Relaxation Phase Labeling: Simulating Perceptual Organization

Winfried A. Fellenz
winfried@fellenz.de

1 Introduction

Visual perception of images is performed easily by the human brain requiring little conscious effort. However, imitating this ability using programmed computers has not been solved, largely due to the complexity of this seemingly simple task. One cause of the problems encountered may be the different processing architectures of serial computers used to emulate intrinsically parallel biological visual systems which also have been adapted by evolution processes to the behavioral environment. Furthermore the visual system of mammals is not completely prewired at birth, but develops during a postnatal period using random visual input and finally sharpens during sensorimotor interactions with the environment. This critical period usually spans only weeks after birth in kittens, but can last for years in humans possibly needing external intervention to establish coordinated stereo vision in young infants.

A clue to what constitutes the processing principles extracted from the environment was given in the phenomenological approach of the Gestalt-psychologists who tried to decipher the "laws" governing visual perception. They established a list of perceptual organization principles which group visual primitives into wholistic elements to be further processed by object recognition structures. The list includes grouping by proximity, similarity, good continuation, and Prägnanz.

To transfer the knowledge of biological visual processing, psychological grouping principles and plastic sensor adjustments into a programmed visual system serveral simplifications have to be made. First, the columnar arrangement of visually sensitive neurons in the visual cortex has to be analysed and transfered into a suitable structure to allow computational processing.

Further, the horizontal connectivity pattern between these feature sensitive neurons has to be encoded into the processing structure to allow the adaptively established connectivity to be used to process the sensory image of the neurons receptive fields.

Physiological examination of the properties of visual cortical neurons revealed certain specifities to special characteristics of the visual light: some neurons respond preferently to elongated bars of specific witdths and length, while others show preference to different spatial frequency and color. The connectivity between these charateristic neurons seem to involve patchy connections between similar orientations and may also include inhibitory activity. The overall processing in the prinary visual cortex appears to be the enhancing of similar properties and the inhibition of incompatible arrangements. Although a temporal analysis of feature specifity reveals a predominant enhancement of selectivity to the expense of broad tuning it is still unsolved how this process sharpens the response tuning of the neurons. We may here take the simplifying assumption that the responses of the neurons are spontaneous without further modifications of their direct response.

A further biological assumption is the temporal all-or-none response of visual neurons which can be observed in the spiking patterns following appropriate stimulation. It has been argued that synchronicity in these patterns may be used to bind features into wholistic objects in the brain to be processed as a single entity in the objet recognition pathway.

2 Perceptual organization

According to the time limitation of preattentive processing, only a few steps must be used by the visual system to segregate the scene into its parts. Due to the locality of this operation, global segmentation techniques can be ruled out being biologically implausible.

Psychophysical experiments concerning the perceptual strength of preattentively grouped parts showed an increased difficulty to direct attention to the individual parts of these groupings. The results suggest that the grouping stage may bind parts into wholes that are difficult or impossible to divide at later stages.

Gestalt psychologists assume that perceptual processes not merely record the visual stimulus but actively construct the percept. Their claim was that "the whole is more than the sum of its parts" giving emphasis to nonlinear interactions between the elementary parts of perception and the emergence of structure and configuration in the wholistic percept.

The Gestaltists have attempted to reveal the laws of perceptual organization underlying the partitioning decisions which group stimulus elements into perceptual wholes. They proposed the general principle of *Prägnanz* which means organization toward the most regular, ordered, stable and balanced state possible.

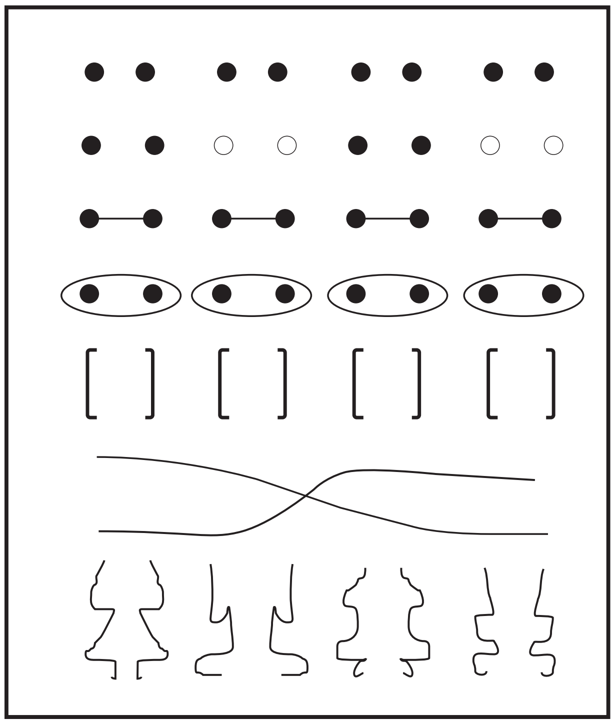

Figure 1: *The proposed Gestalt-laws: grouping by proximity, similarity, connectedness, common region, closure, good continuation and symmetry (top to bottom).*

This principle was accompanied by a collection of *Gestalt laws* depicted in figure 1 which include grouping by proximity, similarity, closure, connectedness, symmetry, simplicity and good continuation. Putting forward an analogy of brain states to the convergence of electric fields to a minimum energy state they proposed a mechanism for reaching a global consistent interpretation of a perceptual impression. The quintessence of the Gestalt argument is that perceptions are organized globally to simplify the representation of the stimulus.

3 The Hypercolumnar Architecture

The proposed perceptual grouping model consists of three successive processing stages which model mechanisms found in visual cortex. The first stage for the hierarchical extraction of contour lines has been adapted from the phase-dependent energy model using quadrature phase Gabor-Filters for the detection of local energy in the image.

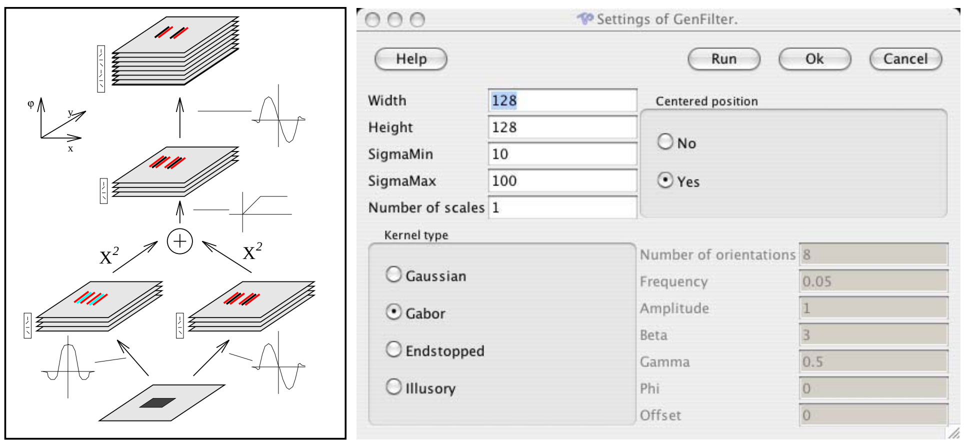

Figure 2: *left: A hypercolumn composed of direction specific ON and OFF channels generated from local energy filters specified as the sum of squared quadrature phase Gabor-filters. right: The RPL module GenFilter used to generate the eight odd symmetric Gabor filters.*

The responses of eight squared even- and odd-symmetric orientation channels are summed pairwise and thresholded to extract the local energy, followed by a differentiation step using odd symmetric Gabor filters to rectify the oriented responses into direction selective ON and OFF channels (figure 2) to build a dualistic representation of intensity changes in the image. The spatial Gabor-filters are defined as:

$$q_+(x,y) = (\alpha + \cos(\rho x)) e^{\left(-\frac{\beta x^2 + \gamma y^2}{\sigma^2}\right)} \quad (1)$$

$$q_-(x,y) = \sin(\rho x) e^{\left(-\frac{\beta x^2 + \gamma y^2}{\sigma^2}\right)} \quad (2)$$

Where $q_+(x,y)$ represents the even symmetric function and $q_-(x,y)$ its odd symmetric Hilbert-transform. The constants β and γ specify the envelope of the oriented Gaussian, ρ sets the appropriate frequency of the modulating sinusoidal, and the offset α is a normalization factor which is needed to allow the symmetric kernel to be zero integrable.

For the localized edge-detection eight filter pairs of size 64×64 were used with orientation difference of $22.5°$. The parameters were set to $\beta = 3$, $\gamma = 0.5$, $\sigma = 1$ and $\rho = 0.05$. The additional parameters for the even symmetric kernel were set to $\phi = \pi/2$ and $\alpha = 0.44$.

In the following processing step the oriented and localized energy responses are convolved using an odd-symmetric Gabor-filter and half-wave rectified into direction specific ON- and OFF-channels (figure 2). The parameters of the utilized Gabor-function were set to $\beta = 3$, $\gamma = 0.5$, $\sigma = 1$ and $\rho = 0.025$. The separation of the local energy follows biological coding principles, where ON- and OFF-channels are used to encode both positive and negative receptive field responses.

4 The Phase Relaxation Procedure

The processing scheme is as follows: smoothly varying constraints on the interaction strength between all direction selective responses of the second processing stage are defined. These constraints support orientation continuity by positive interactions between similar directions thereby implementing the Gestalt-law of good continuation, and decouple both sides of the contour by negative interactions between opposite directions.

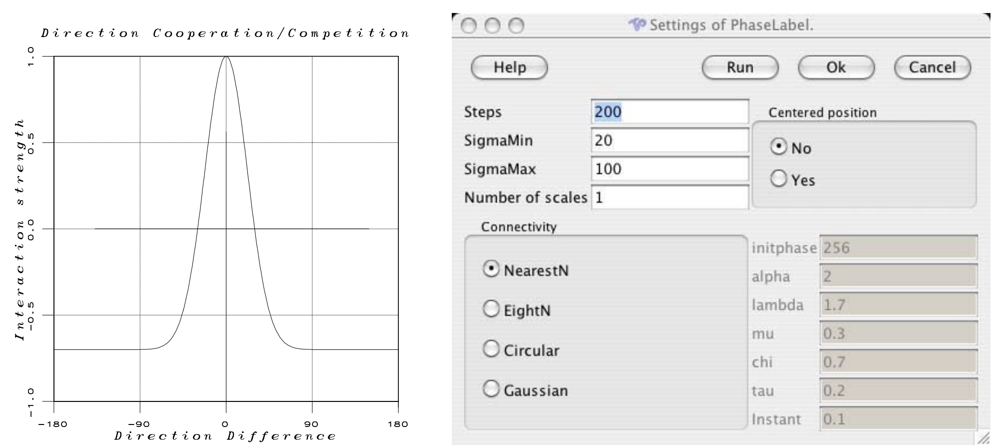

Figure 3: *left: The compatibility function between direction responses modeled as a shifted Gaussian. right: The RPL module PhaseLabel for computing the relaxation phase labeling in the hypercolumnar representation*

This compatibility function $v_{m,n}$ (figure 3), is modeled as a shifted Gaussian:

$$v_{m,n} = \lambda e^{(-\mu(||m-n||)^2)} - \chi \tag{3}$$

using the set of discrete directions $m, n \in M$ and the parameters $\lambda = 1.7$, $\mu = 0.3$ and $\chi = 0.7$. The compatibility function has been set according to analytical and experimental results but could in principle be defined using a learning procedure.

The periodic function $f(x)$ of phase difference $\delta\Phi$ between neighboring columns uses a shifted cosines:

$$f(\delta\Phi) = \frac{\delta\Phi}{\pi}(\beta + cos(\delta\Phi)) \tag{4}$$

$$\delta\Phi = \phi_{i+k,j+l} - \phi_{ij}; \tag{5}$$

$$-\pi <= \delta\Phi < \pi. \tag{6}$$

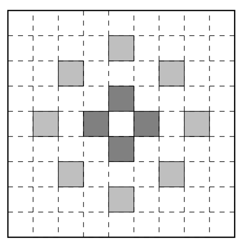

Figure 4: The sparse connectivity scheme $h_{k,l}$ for the propagation of local phase and edge responses between neighboring columns.

The phases ϕ_{ij} of each hypercolumnar vector at position (i, j) are updated according to a Gauß-Seidel procedure, using a sigmoid nonlinearity $g(x)$ for calculating the contributions of neighboring columns $Z_{i,j}(n)$ depending on their phase difference, and employing a sparse horizontal connectivity scheme $h_{k,l}$ (figure 4) to speed up the synchronization between columns:

$$\dot{\phi}_{i,j} = \tau_{i,j} + \sum_{m,n \in M} v_{m,n} E_{i,j}(m) Z_{i,j}(n) \tag{7}$$

$$Z_{i,j}(n) = g\Big(\alpha \sum_{k,l \in N} h_{k,l} E_{i+k,j+l}(n) f(\delta\Phi)\Big) \tag{8}$$

where $\tau_{i,j}$ is a zero mean random variable introducing noise into the phase process, thereby resolving ambiguous situations, and forcing the process to move from an initial equilibrium state with all phases being equal, to a global solution in phase space. $E_{i,j}(m)$ represents the activity in the m-th feature map and $Z_{i,j}(n)$ is the contribution to the n-th feature map from the surrounding hypercolumns using the connectivity matrix $h_{k,l}$. The sigmoid nonlinearity $g(x)$ has been set to $tanh(x)$.

The phases ϕ_{ij} of each hypercolumnar vector change according to the cooperative and competitive interactions defined by the compatibility constraints $v_{m,n}$ between elementary features.The constraints group similar directions using positive values but decouple strongly differing directions using negative values. To allow the spreading of phase labels into the interior regions with no edge response a small instantaneous response is added in a special map. This filling in is similar to brightness diffusion allowing the separation of figure and ground, but instead uses the coherency of the cyclic phases to label the whole scene into objects and background.

5 The RPL Workspace

The emergent forming of perceptual groups, including both contour and region based information is depicted in figure 5 showing the results of the proposed grouping scheme for a scene with three simple objects. Although the objects are defined by different boundary types ranging from intensity discontinuities over lines to dot patterns, the phase gradient shows a common interpretation of all contour types.

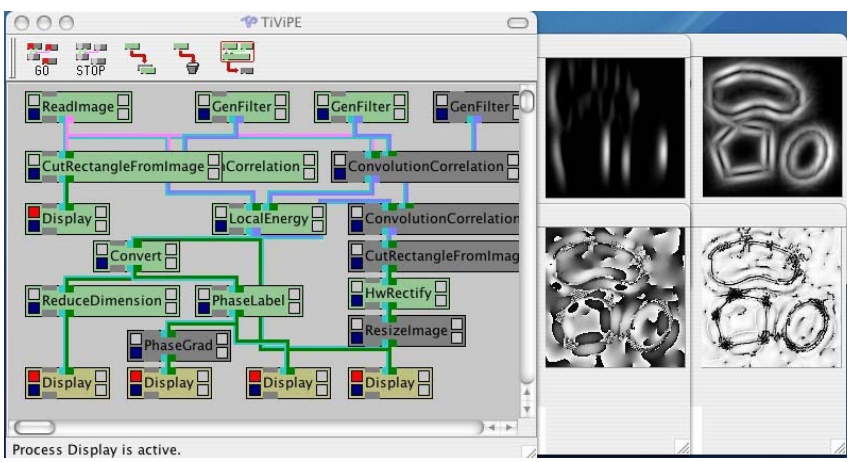

Figure 5: *The RPL workspace for the calculation of the Phase labeling for a simple test image.*

A hypercolumnar processing model for perceptual grouping has been presented which employs a relaxation phase labeling procedure for preattentive segmentation of objects in phase space. By introducing directional responses and local constraints thereupon, serving the grouping of similar directions and the decoupling of both sides of a contour line, the proposed mechanism is able to detect zero-crossings in phase space without an explicit and biological implausible search.

The gradient in phase space is sharpened compared to the edge response or the intensity discontinuity, and the whole scene is labeled into perceptual objects and the background. The relaxation phase labeling process is able to extract the most salient contour lines of perceptual groups in phase space suppressing false responses generated from the preprocessing stage, and can be used to link edges into object contours by closing gaps in the contour lines of the intensity image, or the grouping of perceptual primitives like dots or dashes into perceptual wholes modeling grouping principles proposed by Gestalt psychology.

For a more complete perceptual grouping scheme involving multiple spatial frequencies and multiple feature domains, the system could in principle be expanded by a scale space approach and the integration of parallel texture and color specific processing channels. A further expansion on the feature level is the integration of distinctive maps for two-dimensional features like direction of motion, curvature, endstoppings and junctions.

An extension on the conceptual level iss the inclusion of Helmholtzian principles including top-down inference to resolve ambiguities and the learning of memory traces (previously seen objects) and mental structure (compatibility constraints). The goal of the proposed mechanism is the interpretation of the 3-D world in terms of Gestalt-laws like collinearity and orthogonality for the perception of depth, and the inference of depth structure from the visual image, thereby helping to solve inverse problems in vision.