

A Computational Feature Binding Model of Human Texture Perception

Jörg Ontrup and Helge Ritter
Neuroinformatics Group
Faculty of Technology
Bielefeld University
33501 Bielefeld, Germany

Heiko Wersing
HONDA Research Institute Europe GmbH
Carl-Legien-Straße 30
63073 Offenbach/Main, Germany

We present a computational model for human texture perception which assigns functional principles to the Gestalt laws of similarity and proximity. Motivated by early vision mechanisms, in a first step local texture features are extracted by utilizing multi-scale filtering and non-linear spatial pooling. In the second stage, features are grouped according to the spatial feature binding model of the Competitive Layer Model (CLM) (Wersing, Steil, & Ritter, 2001). The CLM uses cooperative and competitive interactions in a recurrent network, where binding is expressed by the layer-wise coactivation of feature-representing neurons. The Gestalt law of similarity is expressed by a non-euclidean distance measure in the abstract feature space with proximity being taken into account by a spatial component. To choose the stimulus dimensions which allow the most salient similarity-based texture segmentation, the feature similarity metrics is reduced to the directions of maximum variance. We show that our combined texture feature extraction and binding model performs segmentation in strong conformity with human perception. The examples range from classical microtextures and Brodatz textures to other classical Gestalt stimuli, which offer a new perspective on the role of texture for more abstract similarity grouping.

Keywords: Perceptual grouping Texture perception Segmentation Psychophysics

I. Introduction

Perhaps one of the most intriguing capabilities of the brain is its ability to organize the tremendous amount of information supplied by our senses in such a way that the world is not perceived as a chaotic stream of impressions, but as a well structured set of entities. A much discussed topic addressing this issue is the “binding problem”. A classic example for this was given by (Rosenblatt, 1962), in which four neurons individually encode the visual properties *square*, *triangle*, *bottom* and *top*. If more than one object is present, a binding mechanism is required which produces a coherent perception of the object’s feature conjunctions, i.e. answering the question “Is the feature *square* associated with the feature being in the *top* or *bottom* half of the image for a given object?” – The associated features must be correctly *bound* together in order to avoid so called “illusory conjunctions”. These have been reported by (Treisman & Schmidt, 1982) in an experiment where arrays of colored objects of different shapes have been briefly presented to subjects. During the experiment subjects reported seeing objects made up of the color from one array and the shape from a different array. Their findings indicate, that the brain might indeed have difficulties in correct bindings, which we might call a “binding problem”. For a more thorough discussion of the binding problem in human vision see the review of (Wolfe & Cave, 1999).

The question how the brain actually solves the binding problem is still not answered satisfactorily and is receiving continued controversial attention among theorists and experi-

mentalists. A large body of research has focused on temporally correlated neural activity, for which evidence has been established by numerous experiments. (Kreiter & Singer, 1996) showed that cells in the middle temporal (MT) area, associated with motion processing, show selective synchronization if stimulated by a single bar stimulus and no synchronization if stimulated by two coherently moving, but separate bars. For reviews discussing neurophysiological evidence of this “temporal correlation hypothesis” see e.g. (Usrey & Reid, 1999; Singer, 1999).

In case of human texture perception the binding of visual stimuli plays an important role. Experiments (Bergen & Adelson, 1988) have shown that humans are able to segregate certain types of texture stimuli (such as, e.g., depicted in Fig. 7 below) preattentively. Therefore, some mechanism must exist in our early vision system enabling us to perceive a coherent impression of the different texture regions – presumably by some kind of binding.

Much research on visual texture perception was pioneered by Bela Julesz and his colleagues. Julesz was one of the first who systematically investigated the abilities of humans to discriminate between different textures. He proposed, that texture discrimination could be explained in terms of first order differences between local features called *textons* (Julesz, 1981). However, Nothdurft has shown that the *texton*-theory is inconsistent with the segregation of certain patterns (Nothdurft, 1991). Many studies have demonstrated the importance of spatial frequency information in connection with texture segregation phenomena (Dunn, Higgins, & Wakeley,

1994; Graham, Beck, & Sutter, 1992; Rentschler, Hubner, & Caelli, 1988; Sutter, Sperling, & Chubb, 1995; Turner, 1986). Therefore, many researchers have proposed computational models based on a standard model which texture perception scientists “routinely pull out from their back pocket” (Chubb & Landy, 1991). Daugman has suggested that 2D Gabor filters (Daugman, 1985) provide a reasonable fit to the receptive field of simple cells (Jones & Palmer, 1987). Since then this scale- and orientation selective mechanism has frequently been used as a biologically motivated feature extraction process for computational models. Other common filters are based upon wavelet transforms, quadrature mirror filters, discrete cosine transforms, eigenfilters, linear predictors and optimized finite impulse response filters. Randen et al. have evaluated a remarkable range of different filter types and have compared their performance with two nonfiltering approaches (co-occurrence and autoregressive features) (Randen & Husøy, 1999). They conclude that no single approach did perform best for *all* of their test images and could therefore not select a clear “winner” of their study. We therefore believe that the precise shape of the filtering function is not a critical choice.

More decisive for the overall performance are the post processing steps and their integration into a computational model. In the image analysis literature there exist two different approaches to model segmentation problems: *edge based* and *region based* methods. “Back pocket models” based on the former were presented by (Fogel & Sagi, 1989; Malik & Perona, 1990; Landy & Bergen, 1991; Manjunath & Chellappa, 1993). The models of (Jain & Farrokhina, 1991; Grossberg & Williamson, 1999; Hofmann, Puzicha, & Buhmann, 1998) are using region based mechanisms to model texture segregation phenomena.

However, psychophysical studies have revealed that in human texture perception *both* mechanisms are present: Some experiments (Nothdurft, 1985) have demonstrated that for certain stimuli edge based mechanisms are utilized, whereas other studies (Wolfson & Landy, 1998) showed that region based mechanisms are necessary to explain human texture segregation.

All of the edge based models mentioned above work in a similar fashion: After computing local orientation or other feature maps an edge detection mechanism like gradient computation, a Sobel or a Laplacian operator is applied to detect texture boundaries. For region based models there exist several approaches: Jain and Farrokhina use a least square error clustering algorithm to partition a set of texture features (Jain & Farrokhina, 1991), Hofmann et al. a deterministic annealing framework (Hofmann et al., 1998). Both models achieve a good performance, but expose a common problem: The choice of the number of clusters has to be made a priori. The authors suggest a heuristic, where several runs with different choices of the initial number of clusters are used. For each run an energy criterion is used to determine the “true” number of clusters in hindsight. Although the texture segmentation performance of (Hofmann et al., 1998) is very good, this heuristic and the annealing framework itself are biologically not very plausible. A more promising ap-

proach is that from Grossberg and Williamson (Grossberg & Williamson, 1999). Their ARTEX model incorporates a self organizing network which gives remarkable classification results on multiple scale orientational contrast texture features. In this contribution we present another biologically plausible computational model, the Competitive Layer Model (CLM) which addresses the binding problem.

The key principle of the CLM is the encoding of feature bindings by the assignment of feature responses to separate populations of laterally interconnected and locally competitive neurons. The feature binding process is achieved by an energy minimization scheme and exhibits strong similarities to region based relaxation labeling models (Wersing et al., 2001). 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 neural layer itself. Section II introduces the architecture and main properties of the binding dynamics of the CLM. In section III we present the feature extraction mechanism based on multi-scale filtering. Post-processing steps include a non-linear transfer function, spatial pooling and multi-dimensional scaling. Section IV is devoted to the application of the CLM and describes how the neural network architecture can be employed for texture discrimination tasks. A wide variety of application examples is presented in section V. We not only show that patterns which were explicitly designed for region based mechanisms are well separated, but also reproduce phenomena which are usually only explained by edge based properties. A comparison to psychophysical data illustrates that the CLM discriminates different textures in a very similar way as humans and is therefore consistent with human texture perception also from a quantitative point of view. Furthermore we show that our system performs equally well as other state-of-the-art models if applied to standard benchmarks taken from the Brodatz texture album. Additionally a few examples are given which demonstrate, that the CLM also resembles human introspection on more abstract figures which are commonly used to exemplify some well known “Gestalt Laws”. Section VI closes with some concluding remarks.

II. The Competitive Layer Model

The Competitive Layer Model (CLM), introduced by (Ritter, 1990), is an approach to perceptual organization and feature binding that is based on two principles. First, it is assumed that an initial processing stage derives a finite set of features which represent some form of independent, usually local, relevant structure in the sensory input. Second, there is a measure of the mutual compatibility of features which can be interpreted as a functional implementation of the phenomenological Gestalt laws. The CLM uses these compatibilities to partition the input features into salient groups by the recurrent dynamics in a layered neural network with topographically structured competitive and cooperative interactions. In the following we give an overview of the model as presented by (Wersing et al., 2001).

Architecture

Suppose in a preprocessing step a set of image positions $r = (x, y) \in I$ a set of features has been extracted. These are represented as vectors $\mathbf{m}_r \in \mathbb{R}^n$ embedded in a possibly high-dimensional parameter space and can represent arbitrary modalities such as:

- **Spatial information:** A stimulus can simply be described by its position in the image. Therefore, $\mathbf{m}_r = (x, y)^T$, where x and y denote the x - and y -coordinate of the stimulus' position.

- **Texture information:** Surfaces can be characterized by their specific textual appearance. So, $\mathbf{m}_r = \mathbf{t}_{\text{text}}$, where \mathbf{t}_{text} is a suitable local texture description. Section III describes in detail how the features were generated for this contribution.

- **Edge Information:** A common feature for contour integration and grouping models are local edge elements, characterized by $\mathbf{m}_r = (x, y, \theta)^T$, where θ is the orientation and (x, y) is the position of the edge. (Nattkemper, Wersing, Schubert, & Ritter, 2000) used this approach to segment fluorescence microscope images with the CLM.

The above list is only a small selection of practicable features. The combination of different modalities into a new attribute is also feasible. After the features were extracted during a preprocessing stage, to each of them is assigned a scalar intensity value h_r , which can be interpreted as the significance of the detection of feature \mathbf{m}_r . In a neurally plausible setup we may consider a fixed set of topographically ordered features, where the neural activity of a feature-detecting cell is related to the presence of the feature characterized by a particular parameter set, like e.g. an edge at a particular orientation and a particular position. To reduce the complexity for the simulation, however, we assume that at one image position only one of the possibly many feature alternatives is active, e.g. the edge detector at that orientation which receives maximal input. We therefore omit all features with small intensities and choose a set with $h_r = 1$ for all features.

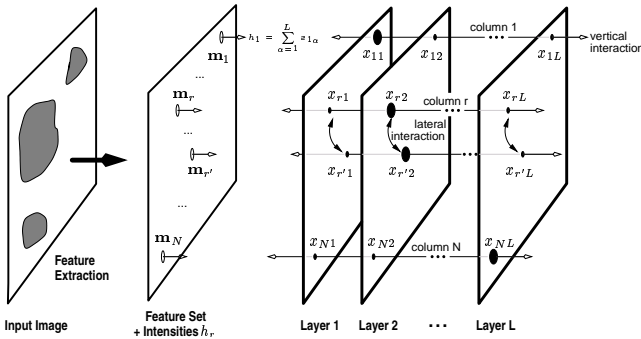


Figure 1. The architecture of the Competitive Layer Model is characterized by two types of interaction: Firstly, all neurons of a column r are *vertically* competing amongst all layers and secondly in every layer α each neuron $x_{r\alpha}$ is *laterally* interacting with all the other neurons in that layer to form the grouping process.

The CLM consists of L layers with index α , each containing N formal neurons with nonnegative activity $x_{r\alpha} \geq 0$. Furthermore, we call all neurons sharing the same position r

the *column* with index r . Therefore, the CLM has N columns with L neurons each (see Figure 1). The feature vectors \mathbf{m}_r are then assigned to their corresponding columns r . Two types of interaction are involved in the construction of the CLM: Firstly, a *vertical* interaction among the neurons of a column enforces approximately the *superposition condition* $\sum_{\alpha=1}^L x_{r\alpha} \approx h_r$. By meeting this constraint the pattern h_r of the input intensities is divided upon the L layers. Deviations from the exact input are used to represent contextual influences of laterally interaction features. Secondly, between neurons in the same layer there is a pairwise *lateral* interaction: Activities belonging to similar features – which therefore might belong to the same perceptual group – are “bound together” by a positive interaction. Activities belonging to diverse features are separated by a negative interaction.

For a set of N features r with parameters \mathbf{m}_r derived from a particular image the pairwise compatibilities can be stored into an $N \times N$ interaction matrix with elements $f(\mathbf{m}_r, \mathbf{m}_{r'}) = f_{rr'}$. To guarantee convergence of the model, the interaction must be symmetric, $f_{rr'} = f_{r'r}$ for all r, r' . Section IV elucidates the design of an interaction function for texture segmentation.

The dynamics of the CLM is given by

$$\dot{x}_{r\alpha} = -x_{r\alpha} + \sigma \left(J \left(h_r - \sum_{\beta} x_{r\beta} \right) + \sum_{r'} f_{rr'} x_{r'\alpha} \right). \quad (1)$$

where $\sigma(x) = \max(x, 0)$ is a linear-threshold transfer function. This corresponds to gradient descent in the following energy function:

$$E = \frac{J}{2} \sum_r \left(\sum_{\beta} x_{r\beta} - h_r \right)^2 - \frac{1}{2} \sum_{\alpha} \sum_{r, r'} f_{rr'} x_{r\alpha} x_{r'\alpha} \quad (2)$$

The first term of (2) is a constraint term corresponding to the superposition condition. The second term measures the sum of all grouping energies of each layer. The parameter J controls the coupling between the superposition constraint and the overall grouping energy.

(Wersing et al., 2001) have shown that

1. If the interaction matrix is symmetric ($f_{rr'} = f_{r'r}$), and additionally J is sufficiently large then the dynamics of the CLM is convergent. The critical value of J is given by $J > \max_r \sum_{r'} |f_{rr'}|$.

2. For positive self-coupling interaction functions ($\forall r : f_{rr} > 0$) stable attractors exhibit a “grouping assignment property”: Within each column r there is maximally one active neuron. Therefore, each feature \mathbf{m}_r is assigned to exactly one layer $\alpha(r) = \text{argmax}_{\alpha} (x_{r\alpha} > 0)$.

In other words, the CLM partitions the feature set \mathcal{M} into L disjoint subsets \mathcal{M}_{α} , $\alpha \in 1, \dots, L$. Note, that the number of layers does not need to correspond to the number of groups in the image, since for sufficiently many layers only those are active which carry a salient segment. This is a fundamental difference to other clustering approaches like k-means clustering. Although the energy (2) is conceptually similar to

other pairwise approaches (Hofmann, Puzicha, & Buhmann, 1996), the dynamical coupling to the input results in more flexible response properties, compared to other spin models and relaxation labeling models of segmentation (Wersing et al., 2001).

The CLM dynamics can be simulated by an iterative solution procedure which rapidly searches for fixed point attractors. This 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). It can be implemented in the following way

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

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

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

$$\xi = \frac{J(h_r - \sum_{\beta \neq \alpha} x_{r\beta}(t)) + \sum_{r' \neq r} f_{rr'} x_{r'\alpha}(t)}{J - f_{rr}}$$

3. Goto step 2 until convergence.

The single activity update in step 2 corresponds to solving the fixed point equation for activity $x_{r\alpha}$ with all other activities held constant. This asynchronous dynamics converges (Wersing et al., 2001) due to a convergence result on asynchronous iteration in neural networks by (Feng, 1997). For a more thorough discussion of the spatio-temporal dynamics involving an eigensubspace analysis of the approach towards the possible attractors of the network the reader is referred to (Wersing et al., 2001).

III. Feature extraction

In this section we describe the feature extraction stages we use to obtain the local texture features which will be subjected to the CLM feature binding model described in the previous section.

Multi-scale filtering

An ensemble of simple cells is best modelled as a family of self-similar 2D Gabor filters (Daugman, 1988)

$$g(x, y) = e^{-\left(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right)} e^{-ik(x-x_0)},$$

where (x_0, y_0) is the center of the receptive field in the spatial domain, σ_x and σ_y are the widths of the Gaussian envelope along the x and y axes, respectively, and $k = 2\pi/\lambda$ is the spatial frequency of a complex plane wave with wave normal along the x -axis and wavelength λ . We can construct a set of self-similar functions, commonly referred to as *Gabor wavelets* by scaling and rotation of the x - y coordinate frame:

$$\begin{aligned} g_{mn}(x, y) &= g(x', y'), \\ x' &= a^{-m}(x \cos \Theta_n + y \sin \Theta_n) \quad \text{and} \\ y' &= a^{-m}(-x \sin \Theta_n + y \cos \Theta_n), \end{aligned}$$

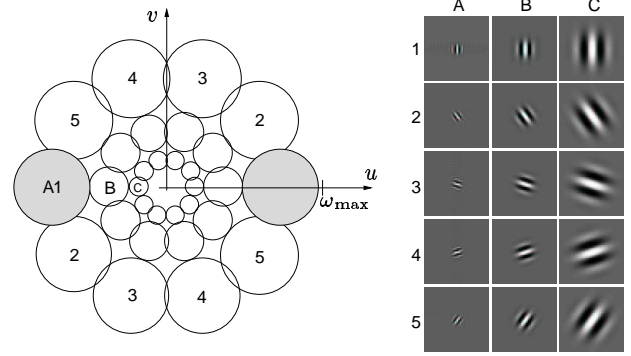


Figure 2. The set of 2D Gabor filters used for the feature extraction: On the left hand side the daisy-like pattern in the frequency domain u, v is shown. The circles denote the half-peak contours of each 2D Gabor. Their corresponding receptive fields in the spatial domain are depicted on the right.

(Watson, 1987) and (Lee, 1996) have demonstrated that a set of filters arranged in a daisy-like pattern as shown in Figure 2 is able to give a good description of an image coding scheme. It is “good” in such a sense that a sparse sampling of the phase space (which is spanned by m, n, x_0 and y_0) is sufficient for a complete representation of arbitrary image data.

According to physiological data (Valois, Yund, & Hessler, 1982; Valois, Albrecht, & Thorell, 1982) the median frequency and orientation bandwidths of simple cells are 1.4 octaves and 40 deg, respectively. By choosing a dyadic filter bank with 5 equally spaced orientations we obtain receptive fields with their optimal stimuli separated by 36 deg. A bandwidth of 1 octave allows an almost complete coverage of the frequency domain without substantial overlap (c.f. Figure 2). We therefore used a dyadic Gabor filter bank with 5 orientations and 3 scales to filter the input image. Since Gabor filters are complex valued it is a priori not clear whether to evaluate their even symmetric, odd symmetric or both parts. (Malik & Perona, 1990) have presented arguments that preattentive texture segregation is based on even symmetric mechanisms only. In addition to their arguments we present an example of an artificially created texture which produces a pop-out effect which can only be explained by the utilization of even symmetric mechanisms.

Nonlinearity

In Figure 3 a texture pair is shown which segregates preattentively, although the average response of a zero d.c. filter is 0 and the local variance is the same for both regions. This motivates to add some sort of nonlinearity to the channels. Commonly used choices are rectification (Bovik, Clark, & Geisler, 1990), energy computation (Landy & Bergen, 1991), intracortical inhibition (Malik & Perona, 1990) or a rectified sigmoid (Jain & Farrokhina, 1991). Albrecht and Hamilton describe a non-linear contrast response function for the majority of measured cells in striate cortex of monkey and cat (Albrecht & Hamilton, 1982). Therefore, a nonlinear function as the hyperbolic tangent seems to be an appropriate

choice, since it is not only biologically plausible, but also enables our model to differentiate between such textures.

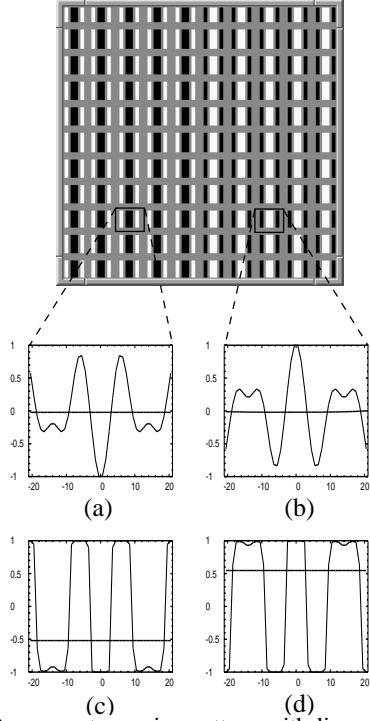


Figure 3. Response to a micropattern with linear and nonlinear spatial pooling. The pattern shown at the top segregates preattentively. **(a)** shows the crosssection along the x axis of the response to a micropattern of the left texture area. The straight line at zero corresponds to the mean value taken over a region about the size of the filter’s receptive field. **(b)** shows the same information for the right texture area. Note, that both the average response and the variance are the same for both regions, which prevents their segregation by filters that are based on these quantities only. In **(c)** and **(d)** an additional nonlinearity is applied. The nonlinear stretching of the curves changes the spatial average values of the signals which now differ between the two patterns.

Spatial pooling

Because texture is a property which is not connected to singular points in an image (like color is) but to a certain area, we need some sort of pooling process which takes an image region as an input and produces some sort of texture measure. Manjunath and Ma proposed a method based upon statistical properties of filter responses in order to gain a meaningful texture description (Manjunath & Ma, 1996). For each channel they compute the unnormalized mean and standard deviation over the whole image:

$$\mu_{mn} = \iint |c_{mn}(x,y)| dx dy$$

$$\sigma_{mn} = \sqrt{\iint (|c_{mn}(x,y)| - \mu_{mn})^2 dx dy},$$

where c_{mn} is the response in channel mn corresponding to scale m and orientation n . They use this feature vector to

compute the global difference of *whole* images. In order to detect local textual differences *within* a single image, a common approach is to divide the image into a set of small overlapping rectangular blocks centered on a regular grid (Hofmann et al., 1996). Then for each block the texture feature vector is computed and associated with the corresponding grid position. The more grid positions we use, the more accurate is the localization of texture borders. Maximum accuracy is achieved if the grid resolution is equal to the resolution of the digital image in pixels. Note, that in this case the computation of the mean value over a small block of size $M \times N$ is equivalent to a convolution of the image with an $M \times N$ filter kernel with entries $\frac{1}{MN}$. Such a convolution in turn corresponds to a smoothing of the image data. In digital image processing it is well known, that the convolution with rectangular filter masks only leads to suboptimal smoothing results (Jähne, 1993; Vernon, 1991). Much better results are obtained with Gaussian kernels. Therefore, we choose the following texture features:

$$\mu_{mn}(x,y) = c_{mn}^{\text{CTF}}(x,y) * g_{smn}(x,y) \quad (3)$$

$$\sigma_{mn}(x,y) = \sqrt{(c_{mn}^{\text{CTF}}(x,y) - \mu_{mn}(x,y))^2 * g_{smn}(x,y)}, \quad (4)$$

where $*$ denotes the convolution operation, c_{mn}^{CTF} is the response in channel mn after the nonlinear scaling with the contrast transfer function, and g_{smn} is the corresponding Gaussian filter kernel given by

$$g_{smn}(x,y) = e^{-\frac{x^2+y^2}{2\rho_{mn}^2}},$$

where ρ_{mn} is the width of the smoothing filter. The parameter ρ_{mn} is a critical choice since texture is a quality which cannot be associated with a single point, but an image region of a certain size. A reliable measurement of texture features calls for large sizes. On the other hand, an accurate localization of texture borders demands smaller sizes. In our experiments we found a heuristical value of three times the size of the receptive field’s Gaussian envelope to be a good choice.

Stages of feature processing

Summing everything up, we arrive at the following stages for the feature extraction process:

1. Compute a set of 2D Gabor filters tuned to 5 different orientations and 3 scales.
2. For each element of the filter bank, compute its response image, or channel, c_{mn} to the given input.
3. Apply the nonlinear contrast transfer function to each channel c_{mn} , yielding c_{mn}^{CTF} .
4. Compute the texture features μ_{mn} and σ_{mn} for each of the 15 channels according to (3) and (4).

Therefore, for each position (x,y) in the input image, we get a 30-dimensional feature vector $\mathbf{t}(x,y)$ describing the local texture at that point:

$$\mathbf{t}(x,y) = (\mu_{11}(x,y), \dots, \mu_{53}(x,y), \sigma_{11}(x,y), \dots, \sigma_{53}(x,y))^T \quad (5)$$

IV. The CLM Texture Segmentation Model

Multi-dimensional scaling

After the preprocessing we obtain a 30-dimensional description of the local image texture. A central question is, if we assume that in the brain a similar separation into many feature channels is performed, how these channels are combined to obtain a segmentation which has largest “Prägnanz” in the Gestalt sense — the partition which is most salient. For the CLM grouping approach this leads to the question to formulate a suitable feature compatibility function.

From a formal point of view, the 30-dimensional feature vector introduces a problem often called the “curse of dimensionality”. In context with classification tasks it refers to the difficulties associated with the exponential growth of hypervolume as a function of dimensionality. In a high dimensional space, data samples quickly become “lost” in the wealth of space. Ideally, only those texture channels should be used for the construction of the feature compatibility which carry the necessary information to segment different textures. It is not clear, however, which components have a high a priori discriminatory power for a given input image. On the other hand, if we take all vector elements into account for the calculation of the distance measure, then for each element with low discriminatory power, noisy information is added. This can severely reduce the segmentation performance - see also (Pichler, Teuner, & Hosticka, 1996) for a discussion on this topic.

For each distinct texture in the input image, we expect a 30-dimensional hyperellipsoid in our feature space. Consequently, the variation of the dataset should be high along those directions which connect the centers of hyperellipsoids corresponding to diverse texture regions, and it should be small along those directions which resemble similar textures according to our texture measure. A simple multi-dimensional scaling technique calculating the directions of maximal variance is principal component analysis (PCA) (Hancock, Baddeley, & Smith, 1992). By projecting the feature vectors \mathbf{t}_r onto their principal components we obtain a new set of feature vectors \mathbf{p}_r .

Feature subsampling

The complexity of the simulation of the CLM’s dynamics (1) is of the order $O(N^2)$, where N is the number of neuron sites r , which equals the number of feature vectors. For an input image of 256×256 pixels the grouping of those 65536 feature vectors would take several hours. Note, that the network type of the CLM could also be implemented in silicon in a biologically plausible way (Hahnloser, Sarpeshkar, Mahowald, Douglas, & Seung, 2000). For the sequential processing of the dynamics on a single processor, we subsampled the 4 feature images obtained by the projection onto the first 4 principal components. By dividing the 256×256 images into $1024 \ 8 \times 8$ regions we arrive at 1024 4-dimensional feature vectors \mathbf{p}_r which are used as inputs for the CLM yielding a tremendous speedup for (1) by a factor of 4096.

After the CLM has labelled this subsampled data set, we step through all original feature vectors \mathbf{p}_r and find its nearest neighbour \mathbf{p}_{near} such that $|\mathbf{p}_r - \mathbf{p}_{\text{near}}| + |\mathbf{r} - \mathbf{r}_{\text{near}}| = \min$, where \mathbf{r} and \mathbf{r}_{near} are the normalized positions in the images. Each \mathbf{p}_r is then labelled according to the label of its nearest neighbour \mathbf{p}_{near} . The usage of this technique allows the localization of texture borders with the same high resolution as in the original input images. Figure 4 depicts an example for the application of this technique.

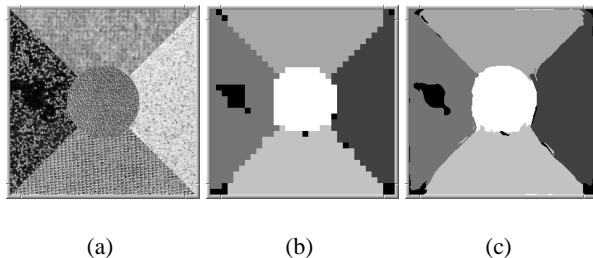


Figure 4. (b) shows the grouping result of the CLM to the input (a). The resolution enhancement (c) provides a more accurate localization of texture boundaries. Note that the black area in the left region does not depict a salient group. The activities in this area do not exceed the threshold of $1.2h_r$ (cf. Fig. 5).

Interaction function

Based upon a distance measure proposed by (Manjunath & Ma, 1996) we construct the following distance measure between two projected feature vectors \mathbf{p}_r and $\mathbf{p}_{r'}$:

$$d_{\text{text}}(r, r') = \sqrt[n]{\sum_{i=1}^4 \left(\frac{|p_r^i - p_{r'}^i|}{\sqrt{\sigma(p^i)}} \right)^n}, \quad (6)$$

where m_r^i is the i th component of vector \mathbf{m}_r , $\sigma(m^i)$ is the standard deviation of the i th component of all 1024 \mathbf{m}_r , and n is the dimension parameter of the Minkowski norm. In cases of small values for n , the Minkowski norm is less sensitive to differences in single channels. Following (Hofmann et al., 1996; Ma & Manjunath, 1996) we have chosen $n = 1$ which yielded good results in all of our experiments. The interaction function we propose for the CLM consists of two parts:

$$f_{rr'} = e^{-\frac{d_{\text{text}}^2(r, r')}{R_{\text{sim}}^2}} + c e^{-\frac{|\mathbf{x}_r - \mathbf{x}_{r'}|^2}{R_{\text{prox}}^2}} - k \quad (7)$$

Speaking in terms of Gestalt psychology, the first part can be identified with the Gestalt law of similarity, which states that similar features are bound together; the second part corresponds to the Gestalt law of proximity (stimuli are grouped together based on their physical proximity). The constant c controls the weighting between these principles. The summation of the two parts corresponds to a logical “or” which combines the two principles: $f_{rr'}$ is positive if the two image regions corresponding to r and r' have a similar textual appearance or if they are close together. The parameter $k > 0$

characterizes a global inhibitory influence for features which are both dissimilar and far apart. Note, that all simulation examples shown below were obtained with a constant set of the four parameters R_{sim} , c , R_{prox} , and k .

Interpreting the CLM's output

Once the dynamic of the CLM has reached an equilibrium point, we get for each feature vector \mathbf{m}_r the number $\alpha(r)$ of its active layer as an output from the CLM and use this to label (or colorize) the input image. According to (Wersing et al., 2001) for an equilibrium point of the dynamics either holds

$$x_{r\alpha} = h_r + \frac{1}{J_1} \sum_{r'} f_{rr'} x_{r'\alpha} \quad \text{or} \quad x_{r\alpha} = 0. \quad (8)$$

Therefore, $0 < x_{r\alpha} < h_r$ means, that there is a significant amount of activities in the layer α which belong to dissimilar feature vectors, because in that case the sum over all r' is negative. Consequently, only layers with activities greater than h_r correspond to clear and distinct perceptual groups. This motivates a heuristical threshold of $1.2h_r$. It proved to be a good value to discriminate between significant groupings and layers containing noisy and ambiguous information. The effects of this thresholding operation can be seen in Figure 5.

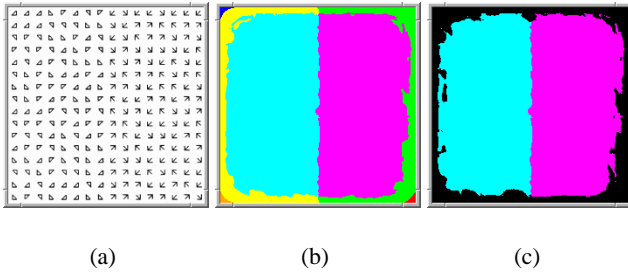


Figure 5. Effects of Thresholding: A CLM with 10 layers perceptually groups the image (a). (b) shows that actually 7 of the 10 layers are occupied with active neurons. By using only those layers for the labelling process which contain activities above the proposed threshold, we gain the result as shown in (c): Only the two distinct texture areas remain.

V. Benchmark images and results

So far we have motivated the feature extraction process, the construction of the interaction function for the CLM, and a method to obtain perceptually grouped images with the same resolution as the input images. Taking all these stages together we arrive at an artificial model of visual perception which describes the organization of visual stimuli according to the two Gestalt principles of Similarity and Proximity.

In this section we present a set of images the model is applied to. It can be divided into three parts:

1. Artificial textures were created by using different types of micropatterns. Following this way, we are able to construct well defined test images.

2. For the testing of real world textures we use a database of textured images containing samples from the popular Brodatz album (Brodatz, 1966). Since many texture segmentation algorithms are tested with these images, we are able to evaluate and compare the CLMs performance with other state of the art models.

3. Illustrations of well-known grouping phenomena were employed in order to show that our texture-based approach provides alternative explanations for classical visual stimuli for Gestalt perception of similarity and proximity.

Note, that all results presented in this section were obtained using the same set of parameters as described above. In certain cases we additionally apply the model with other parameters to demonstrate their effects on the grouping result. In all of these cases the change of parameters is explicitly described. Furthermore, if not otherwise noted, the CLM is applied with a constant number of 10 layers.

Artificial Textures

First, we show the perceptual organization the model produces if applied to artificial textures consisting of mirror symmetric micropatterns. (Malik & Perona, 1990) have presented arguments, that preattentive texture segregation is based on even symmetric mechanisms only. We discovered an interesting effect which strengthens their arguments: Consider Figure 6: At the top we can see two different test im-

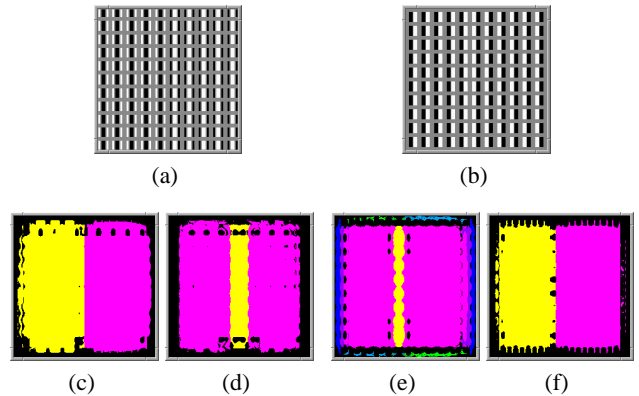


Figure 6. Grouping of textures consisting of mirror symmetric micropatterns. (c) and (e) are the grouping results for (a) and (b), respectively. In (d) and (f) the preprocessing pathway was modified as discussed in the text.

ages which consist of mirror symmetric micropatterns. For humans (a) segregates preattentively, (b) does not. (b) however, produces a “pop out” effect where the border region of the two areas is accentuated. This phenomenon is also reproduced by our model which can be seen in (e). The two grouping results shown in (d) and (f) were obtained with exactly the same processing pathway, except that the even symmetric 2D Gabors were exchanged by odd symmetric Gabors.

In this case, the results are reversed: The “pop out” effect occurs for the first pattern, and the second segregates. Since humans observe the “pop out” effect only for (b), this might be additional evidence, that in human texture perception indeed only even symmetric mechanisms are utilized. We are not aware that this kind of “pop out” effect was reported previously by other authors of texture segmentation models.



Figure 7. Grouping of textures consisting of L’s T’s and x’s: Similar to the untrained human observer the CLM only distinguishes between three different regions in (b). In (c) we decreased the parameter R_{sim} in (7), such the model “looks closer at texture differences” and perceptually organizes the image in four distinct groups.

The image shown in Figure 7(a) is composed of four different regions. Because orientation is an important criterion to detect similarity, the regions with upright T’s and L’s look similar. Consequently, the untrained observer usually sees only three different regions in this figure. Interestingly enough, the region with equally oriented x’s and L’s segregates well to the human observer. To get an idea of how our model segregates these two regions we plot the first four principal components of the texture feature vectors. As we see in Figure 8, mainly the differences of the feature vectors along the direction of the 3rd principal component are responsible for the segregation of the two regions. The corresponding linear combination of the 2D Gabor filters in the frequency domain is dominated by the component μ_{25} of the feature vectors $\mathbf{h}(x,y)$ as defined in (5).

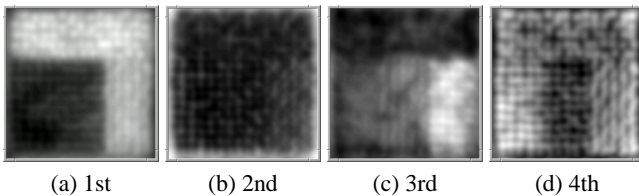


Figure 8. First four principal components of the texture features extracted from Figure 7(a). It can be seen that the 3rd p.c. is responsible for the segregation of the top and rightmost region in Figure 7(b).

The corresponding receptive field of the filter from which μ_{25} is extracted is sensitive to gratings orientated diagonally from the upper right to the lower left. If we project this receptive field on the input image, as indicated in Figure 9(a), we see that in the region with tilted L’s a large number of neurons with this receptive field have zero response. This is not the case for the region constructed of the x’s. If we try to position the fields between the x’s, there is always some part

of the pattern extending into the inhibitory region of the receptive field. Therefore, the average response in the x-region should be lower. This is indeed the case, if we inspect μ_{25} alone, as plotted in Figure 9(b).

Therefore, our model segments the two regions not only because of different responses to the micropatterns themselves, but also due to different responses to the *background* generated by the different patterns. One might speculate that this is also the case for the human observer.

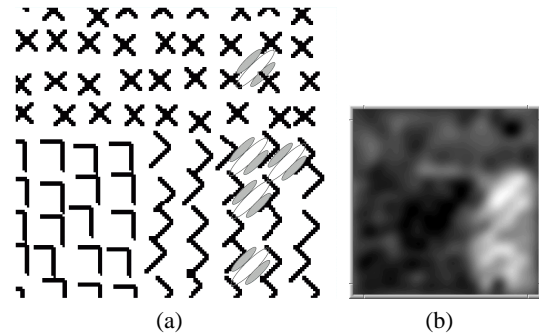


Figure 9. The sketch in (a) indicates, that there are a large number of neurons in the tilted L region whose receptive fields generate a zero response. This is not the case for the x region. In (b) the average response in the corresponding channel B5 is shown – see Figure 2 for the naming of the channels.

Edge and region based phenomena

Nothdurft studied human texture discrimination using patterns with different orientated line segments (Nothdurft, 1985). He systematically measured the influence of structure density on human segregation performance. According to his experiments texture discrimination depends not only on form, but also on the spacing of texture elements. Humans commonly fail to segment widely spaced texture elements, despite their instantaneous segregation in close arrangements. He suggests that texture segmentation is achieved by the evaluation of gradients and concluded that edge based mechanisms probably play a significant role for the detection of certain texture patterns. Some of his experiments and the output generated by our CLM approach are shown in Figure 10. For these experiments he created texture patterns with global figures consisting of small line elements in which line orientation differed from that in the surrounding texture field. The results (Nothdurft, 1985) showed that textures with strong differences in line orientation can be discriminated down to shorter line lengths than textures with smaller orientation difference.

Nothdurft presents another experiment which strengthens the assumption that edge based mechanisms are employed in human texture perception (Nothdurft, 1985). A texture pair is constructed analogous to the Craik-Cornsweet illusion of luminance perception. The illusion is caused by the fact that the human visual system has only limited sensitivity for absolute luminance levels. Subthreshold

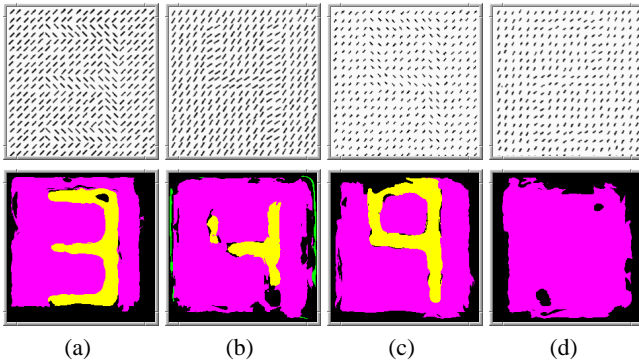


Figure 10. The patterns contain global figures in which line orientation differs from that in the surrounding texture field by 90° ((a),(c)) and 24° ((b),(d)). For each difference, two patterns with different line lengths are shown ((a),(b): 0.8 raster width, (c),(d): 0.5 raster width). The CLM grouping results are consistent with Nothdurft’s observations.

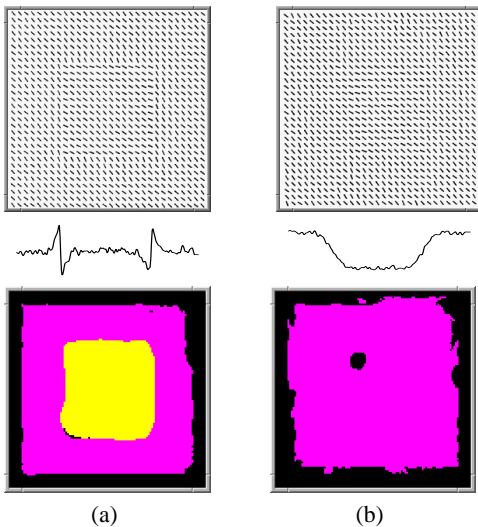


Figure 11. Craik-Cornsweet illusion for textured images (taken from [29]): In (a) lines are arranged to generate a maximal texture gradient, in (b) line orientation is varied in a smooth way (as indicated by the two graphs below the patterns mirroring the line orientations across the center). The human observer perceives a distinct and homogenous square in (a), and just a uniform background in (b).

variation of luminance may remain undetected and areas displaying identical luminance values on an absolute scale may appear different when the noticeable luminance contrast to neighbouring areas is different. The analogous textures corresponding the Craik-Cornsweet illusion are shown in Figure 11. In (a) the orientation of the line elements changes continuously from the periphery to the centre, with a sudden step in midway between center and perimeter: “As far as texture is concerned, the central square appears to be homogeneous and obviously distinct from the background,

even though lines in the centre of the square have the same orientation as lines at the pattern’s edges.” (Nothdurft, 1985). In (b) the global variation of texture differences is identical as in (a), but becomes imperceptible with respect to texture perception. As can be seen in Figure 11 the perceptual organization achieved by the CLM model is consistent with Nothdurft’s observations.

In addition to the experiments above, other studies were conducted which were specifically designed to demonstrate that also region based mechanisms are utilized in human texture perception (Landy & Bergen, 1991). In order to show that the CLM grouping behaviour also exhibits region based properties, we constructed a pattern according to (Landy & Bergen, 1991) as shown in Figure 12. The two texture regions consist of elements which have the same mean orientation but differ in their standard deviation. Additionally, the two regions do not share a common border such that a pure utilization of edge based mechanisms could not explain why humans are able to tell that two distinct patterns are visible in the image. In this case, the CLM model basically identifies three distinct groups: the background and the left and right half circle. As can be seen, the CLM model also mimics human perception in this case.

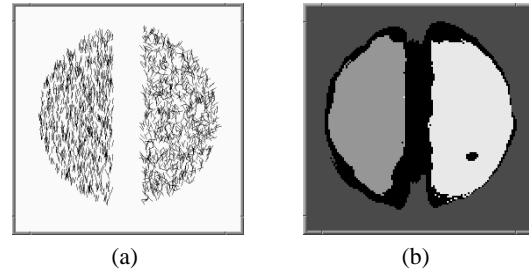


Figure 12. This result shows that the model is also able to detect different texture stimuli which were specifically designed for region based mechanisms - as described by (Landy & Bergen, 1991).

Why is the model able to distinguish between the two regions differing in their standard deviation of bar orientation? The CLM architecture binds stimuli together which share common features with respect to their interaction as defined by Eq. (7). By analyzing the eigenvectors of the texture features – akin to the procedure as shown in Fig. 8 – we found that not a feature related to the deviation measure as defined by Eq. (4) contributes most to the segregation, but mainly a component based on the mean responses in the low frequency channel C3 (as shown in Fig. 2). Therefore, the CLM models separates the two half circles mainly because a channel sensitive to *horizontal* gratings produces different responses to *the backgrounds* generated by the bars. An interesting experimental question would be if the suppression of certain orientation selective channels by GABA injections could switch on/off texture segregations like this.

Taking these results together we can summarize, that our texture segmentation model based on the CLM resembles hu-

man texture perception connected to edge based *and* region based phenomena.

Comparison to psychophysical data

In order to give a quantitative comparison of our model’s performance with human data, we have constructed a set of texture pairs according to (Kröse, 1987). The images used for the benchmark are shown in Figure 13. For each micropattern pair, 10 texture images with a size of 512x512 pixels were created. The elements were randomly oriented and positioned on a rectangular grid superimposed with a random jitter of 4 pixels. Then for each input image the CLM’s output was evaluated. The numbers given in Table 1 show the classification rates of the CLM model, which denotes the percentage of correctly classified pixels. Note, that a classification rate of 90% is almost perfect for these examples, since border effects cause a classification error of about 10% as can be seen in Figure 13(h). The rank order of discriminability for the CLM model matches the data from (Kröse, 1987) remarkably well and we therefore conclude that the CLM texture model also resembles human texture perception from a quantitative point of view.

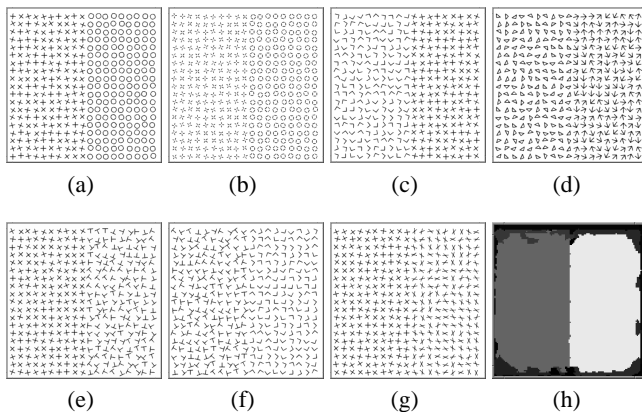


Figure 13. (a)–(g) depict the types of testimages used for the benchmark. (h) is an example for a typical grouping result for (b) with a classification error of approximately 86% mainly caused by border effects.

Natural Textures

In this section we present the application of the CLM to a set of images constructed of natural textures taken from the popular Brodatz album (Brodatz, 1966). This album contains photographs from natural textures, such as water, grass, leather, sand, bricks etc. Test images from this album are of common usage in the field of texture segmentation. We use images from a database taken from (Hofmann et al., 1996) which contains pictures each assembled from five different textures. We will present a few examples, which show the most important properties of our model if applied to this kind of natural images.

| Texture Pair | Psychophysical Data (Discriminability) | CLM Model (Classification rate) |
|--------------|--|---------------------------------|
| (a) | 100 | 89.1 |
| (b) | 88.1 | 89.3 |
| (c) | 68.6 | 82.1 |
| (d) | 52.3 | 79.7 |
| (e) | 37.6 | 74.7 |
| (f) | 30.6 | 69.7 |
| (g) | 30.3 | 69.0 |

Table 1

Comparison of psychophysical data from (Kröse, 1987) with CLM Model’s performance. The rank order of the classification rate of the CLM matches the rank order of the psychophysical data remarkably well.

The images presented in this section show, that the grouping results obtained with the CLM in connection with the proposed feature extraction mechanisms are generally in good accordance with human texture perception. In some cases the model has difficulties to detect the borders of different textures properly. These misclassifications are occurring in those examples, where different texture regions have a great similarity in visual appearance. Figure 4 shows that local differences within otherwise uniform textures are detected. Figure 14 presents some examples where the classification rate was not close to 100%. A visual inspection reveals that also the human observer has difficulties to detect the exact border of two textures if no a priori knowledge about the geometrical properties is assumed.

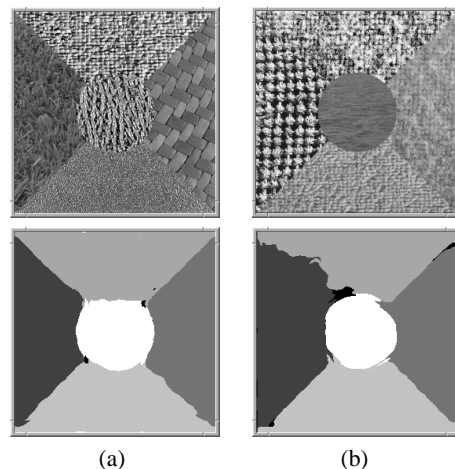


Figure 14. Two examples where the CLM achieved a classification rate not close to 100%. In (a) the center region is not well separated from the top, in (b) the top and left region seem to be interwaved. Note, that these “errors” resemble quite well the human perception of these texture examples.

Gestalt Laws

In this section we will present the grouping results of the CLM if applied to some of the images which illustrate the Gestalt laws proposed by the Gestalt psychologists. We have chosen these abstract figures to demonstrate that the feature set generated by the multi-scale filtering contains directions which might be interpreted as the largest “Prägnanz” in the Gestalt sense. The term “Prägnanz” can then be identified with the direction of maximal variance in the feature set.

Consider Figures 15: They show arrangements of dot patterns which exemplify the Laws of Proximity and Similarity. Although no explicit texture is contained in the images, the feature binding process of the CLM generates groups which are meaningful and correspond well to human introspection.

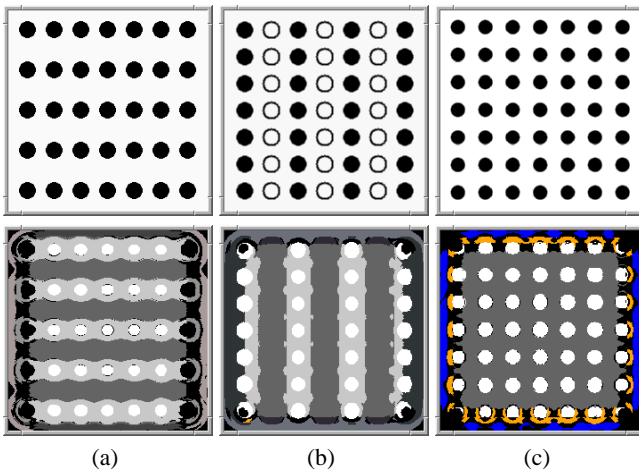


Figure 15. (a) shows an image which is often used to illustrate the Law of Proximity [36]: Because the horizontal spacing between the elements is smaller than the vertical spacing, we perceive five horizontal groups. The CLM grouping result below identifies three distinct groups: The first (white) corresponds to the elements themselves, the second (bright gray) connects the elements together, forming a horizontal structure. The third group (dark gray) expresses the horizontal structure generated from the background. Therefore, the perceptual grouping of this image mirrors the human introspection very well. (b) exemplifies the Law of Similarity: The spacing between elements is constant, but the elements themselves consist of different stimuli arranged in vertical groups. (c) shows equally spaced stimuli which do not differ in appearance. Consequently the human observer just perceives black dots on a uniform background - exactly as the CLM model as shown below.

Figure 16(a) shows an example where the CLM grouping result forms an illusory contour. Note, that the creation of such an illusory contour depends on the line spacing. If we increase the spacing as indicated in Figure 16(c), the grouping shows a complete different behaviour. In the latter case not the wavelike contour is predominant, but the background structure generated by the vertical lines. This example could inspire a psychophysical experiment which investigates the correlation between background line spacing and sensation of illusory contours. Results of such an experiment could be

used to fit the set of Gabor filters to the visual angle which is needed to perceive that phenomenon.

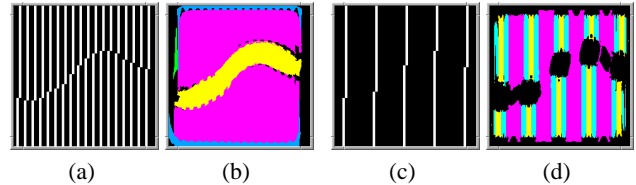


Figure 16. In (a) a pattern of broken lines is shown. Because the points marked by the discontinuities lie on a smooth line, we perceive a wavelike illusory contour. The grouping result obtained with the CLM also shows this wavelike contour in (b).

Time-varying Input

All examples presented above involved static stimuli, and it remains an interesting question, how the model can be extended to dynamic stimuli. If we consider a sequence of frames, where the grouping dynamics does not start from a fully initialized state, but from an already converged grouping state, we can expect an effect of hysteresis, that will lead to a perceptually plausible temporal integration. Only a major change in the visual input will induce a complete reorganization of the grouping. See also (Giese, 1999) for a comparison of similar perceptual hysteresis effects in related dynamical models of motion perception. The grouping of local optical flow features using the CLM has also been investigated in (Wersing, 2000).

VI. Discussion & Conclusion

In the following two sections we discuss the relation of the CLM to other binding models and shed some light on its biological relevance.

Relation to other Binding Models

Since synchronous activity in the visual cortex has been established by several experiments, a variety of synchronization-based modeling approaches have emerged. A large class of models uses local activity-based oscillatory circuits which are coupled by long-range horizontal interactions to facilitate the synchronization of the single oscillatory modules. 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. (Terman & Wang, 1995) proposed the LEGION model which uses relaxation oscillators to achieve fast synchronization and desynchronization with local excitatory and global inhibitory coupling schemes. This is very similar to the CLM grouping we are considering here, where the lateral interactions are based on a local similarity measure superimposed by a weak global inhibition. 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 winner-take-all columns.

Biological Relevance

The LEGION model has successfully been applied to a wide range of problems such as greyscale image segmentation, range image segmentation, auditory processing and texture segmentation (Cesmeli & Wang, 2001) as well. However, experiments indicate that texture segregation may not be represented by neuronal synchrony: (Lamme & Spekreijse, 1998) have presented various texture based figure-ground stimuli to monkey. Recordings of neural activity in the primary visual cortex (V1) showed “no systematic relationship between the synchrony of firing of pairs of neurons and the perceptual organization of the scene.” Therefore, they conclude that “synchrony in V1 does not represent the binding of local features into segregating textures, but is instead a reflection of the horizontal connections in V1.” Hence, for the case of texture segregation, the CLM architecture presents an alternative which in a biological way may appear more plausible than synchronization models.

The filtering stage of our architecture corresponds to responses of simple cells in V1. It is widely accepted (Valois & Valois, 1988) that these resemble Gabor-like receptive fields as proposed for the early processing stages of our model (c.f. Fig 2). The computation of statistical quantities during a pooling process as suggested by Eq. (3) and (4) requires neurons with larger receptive fields and might be carried out by cortical complex cells (Sakai & Tanaka, 2000). Neurophysiological experiments (Nothdurft, Gallant, & Van Essen, 2000) support the notion that responses in V1 can be strongly modulated by texture borders. Further experiments suggest that these response modulations in V1 do not depend on feedback connections from V2 (Hupe, James, Girard, & Bullier, 2001). These data suggest that local feature contrast in V1 is sufficient to allow for texture segregation. Based on recent findings where awake monkey reported figure-ground perception based on texture segregation (Super, van der Togt, Spekreijse, & Lamme, 2003) suggest “that perception depends strongly on recurrent interactions between visual areas [...]. This idea is supported by the finding that anesthesia has relatively little effect on feedforward responses in V1, [...] whereas figure-ground modulation is abolished by anesthesia”.

Considering these results the CLM model might well be biologically plausible. The CLM structure 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 synchronization-based approaches, which carry 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 therefore suggest the principle of topological segregation, as used in the CLM, as an additional binding principle which could improve the robustness of the feature binding processes.

Neurons of other areas could use different strategies to

read out a grouping result from the CLM architecture. One possibility as suggested by (Roelfsema, Lamme, Spekreijse, & Bosch, 2002) could employ a layer of nonmodulated cells. Perceptual groups can then be read out if neurons in other areas receive excitatory input from the CLM group layers and inhibitory input from the nonmodulated cells. An alternative could employ the combination of topological and temporal segregation by introducing layers of neural oscillators which receive input from the activation-based CLM layers.

Conclusion

We have presented a novel computational model for human texture perception. Based upon standard filter techniques which are commonly used in the texture segmentation literature we have developed a feature extraction system effectively producing a meaningful texture description. The usage of these features as inputs for the CLM have demonstrated that the model is able to reproduce edge based as well as region based phenomena which are both present in human texture perception. A quantitative analysis has shown that the model not only behaves qualitatively correct, but also closely resembles psychophysical data. Natural textures from the widely cited Brodatz album have been successfully segmented and proven that the CLM’s performance is comparable with other state-of-the-art texture segmentation approaches. The results obtained from the application to more conceptual figures exemplifying some well known “Gestalt Laws” offer a new perspective on the role of texture for more abstract saliency grouping. Altogether, the CLM was applied to a wide variety of different visual stimuli and produced strikingly similar results to human perception in all these cases. All grouping results were obtained with the same constant set of parameters, indicating that our model does not need a “parameter tuning process” many other models suffer from. Therefore, the Competitive Layer Model appears as a biologically plausible recurrent neural network implementing a feature binding process which can mimic a large variety of human perception phenomena.

References

- Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology*, 48(1), 217–237.
- Bergen, J., & Adelson, E. (1988). Early vision and texture perception. *Nature*, 333, 363–364.
- Besag, J., Green, P., Higdon, D., & Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science*, 1, 3–66.
- Bovik, A. C., Clark, M., & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1), 55–73.
- Brodatz, P. (1966). *Texture: A photographic album for artists and designers*. New York: Dover.
- Cesmeli, E., & Wang, D. (2001). Texture segmentation using gaussian-markov random fields and neural oscillator networks. *IEEE Transactions on Neural Networks*, 12, 394–404.
- Chubb, C., & Landy, M. S. (1991). Orthogonal distribution analysis: A new approach to the study of texture perception. In M. S.

- Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 291–301). MIT Press.
- Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2(7), 1160–1169.
- Daugman, J. G. (1988). Complete discrete 2D Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(7), 1169–1179.
- Dunn, D., Higgins, W. E., & Wakeley, J. (1994). Texture segmentation using 2D Gabor elementary functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2), 130–149.
- Feng, J. (1997). Lyapunov functions for neural nets with nondifferentiable input-output characteristics. *Neural Computation*, 9, 43–49.
- Fogel, I., & Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*, 61, 103–113.
- Giese, M. A. (1999). *Dynamic neural field theory for motion perception*. Kluwer Academic Publishers.
- Graham, N., Beck, J., & Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision Research*, 32(4), 719–743.
- Grossberg, S., & Williamson, J. R. (1999). A self-organizing neural system for learning to recognize textured scenes. *Vision Research*, 39(7), 1385–1406.
- Hahnloser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., & Seung, H. S. (2000). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405, 947–951.
- Hancock, P. J., Baddeley, R. J., & Smith, L. S. (1992). The principal components of natural images. *Network*, 3, 61–70.
- Hofmann, T., Puzicha, J., & Buhmann, J. (1996). *A deterministic annealing framework for unsupervised texture segmentation* (Tech. Rep. No. IAI-TR-96-2). University of Bonn.
- Hofmann, T., Puzicha, J., & Buhmann, J. M. (1998). Unsupervised texture segmentation in a deterministic annealing framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 803–818.
- Hupe, J., James, A., Girard, P., & Bullier, J. (2001). Response modulations by static textures surround in area V1 of the macaque monkey do not depend on feedback connections from V2. *Journal of Neurophysiology*, 85(1), 146–163.
- Jähne, B. (1993). *Digitale bildverarbeitung*. Berlin: Springer-Verlag.
- Jain, A. K., & Farrokhi, F. (1991). Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24(12), 1167–1186.
- Jones, J., & Palmer, L. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58, 1233–1258.
- Julesz, B. (1981). Textons, the elements of texture perception and their interaction. *Nature*, 290, 91–97.
- Kreiter, A., & Singer, W. (1996). Stimulus-dependent synchronization of neural responses in the visual cortex of the awake macaque monkey. *Journal of Neuroscience*, 16, 2381–2396.
- Kröse, B. (1987). Local structure analyzers as determinants of preattentive pattern discrimination. *Biological Cybernetics*, 55, 289–298.
- Lamme, V., & Spekreijse, H. (1998). Neuronal synchrony does not represent texture segregation. *Nature*, 396, 362–366.
- Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research*, 31(4), 679–691.
- Lee, T. S. (1996). Image representation using 2D Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10), 959–971.
- Ma, W., & Manjunath, B. (1996). Texture features and learning similarity. In *Proceedings of IEEE Int. Conference on Computer Vision and Pattern Recognition*.
- Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, 7(5).
- Manjunath, B., & Chellappa, R. (1993). A unified approach to boundary perception: Edges, textures, and illusory contours. *IEEE Transactions on Neural Networks*, 4(1), 96–107.
- Manjunath, B., & Ma, W. (1996). Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8), 837–842.
- Nattkemper, T. W., Wersing, H., Schubert, W., & Ritter, H. (2000). Fluorescence micrograph segmentation by gestalt-based feature binding. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)* (pp. 248–254).
- Nothdurft, H. (1985). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research*, 25(12), 1957–1968.
- Nothdurft, H. (1991). Different effects from spatial frequency masking in texture segregation and texture detection tasks. *Vision Research*, 31, 299–320.
- Nothdurft, H., Gallant, J., & Van Essen, D. (2000). Response profiles to texture border patterns in area V1. *Journal of Neurophysiology*, 16(3), 421–436.
- Pichler, O., Teuner, A., & Hosticka, B. J. (1996). A comparison of texture feature extraction using adaptive Gabor filtering, pyramidal and tree structured transforms. *Pattern Recognition*, 29(5), 733–742.
- Randen, T., & Husøy, J. H. (1999). Filtering for texture classification: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4), 291–310.
- Rentschler, I., Hubner, M., & Caelli, T. (1988). On the discrimination of compound Gabor signals and textures. *Vision Research*, 28, 279–291.
- Ritter, H. (1990). A spatial approach to feature linking. In *Proc. of the International Neural Network Conference* (Vol. 2, pp. 898–901). Paris.
- Robert, A. (1997). *From contour completion to image schemas: A modern perspective on Gestalt psychology* (Tech. Rep.). Department of Cognitive Science, University of California, San Diego.
- Roelfsema, P., Lamme, V., Spekreijse, H., & Bosch, H. (2002). Figure-ground segregation in a recurrent network architecture. *Journal of Cognitive Neuroscience*, 14, 525–537.
- Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptions and the theory of brain mechanics*. Washington DC: Spartan Books.
- Sakai, K., & Tanaka, S. (2000). Spatial pooling in the second-order spatial structure of cortical complex cells. *Vision Research*, 40, 855–871.
- Schillen, T., & König, P. (1994). Binding by temporal structure in multiple feature domains of an oscillatory network. *Biological Cybernetics*, 70, 397–405.

- Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24, 49–65.
- Super, H., van der Togt, C., Spekreijse, H., & Lamme, V. (2003). Internal state of monkey primary visual cortex(V1) predicts figure-ground perception. *The Journal of Neuroscience*, 23(8), 3407–3414.
- Sutter, A., Sperling, G., & Chubb, C. (1995). Measuring the spatial frequency selectivity of second-order texture mechanisms. *Vision Research*, 35(7), 915–924.
- Terman, D., & Wang, D. (1995). Global competition and local cooperation in a network of neural oscillators. *Physica D*, 81, 148–176.
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107–141.
- Turner, M. (1986). Texture discrimination by Gabor functions. *Biological Cybernetics*, 55, 71–82.
- Usrey, W., & Reid, R. (1999). Synchronous activity in the visual system. *Annual Review of Neuroscience*, 61, 194–214.
- Valois, R. L. D., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22, 545–559.
- Valois, R. L. D., & Valois, K. K. D. (1988). *Spatial vision*. New York: Oxford University Press.
- Valois, R. L. D., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, 22, 531–544.
- Vernon, D. (1991). *Machine vision*. New York: Prentice Hall.
- von der Malsburg, C., & Buhmann, J. (1992). Sensory segmentation with coupled oscillators. *Biological Cybernetics*, 54, 29–40.
- Watson, A. B. (1987). Efficiency of a model human image code. *Journal of the Optical Society of America A*, 4(12), 2401–2417.
- Wersing, H. (2000). *Spatial feature binding and learning in competitive neural layer architectures*. Goettingen: Cuvillier.
- Wersing, H., Steil, J. J., & Ritter, H. (2001). A competitive layer model for feature binding and sensory segmentation. *Neural Computation*, 13(2), 357–387.
- Wolfe, J., & Cave, K. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron*, 24, 11-17.
- Wolfson, S. S., & Landy, M. S. (1998). Examining edge- and region-based texture analysis mechanisms. *Vision Research*, 38(3), 439–446.