, \theta_2)$ evaluates the combination of the two angles by discretizing them to a number of orientation intervals by a function $s(\theta) = \text{int}(S\theta/(2\pi)) \in \{0, \dots, S-1\}$, where $\text{int}(x)$ is the largest integer number n for which $n < x$. The complete function must be symmetric under feature exchange, that is $g^j(\mathbf{m}_r, \mathbf{m}'_r) = g^j(\mathbf{m}'_r, \mathbf{m}_r)$. As is illustrated in Figure A.1 under an exchange of features $\mathbf{m}'_1 = \mathbf{m}_2$ and $\mathbf{m}'_2 = \mathbf{m}_1$ the corresponding angles are changed by $\theta'_1 = \theta_2 + \pi$ and $\theta'_2 = \theta'_1 + \pi$. The symmetry relation is encoded into a lookup table $Q_{ss'}, s = 1, \dots, S$ which carries index of the independent parameters. Hence

$$
q(\theta_1, \theta_2) = Q_{s(\theta_1)s(\theta_2)}. \tag{A.7}
$$

The angles are quantized into $S = 8$ intervals. Due to the symmetry condition this results in 36 free parameters and the corresponding lookup table matrix is given by

$$
Q = \begin{pmatrix}
1 & 2 & 3 & 4 & 27 & 28 & 29 & 30 \\
5 & 6 & 7 & 8 & 28 & 31 & 32 & 33 \\
9 & 10 & 11 & 12 & 29 & 32 & 34 & 35 \\
13 & 14 & 15 & 16 & 30 & 33 & 35 & 36 \\
17 & 18 & 19 & 20 & 1 & 5 & 9 & 13 \\
18 & 21 & 22 & 23 & 2 & 6 & 10 & 14 \\
19 & 22 & 24 & 25 & 3 & 7 & 11 & 15 \\
20 & 23 & 25 & 26 & 4 & 8 & 12 & 16
\end{pmatrix}. \tag{A.8}
$$

Lateral Interaction 3

The third lateral interaction template is identical to the previous one regarding the background layer, global inhibition and the angular component with its lookup table representation. The only difference is a further distance-dependent subdivision in the local neighbourhood within range R. As a consequence, the whole superimposed interaction has 74 free parameters and is given by

$$g^{1}_{\boldsymbol{\nu}} = g^{1\alpha}_{rr'} = \delta_{\alpha 1}\delta_{rr'} \tag{A.9}$$

$$g^{2}_{\boldsymbol{\nu}} = g^{2\alpha}_{rr'} = \begin{cases} 1 & \text{if} \quad \alpha > 1 \text{ and } d > R \\ 0 & \text{else} \end{cases}$$

$$g^{3\leq j\leq 38}_{\boldsymbol{\nu}} = g^{j\alpha}_{rr'} = \begin{cases} 1 & \text{if} \quad \alpha > 1, d < \frac{R}{2} \text{ and } j = 2 + q(\theta_1, \theta_2) \\ 0 & \text{else} \end{cases}$$

$$g^{39\leq j\leq 74}_{\boldsymbol{\nu}} = g^{j\alpha}_{rr'} = \begin{cases} 1 & \text{if} \quad \alpha > 1, \frac{R}{2} < d < R \text{ and } j = 2 + q(\theta_1, \theta_2) \\ 0 & \text{else} \end{cases} \tag{A.10}$$

Appendix B

Backtracking Deterministic Annealing Algorithm

The BDA algorithm is based on the simulation algorithm stated in Section 4.3.1. The global temperature T is replaced by a set of local temperatures T_r, which control the softness of the assignment for each column r separately.

1. Set $T_r(0) = T_c$ for all r, where $T_c = \lambda_{max}(\{\mathsf{f}_{rr'}^{\alpha\beta} - J\delta_{rr'}\})$.
   Initialise all $x_{r\alpha}$ with random values

   $$x_{r\alpha}(t = 0) \in [1/L - \epsilon, 1/L + \epsilon].$$

2. Do $N \cdot L$ times:

   (a) Choose (r, α) randomly and update $x_{r\alpha}(t{+}1) = \max(0, \xi)$, where

   $$\xi = \frac{J\Big(1 - \sum_{\beta \neq \alpha} x_{r\beta}(t)\Big) + \sum_{r'\beta} \mathsf{f}_{rr'}^{\alpha\beta} x_{r'\beta}(t) - \mathsf{f}_{rr}^{\alpha\alpha} x_{r\alpha}(t)}{J - \mathsf{f}_{rr}^{\alpha\alpha} + T_r}.$$

   (b) If $T_r < T^{(1)}$ and $x_{r\alpha} > \theta$, then do for all r', β:
   If $\mathsf{f}_{rr'}^{\alpha\beta} < 0$ and $x_{r'\beta} > \theta$ then set $\mathsf{f}_{rr'}^{\alpha\beta}(t+1) = \mathsf{f}_{rr'}^{\alpha\beta}(t) - \Delta$ and reset $T_r(t+1) = T_{r'}(t+1) = T^{(2)}$.

3. Set $T_r(t{+}1) = (1 - \eta)T_r(t)$ for all r with $0 < \eta < 1$. Go to step 2 until convergence.

In step 2 (b) the BDA weight modification is done if the local temperature is below the threshold $T^{(1)}$ and the two activities are greater than a threshold parameter $0 < \theta < 1$ such that an almost unambiguous assignment has already occurred for the two features.

Bibliography

Adams, R. & Bischof, L. (1994). Seeded region growing. IEEE Trans. Pattern Analysis and Machine Intelligence, 16, 641–647.

Adorjan, P., Levitt, J. B., Lund, J. S., & Obermayer, K. (1999). A model for the intracortical origin of orientation preference and tuning in macaque striate cortex. Visual Neuroscience, 16, 303–318.

Anderson, J. A. (1983). Cognitive and psychological computation with neural models. IEEE Trans. Systems, Man, Cybernetics 13(5), 799–815.

Baldi, P. & Meir, R. (1990). Computing with arrays of coupled oscillators: An application to preattentive texture discrimination. Neural Computation 2(4), 458–471.

Barlow, H. B. (1972). Single units and cognition: A neuron doctrine for perceptual psychology. Perception, 1, 371–394.

Barlow, H. B. (1985). The twelfth Bartlett memorial lecture: the role of single neurons in the psychology of perception. Quart. J. Exp. Psychol., 37, 121–145.

Bauer, U., Scholz, M., Levitt, J. B., Obermayer, K., & Lund, J. S. (1999). A biologically-based neural network model for geniculocortical information transfer in the primate visual system. Vision Research, 39, 613–629.

Ben-Yishai, R., Lev Bar-Or, R., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. Proc. Nat. Acad. Sci. USA, 92, 3844–3848.

Bergen, J. R. & Adelson, E. H. (1988). Early vision and texture perception. Nature, 333, 363–364.

Bergen, J. R. & Julesz, B. (1983). Rapid discrimination of visual patterns. IEEE Transactions on Systems, Man, and Cybernetics, 13, 857–863.

Berman, A. & Plemmons, R. J. (1979). Nonnegative matrices in the mathematical sciences. Academic Press.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. J. Roy. Statist. Soc. Ser. B, 36, 192–236.

Besag, J., Green, P., Higdon, D., & Mengersen, K. (1995). Bayesian computation and stochastic systems. Statistical Science, 1, 3–66.

Biederman, I. (1987). Recognition by components - a theory of human image understanding. Psychological Review, 94, 115–147.

Bitner, J. R. & Reingold, E. M. (1975). Backtrack programming techniques. Comm. ACM, 651–656.

Black, M. J. & Anandan, P. (1996). The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. Computer Vision and Image Understanding 63(1), 7075–7104.

Blasdel, G. G. & Fitzpatrick, D. (1984). Physiological organization of layer 4 in macaque striate cortex. J. Neurosci., 4(3), 880–895.

Blatt, M., Wiseman, S., & Domany, E. (1997). Data clustering using a model granular magnet. Neural Computation 9(8), 1805–1842.

Brodatz, P. (1966). Texture: A Photographic Album for Artists and Designers. New York: Dover.

C. Aykanat, T. B. & Haritaoglu, I. (1998). A fast neural network algorithm for VLSI cell placement. Neur. Netw. 11(9), 1671–1684.

Chang, M. M., Tekalp, A. M., & Sezan, M. I. (1997). Simultaneous motion estimation and segmentation. IEEE Trans. Image Processing 6(9), 1326–1333.

Chou, P. B. & Brown, C. M. (1990). The theory and practice of bayesian image labeling. International Journal of Computer Vision, 4, 185–210.

Cohen, M. A. & Grossberg, S. (1983). Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. IEEE Trans. Systems, Man, Cybernetics, 13, 815–826.

DeAngelis, G. C., Freeman, R. D., & Ohazawa, I. (1994). Length and width tuning of nerons in the cat's primary visual cortex. Neurophysiology, 71, 347–374.

Douglas, R., Koch, C., Mahowald, M., Martin, K., & Suarez, H. (1995). Recurrent excitation in neocortical circuits. Science, 269, 981–985.

Douglas, R., Mahowald, M., & Martin, K. (1994). Hybrid analog-digital architectures for neuromorphic systems. In IEEE International Conference on Neural Networks, Volume 3, Orlando, pp. 1848–1853.

Eckes, C. & Vorbrüggen, J. C. (1996). Combining data-driven and model-based cues for segmentation of video sequences. In Proceedings WCNN96, pp. 868–875. INNS Press & Lawrence Erlbaum Ass.

Elder, J. H. & Zucker, S. W. (1993). The effect of contour closure on the rapid discrimination of two-dimensional shapes. Vision Research, 33, 981–991.

Engel, A. K., König, P., & Singer, W. (1991). Direct physiological evidence for scene segmentation by temporal coding. Proc. Nat. Acad. Sci. USA, 88, 9136–9140.

Eremin, I. I. (1962). Iteration method for Chebyshev approximations for sets of incompatible linear inequalities. Soviet Math. Doklady, 3, 570–572.

Fellenz, W. (1997). Ein neuromorphes System für die datengetriebene Szenenanalyse. Ph. D. thesis, Universit"at Paderborn.

Fellenz, W. A. & Hartmann, G. (1995). Image segmentation by phase label diffusion. In Proc. Int. Conf. Artificial Neural Networks ICANN, Paris, pp. 309–314.

Feng, J. (1997). Lyapunov functions for neural nets with nondifferentiable input-output characteristics. Neural Computation 9(1), 43–49.

Feng, J. & Hadeler, K. P. (1996). Qualitative behaviour of some simple networks. Journal of Physics A, 29, 5019–5033.

Feng, J., Pan, H., & Roychowdhury, V. P. (1996). On neurodynamics with limiter function and Linsker's development model. Neural Computation, 8, 1003–1019.

Field, D. J., Hayes, A., & Hess, R. F. (1992). Contour integration by the human visual system: Evidence for a local "association field". Vision Research 33(2), 173–193.

Galbraith, W., Wagner, M., Chao, J., Abaza, M., Ernst, L., Nederlof, M., Hartsock, R., Taylor, D., & Waggoner, A. (1991). Imaging cytometry by multiparameter fluorescence. Cytometry, 12, 579–596.

Geiger, A. & Yuille, A. (1991). A common framework for image segmentation. International Journal of Computer Vision 6(3), 227–243.

Geman, D., Geman, S., Graffigne, C., & Dong, P. (1990). Boundary detection by constrained optimization. IEEE Trans. Pattern Analysis and Machine Intelligence 12(7), 609–628.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Analysis and Machine Intelligence, 6, 721–741.

Gerstner, W., van Hemmen, J. L., & Cowan, J. D. (1996). What matters in neuronal locking? Neural Computation 8(8), 1653–1676.

Ghose, G. M. & Maunsell, J. (1999). Specialized representations in visual cortex: A role for binding? Neuron, 24, 79–85.

Gilbert, C. D. (1992). Horizontal integration and cortical dynamics. Neuron 9(1), 1–13.

Gislén, L., Peterson, C., & Söderberg, B. (1992). Rotor neurons: Basic formalism and dynamics. Neural Computation 4(5), 737–745.

Golomb, S. W. (1994). *Polyominoes: puzzles, patterns, problems and packings* (2nd ed.). Princeton Univ. Press.

Golub, G. H. & van Loan, C. (1989). Matrix Computations. John Hopkins Press.

Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration: Still alive and well. Neuron, 24, 31–47.

Hahnloser, R. L. T. (1998). On the piecewise analysis of linear threshold neurons. Neur. Netw., 11, 691–697.

Hanahara, K. & Hiyane, M. (1990). A circle-detection algorithm simulating wave propagation. Machine Vision and Applications, 3, 97–111.

Hartline, H. K. & Ratliff, F. (1958). Spatial summation of inhibitory influence in the eye of limulus and the mutual interaction of receptor units. Journal of General Physiology, 41, 1049–1066.

Hawken, M. J. & Parker, A. J. (1984). Contrast sensitivity and orientation selectivity in lamina iv of the striate cortex of old world monkeys. Exp. Brain Res., 54, 367–372.

Herault, L. & Horaud, R. (1993). Figure-ground discrimination: A combinatorial optimization approach. IEEE Trans. Pattern Analysis and Machine Intelligence 15(9), 899–914.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). An Introduction to the Theory of Neural Computation. Lecture Notes Volume I. Addison Wesley.

Hirsch, M. W. (1989). Convergent activation dynamics in continuous time networks. Neur. Netw., 2, 331–349.

Hofmann, T., Puzicha, J., & Buhmann, J. (1998). Unsupervised texture segmentation in a deterministic annealing framework. IEEE Trans. Pattern Analysis and Machine Intelligence 20(8), 803–818.

Holt, G. R., Softky, W. R., Koch, C., & Douglas, R. J. (1996). A comparison of discharge variability *in vitro* and *in vivo* in cat visual cortex neurons. Journal of Neurophysiology, 75, 1806–1814.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proc. Nat. Acad. Sci. USA, 79, 2554–2558.

Hopfield, J. J. (1984). Neurons with graded responses have collective computational properties like those of two-state neurons. Proc. Nat. Acad. Sci. USA, 81, 3088–3092.

Hopfield, J. J. & Tank, D. D. (1985). Neural computation of decisions in optimization problems. Biol. Cyb., 52, 141–152.

Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction adn functional architecture in cat's visual cortex. J. Physiol., 160, 106–154.

Hubel, D. H. & Wiesel, T. N. (1972). Laminar and columnar distribution of geniculocortical fibers in macaque monkey. J. Comp. Neurol., 146, 421–450.

Hummel, R. A. & Zucker, S. W. (1983). On the foundations of relaxation labeling processes. IEEE Trans. Pattern Analysis and Machine Intelligence 5(3), 267–286.

Jacobs, D. W. (1996). Robust and efficient detection of salient convex groups. IEEE Trans. on Pattern Analysis and Machine Intelligence 18(1), 23–37.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. Nature, 290, 91–97.

Julesz, B. (1986). Texton gradients: the texton theory revisited. Biological Cybernetics, 54, 245–251.

Kandel, E. R., Schwartz, J. H., & Jessel, T. M. (1991). *Principles of Neural Science* (3 ed.). Prentice-Hall International Inc.

Kapadia, M. K., Ito, M., & Westheimer, G. (1995). Improvement of visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. Neuron, 15, 843–856.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimisation by simulated annealing. Science, 220, 671–680.

Kittler, J. & Illingworth, J. (1985). Relaxation labeling algorithms - a review. Image and Vision Computing 3(4), 206–216.

Koffka, K. (1935). Principles of Gestalt Psychology. New York: Harcourt Brace.

Köhler, W. (1929). Gestalt Psychology. New York: Liveright.

Kosowsky, J. J. & Yuille, A. L. (1994). The invisible hand algorithm: Solving the assignment problem with statistical physics. Neur. Netw. 7(3), 477–490.

Kovács, I. & Julesz, B. (1993). A closed curve is much more than an incomplete one: effect of closure in figure-ground segmentation. Proc. Nat. Acad. Sci. USA, 90, 7495–7497.

Kreiter, A. K. & Singer, W. (1996). Stimulus-dependent synchronization of neuronal responses in the visual cortex of the awake macaque monkey. Journal of Neuroscience, 16, 2381–2396.

Krishnamachari, S. & Chellappa, R. (1996). Delineating buildings by grouping lines with MRFs. Trans. on Image Processing, 5, 164–168.

Kuhn, D. & Loehwen, R. (1987). Piecewise affine bijections of R^n and the equation $Sx_+ - Tx_- = y$. Linear Algebra and its Appl., 96, 109–129.

Li, S. Z. (1995). Markov Random Field Modeling in Computer Vision. Springer.

Li, S. Z., Soh, W. Y. C., & Teoh, E. K. (1998). Relaxation labeling using augmented Lagrange-Hopfield method. Pattern Recognition 31(1), 73–81.

Li, Z. (1998). A neural model of contour integration in the primary visual cortex. Neural Computation, 10, 903–940.

Li, Z. & Dayan, P. (1999). Computational differences between asymmetrical and symmetrical networks. Network, 10, 59–77.

Linsker, R. (1992). Local synaptic learning rules suffice to maximize mutual information in a linear network. Neural Computation, 1, 402–11.

Lippmann, R. (1987). An introduction to computing with neural nets. IEEE ASSP Mag., 4, 4–22.

Littmann, E. & Ritter, H. (1997, January). Adaptive color segmentation—a comparison of neural and statistical methods. IEEE Transactions on Neural Networks 8(1), 175–185.

Liu, X. & Wang, D. L. (1999). Range image segmentation using a relaxation oscillator network. IEEE Trans. Neur. Netw. 10(3), 564.

Logothetis, N. K. & Pauls, J. (1995). Psychophysical and physiological evidence for viewer-centered object representations in the primate. Cerebral Cortex, 5, 270–288.

Lund, J. S. (1987). Local circuit neurons of macaque monkey striate cortex: I. neurons of laminae 4c and 5a. Journal of Comparative Neurology, 257, 60–92.

Massmann, A., Posch, S., Sagerer, G., & Schlüter, D. (1997). Using Markov random fields for contour-based grouping. In International Conference on Image Processing, Volume II, Santa Barbara, pp. 207–210. IEEE.

McCafferty, J. D. (1990). Human and Machine Vision. Ellis Horwood.

Meyer, F. & Beucher, S. (1990). Morphological segmentation. J. Visual Communications and Image Representation 1(1), 21–46.

Milner, P. (1974). A model for visual shape recognition. Psychological Review, 81, 521–535.

Minsky, M. & Papert, S. (1969). Perceptrons. Cambridge, MA: MIT Press.

Mozer, M., Zemel, R. S., Behrmann, M., & Williams, C. K. I. (1992). Learning to segment images using dynamic feature binding. Neural Computation 4(5), 650–665.

Mundel, T., Dimitrov, A., & Cowan, J. D. (1997). Visual cortex circuitry and orientation tuning. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), Advances in Neural Information Processing Systems, Volume 9, pp. 887. The MIT Press.

Murino, V., Regazzoni, C. S., & Foresti, G. L. (1996). Grouping as a searching process for minimum-energy configurations of labelled random fields. Computer Vision and Image Understanding 64(1), 157–174.

Nattkemper, T. W., Ritter, H., & Schubert, W. (1999). Extracting patterns of lymphocyte fluorescence from digital microscope images. In Intelligent Data Analysis in Medicine and Pharmacology 99 (IDAMAP), Workshop Notes, pp. 79–88.

Nattkemper, T. W., Wersing, H., Schubert, W., & Ritter, H. (2000a). Fluorescence micrograph segmentation by gestalt-based feature binding. In Proc. of the Int. Joint Conf. on Neur. Netw. (IJCNN). submitted.

Nattkemper, T. W., Wersing, H., Schubert, W., & Ritter, H. (2000b). A neural network architecture for automatic segmentation of fluorescence micrographs. In Proc. of the 8th Europ. Symp. on Art. Neur. Netw. (ESANN). accepted.

Neisser, U. (1963). Decision-time without reaction-time: experiments in visual scanning. American Journal of Psychology, 76, 376–385.

Neisser, U. (1967). Cognitive Psychology. New York: Appleton-Century-Crofts.

Ontrup, J. & Ritter, H. (1998). Perceptual grouping in a neural model: Reproducing human texture perception. Technical Report SFB 360 98-6, University of Bielefeld.

Opara, R. & Wörgötter, F. (1998). A fast and robust cluster update algorithm for image segmentaion in spin-lattice models without annealing - visual latencies revisited. Neural Computation 10(6), 1547–1566.

Pal, N. R. & Pal, S. K. (1993). A review on image segmentation techniques. Pattern Recognition 26(9), 1277–1294.

Parent, P. & Zucker, S. W. (1989). Trace inference, curvature consistency, and curve detection. IEEE Trans. Pattern Analysis and Machine Intelligence 46(17), 763–770.

Park, J., Cho, H., & Park, D. (1999). On the design of BSB associative memories using semidefinite programming. Neural Computation, 11, 1985–1994.

Peleg, S. & Rosenfeld, A. (1978). Determining compatibility coefficients for curve enhancement relaxation processes. IEEE Trans. Systems, Man, Cybernetics 8(7), 548–555.

Pelillo, M., Abbatista, F., & Maffione, A. (1995). An evolutionary approach to training relaxation labeling processes. Pattern Recognition Letters, 16, 1069–1078.

Pelillo, M. & Fanelli, A. M. (1997). Autoassociative learning in relaxation labeling networks. Pattern Recognition Letters, 18, 3–12.

Pelillo, M. & Refice, M. (1994). Learning compatibility coefficients for relaxation labeling processes. IEEE Trans. Pattern Analysis and Machine Intelligence 16(9), 933–945.

Perfetti, R. (1995, September). A synthesis procedure for Brain-State-in-a-Box neural networks. IEEE Transactions on Neural Networks 6(5), 1071–1080.

Peterson, C. & Soderberg, B. (1989). A new method for mapping optimization problems onto neural networks. Int. J. Neural Systems 1(1), 3–22.

Polat, U. & Sagi, D. (1994). The architecture of perceptual spatial interactions. Vision Research, 34, 73–78.

Pomerantz, J. (1981). Perceptual organization in information processing. In Perceptual Organization. Hillsadle, NJ: Lawrence Erlbaum Associates.

Posch, S. & Schlüter, D. (1998). Perceptual grouping using Markov random fields and cue integration of contour and region information. Technical Report SFB 360 98-10, University of Bielefeld.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). Numerical Recipes in C. Cambridge University Press.

Reynolds, J. H. & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. Neuron, 24, 19–29.

Riesenhuber, M. & Poggio, T. (1999). Are cortical models really bound by the "binding problem"? Neuron, 24, 87–93.

Ritter, H. (1990). A spatial approach to feature linking. In Proc. Int. Neur. Netw. Conf. Paris Vol.2, pp. 898–901.

Ritter, H., Martinetz, T., & Schulten, K. (1992). Neural Computation and Self-Organizing Maps: An Introduction. Reading, MA: Addison-Wesley.

Ritz, R., Gerstner, W., Fuentes, U., & van Hemmen, J. L. (1994). A biological motivated and analytically soluble model of collective oscillations in the cortex. II. Application to binding and pattern segmentation. Biol. Cyb., 71, 349–358.

Rock, I. & Palmer, S. (1990, December). The legacy of Gestalt psychology. Scientific American, 84–90.

Rockland, K. S. & Lund, J. S. (1983). Intrinsic laminar lattice connections in primate visual cortex. J. Comp. Neurol., 216, 303–318.

Rose, K. (1998). Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. Proc. IEEE 86(11), 2210–2239.

Rosen, J. B. (1960). The gradient projection method for linear programming – Part I: Linear constraints. J. Soc. Industr. Appl. Math.

Rosenblatt, F. (1962). Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Washington DC: Spartan Books.

Rosenfeld, A., Hummel, R. A., & Zucker, S. W. (1976). Scene labeling by relaxation operations. IEEE Trans. Systems, Man, Cybernetics 6(6), 420–433.

Sajda, P. & Finkel, L. H. (1994). Dual mechanisms for neural binding and segmentation. In J. D. Cowan, G. Tesauro, & J. Alspector (Eds.), Advances in Neural Information Processing Systems, Volume 6, pp. 993–1000. Morgan Kaufmann Publishers, Inc.

Salinas, A. & Abbott, L. F. (1996). A model of multiplicative responses in parietal cortex. Proc. Nat. Acad. Sci. USA, 93, 11956–11961.

Schillen, T. B. & König, P. (1994). Binding by temporal structure in multiple feature domains of an oscillatory network. Biol. Cyb., 70, 397–405.

Shadlen, M. & Movshon, J. A. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis. Neuron, 24, 67–77.

Shareef, N., Wang, D. L., & Yagel, R. (1999). Segmentation of medical images using LEGION. IEEE Trans. Medical Imaging, 18, 74–91.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? Neuron, 24, 49–65.

Somers, D. & Kopell, N. (1995). Waves and synchrony in networks of oscillators of relaxation and non-relaxation type. Physica D, 89, 169–183.

Sompolinsky, H., Golomb, D., & Kleinfeld, D. (1991). Cooperative dynamics in visual processing. Physical Review A 43(12), 6990–7011.

Sonka, M., Hlavac, V., & Boyle, R. (1993). Image Processing, Analysis and Machine Vision. Chapman & Hall.

Steil, J. J. (1999). Input-Output Stability of Recurrent Neural Networks. Ph. D. thesis, University of Bielefeld, Faculty of Technology.

Takefuji, Y. & Lee, K.-C. (1990, March). A parallel algorithm for tiling problems. IEEE Trans. Neur. Netw. 1(1), 143–145.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. Science, 262, 685–688.

Tanaka, K. (1996). Inferotemperal cortex and object vision: stimulus selectivity and columnar organization. Annual Review of Neuroscience, 19, 109–139.

Terman, D. & Wang, D. L. (1995). Global competition and local cooperation in a network of coupled oscillators. Physica D, 81, 148–176.

Terzopoulos, D. (1986). Regularization of inverse visual problems involving discontinuities. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8, 413–424.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. Journal of Experimental Psychology: Human Perception and Performance 8(2), 194–214.

Usrey, W. M. & Reid, R. C. (1999). Synchronous activity in the visual system. Annual Review of Neuroscience, 61, 435–456.

Vapnik, V. (1995). The nature of statistical learning theory. New York: Springer.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. Kybernetik, 14, 85–100.

von der Malsburg, C. (1981). The correlation theory of brain function. Technical Report 81-2, MPI Göttingen.

von der Malsburg, C. (1999). The what and why of binding: The modeler's perspective. Neuron, 24, 95–104.

von der Malsburg, C. & Buhmann, J. (1992). Sensory segmentation with coupled oscillators. Biol. Cyb., 54, 29–40.

von der Malsburg, C. & Schneider, W. (1986). A neural cocktail-party processor. Biological Cybernetics, 54, 29–40.

Vorbrüggen, J. C. (1995). Zwei Modelle zur datengetriebenen Segmentierung visueller Daten, Volume 47 of Reihe Physik. Thun, Frankfurt am Main: Verlag Harri Deutsch.

Waltz, D. (1975). Understanding line drawings of scenes with shadows. In P. H. Winston (Ed.), The Psychology of Computer Vision. McGraw Hill.

Wang, D. & Terman, D. (1997). Image segmentation based on oscillatory correlation. Neural Computation 9(4), 805–836.

Wang, D. L. & Brown, G. J. (1999). Separation of speech from interfering sounds based on oscillatory correlation. IEEE Trans. Neur. Netw. 10(3), 684.

Wang, J. Y. A. & Adelson, E. H. (1994). Representing moving images with layers. IEEE Trans. Image Processing 3(5), 625–638.

Weliky, M., Kandler, K., Fitzpatrick, D., & Katz, L. C. (1995). Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to orientation columns. Neuron, 15, 541–552.

Wersing, H. (1996). A neurodynamical model of perceptive grouping. Diplom Thesis, Universität Bielefeld.

Wersing, H., Beyn, W.-J., & Ritter, H. (2000). Dynamical stability conditions for recurrent neural networks with unsaturating piecewise linear transfer functions. submitted to Neural Computation.

Wersing, H. & Ritter, H. (1999a). Backtracking deterministic annealing for constraint satisfaction problems. In Int. Conf. Art. Neur. Netw. ICANN, pp. 868–875.

Wersing, H. & Ritter, H. (1999b). Feature binding and relaxation labeling with the competitive layer model. In Proc. Eur. Symp. on Art. Neur. Netw. ESANN, pp. 295–300.

Wersing, H., Steil, J. J., & Ritter, H. (1997). A layered recurrent neural network for feature grouping. In W. Gerstner, A. Germond, M. Hasler, & J.-D. Nicoud (Eds.), Proc. ICANN, pp. 439–444.

Wersing, H., Steil, J. J., & Ritter, H. (2000). A competitive layer model for feature binding and sensory segmentation. Neural Computation (in press).

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. Psychologische Forschung, 4, 301–350.

Wertheimer, M. (1938). Principles of perceptual organisation. In W. H. Ellis (Ed.), Source Book of Gestalt Psychology. London.

Wilson, H. R. & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J., 12, 1–24.

Wu, F. Y. (1983). The Potts model. Rev. Mod. Physics, 54, 235.

Yabuta, N. H. & Callaway, E. M. (1998). Functional streams and local connections of layer 4c neurons in primary visual cortex of the macaque monkey. J. Neurosci., 18(20), 9489–9499.

Yamamoto, H., Ninomiya, H., & Asai, H. (1998). A neuro-based optimization algorithm for rectangular puzzles. IEICE Trans. Fundamentals. E 81-A(6), 1113–1117.

Yen, S.-C. & Finkel, L. (1996). Salient contour extraction by temporal binding in a cortically-based network. In M. Jordan, M. Mozer, & T. Petsche (Eds.), Advances in Neural Information Processing Systems 9, pp. 915–921. MIT Press.

Yen, S.-C. & Finkel, L. H. (1998). Extraction of perceptually salient contours by striate cortical networks. Vision Research 38(5), 719–741.

Yoshioka, T., Levitt, J. B., & Lund, J. S. (1994). Independence and merger of thalamocortical channels within macaque monkey primary visual cortex: anatomy of interlaminar projections. Vis. Neurosci., 11, 467–489.

Zemel, R. S., Williams, C. K. I., & Mozer, M. C. (1993). Directional-unit Boltzmann machines. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), Advances in Neural Information Processing Systems, Volume 5, pp. 172–179. Morgan Kaufmann, San Mateo, CA.

Zucker, S. W. (1976). Region growing: Childhood and adolescence. Computer Graphics and Image Processing 5(3), 382–399.

Zucker, S. W., David, C., Dobbins, A., & Iverson, L. (1988). The organization of curve detection: Coarse tangent fields and fine spline coverings. In 2nd Int. Conf. on Computer Vision, Washington, DC,, pp. 568–577. Comp. Soc. Press.

Zucker, S. W., Dobbins, A., & Iverson, L. (1989). Two stages of curve detection suggest two styles of visual computation. Neural Computation 1(1), 68–81.