Proceedings of the 5th WSEAS Int. Conf. on SIGNAL, SPEECH and IMAGE PROCESSING, Corfu, Greece, August 17-19, 2005 (pp292-297)

Perceptual Grouping using Hypercolumnar Relaxation Phase Labeling

WINFRIED A. FELLENZ
Department of Cognitive and Linguistic Sciences
Brown University
Providence, R.I., 02912

Abstract: - A network for the extraction of salient image structure in dot patterns is proposed which uses a phase diffusion process to label the image into holistic perceptual objects and the background. The image is processed using three successive stages, copying the design of biological visual mechanisms. The population coded direction specific edge representation generated from phase-dependent energy filters is used by a competitive and cooperative phase process to allow boundary completion, grouping of perceptual elements and the suppression of false responses generated by the edge detection stage.

Key-Words: - Preattentive grouping, boundary completion, visual perception, Gestalt

1 Introduction

The partitioning of an image into its elementary objects and the background is the first task for an image recognition system linking low-level feature extraction to high-level scene interpretation. The robust segmentation of the image is difficult due to frequent intensity variations caused by discontinuities of the material reflectance and orientation of the surface shape, differing textures, and illumination effects. As object boundaries normally coincide with intensity discontinuities in the image, many segmentation methods rely on a low-level edge detection stage and group the extracted edges into region boundaries using edge following and linking methods. To extract globally salient structures in contour images, several parallel bottom-up mechanisms have been proposed.

A general mechanism for line and curve enhancement using local constraints has been proposed [14] which achieves a globally consistent interpretation of an image by the application of a relaxation process. The labels which are finally assigned to the image primitives have to be mutually consistent to maximize an explicit functional which depends on a multidimensional matrix of compatibilities. A related low-level mechanism was described which groups contour segments into salient groupings by iteratively computing a saliency measure based on simple geometrical attributes at each point in the image [27].

In an influential theory proposed by David Marr [18] the processing of visual information is supposed to take place in a hierarchy of successive representations called sketches which extract 3-D models of objects in a stepwise process of information integration. This coding proceeds bottom-up and involves the extraction of attributed primitives in a primal sketch which are then grouped by a mechanism called theta-aggregation. The intrinsic network architecture for early vision proposed in [3] supposes the integration of multiple cues in a parallel fashion with only local interactions acting within and between registered arrays, representing different aspects of the original intensity image. The imposed constraints and edge modification processes operate continuously between the maps to recover accurate intrinsic scene characteristics, at the same time maintaining consistency with the observed intensity at each point.

Computational models for extracting perceptual structure from dot patterns have been proposed using techniques ranging from oriented receptive fields [30, 12], cooperative and competitive feedback interactions of local features [11], virtual lines [26, 25], grouping by proximity [5], mean field annealing [13] to the use of Voronoi tessellations [1], and geometrical regularities [17, 28, 19].

A biologically motivated model for perceptual grouping and image segmentation incorporating multiple processing stages for sequential object extraction was proposed by the author [7, 8, 9]. The first processing stages extract an early visual representation of the image using direction specific edge responses. The next stage, which will be further discussed in section 4 involves the application of a phase relaxation scheme for the emergent extraction of perceptual groups by the propagation of local constraints.

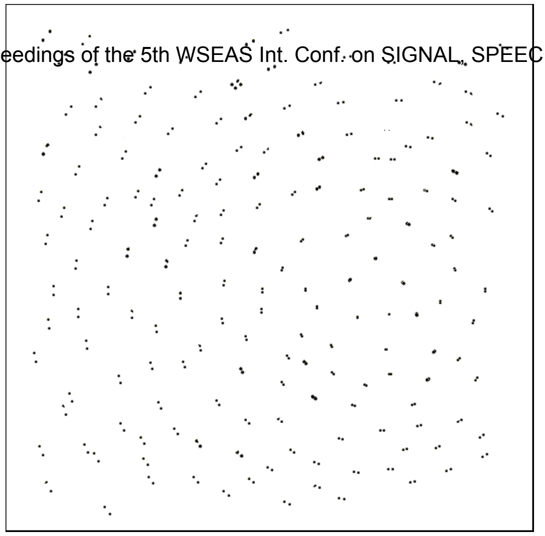

Figure 1. *Glass-pattern constructed from a random dot pattern and a superimposed slightly rotated copy.*

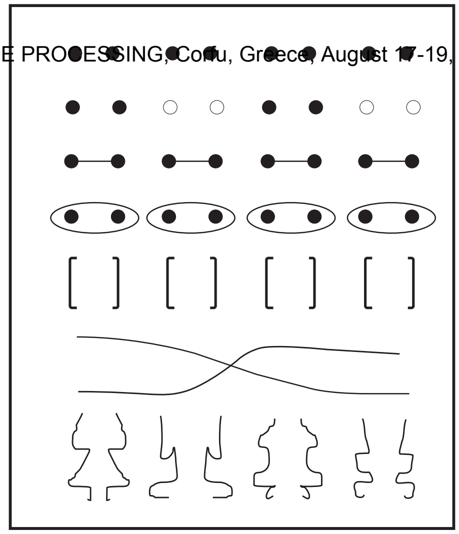

Figure 2. *The proposed Gestalt-laws: grouping by proximity, similarity, connectedness, common region, closure, good continuation and symmetry (top to bottom).*

2 Perceptual Organization

Perceptual organization is the process by which particular relationships among potentially separate stimulus elements are perceived and guide the interpretation of those elements. An example of the preattentive emergence of a holistic percept can be seen in the Glass pattern [10] of figure 1 where a simple random dot pattern has been slightly rotated and overlayed on the original pattern giving rise to the holistic perception of a rotating pattern going beyond the simple local grouping of nearby dots.

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 biological implausible.

Psychophysical experiments concerning the perceptual strength of preattentively grouped parts [24] 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.

There have been four major theoretical approaches to perceptual organization:

2.1 The Structuralist View

The structuralist view denies the existence of perceptual organization and assumes that innately endowed sensory organs associate sensations with memory traces. Any nonlinear effects have to be attributed to the interaction of these memory images and not to the elementary sensations.

2.2 The Helmholtzean View

This view extends the structuralist view by assuming a likelihood principle during unconscious inference, allowing the perceptual process to solve ambiguities and to organize the sensory elements into the most probable object. Furthermore, great emphasis is placed on the acquisition of a mental structure during perceptual learning.

2.3 The Gestalt View

Gestalt psychologists assume on the contrary 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 holistic percept.

The Gestaltists have attempted to reveal the laws of perceptual organization underlying the partitioning decisions which group stimulus elements into perceptual wholes [29, 15, 16]. They proposed the general principle of *Prägnanz* which means organization toward the most regular, ordered, stable and balanced state possible.

This principle was accompanied by a collection of *Gestalt laws* depicted in figure 2 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

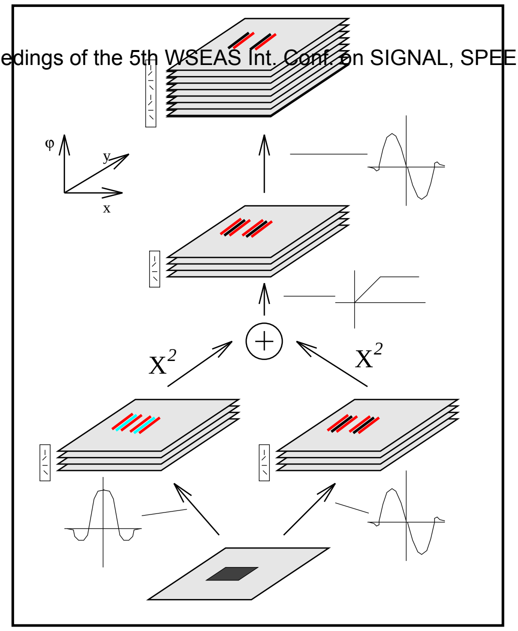

Figure 3. *A hypercolumn composed of direction specific ON and OFF channels generated from local energy filters specified as the sum of squared quadrature phase Gabor-filters.*

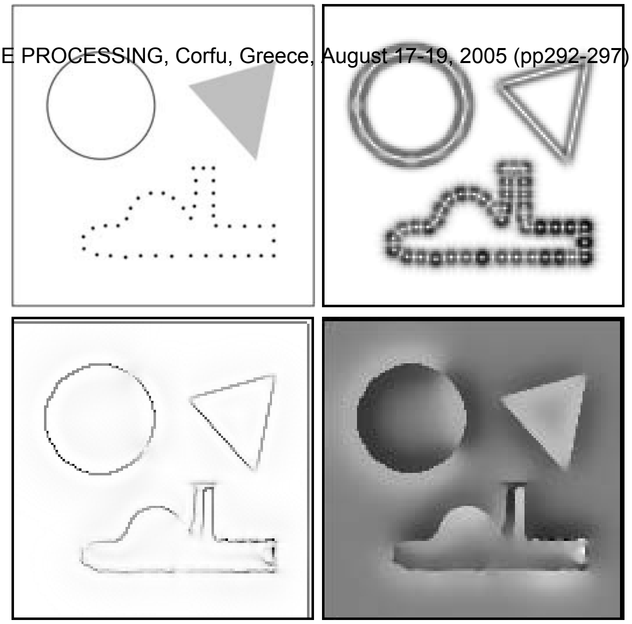

Figure 4. *Clockwise from top left: a) Scene with simple contour defined objects; b) Sum of ON and OFF channels; c) Phase image; d) Phase gradient.*

impression. The quintessence of the Gestalt argument is that perceptions are organized globally to simplify the representation of the stimulus.

2.4 The Gibsonian View

The direct perception view denies the existence of a mental structure during unconscious perception and assumes that the organization resides in the stimulus. The perceptual system only needs to be tuned accordingly to pick up the structure already available in the stimulus array.

3 The Hypercolumnar Architecture

The proposed perceptual grouping model consists of three successive processing stages which model mechanisms found in visual cortex [7]. The first stage for the hierarchical extraction of contour lines has been adapted from the phase-dependent energy model [20] using quadrature phase Gabor-Filters [6] for the detection of local energy in the image. The responses of six 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 3) to build an dualistic representation of intensity changes in the image. The Gabor-filters are defined as:

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

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

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 α is a normalization factor which is needed to allow the symmetric kernel to be zero integrable.

4 The Relaxation Procedure

The direction selective responses at each image position are combined into hypercolumns and local constraints within and between these columns are defined to allow the emergent extraction of perceptual objects. Each hypercolumn is tagged with an associated phase label to allow phase discontinuities to emerge at object boundaries which correspond to zero crossings of the smoothed second intensity derivative.

The emergent forming of perceptual groups, including both contour and region based information is depicted in figure 4 showing the results of the proposed grouping scheme for a scene with three simple objects. Although the objects are defined

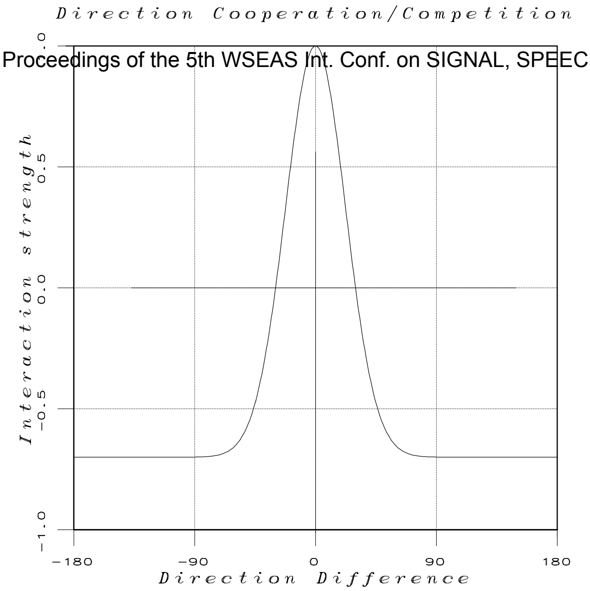

Figure 5. *The compatibility function between direction responses modeled as a shifted Gaussian.*

by different boundary types ranging from intensity discontinuities over lines to dot patterns, the phase gradient shows a common interpretation of all contour types.

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 5). This compatibility function $v_{m,n}$, is modeled as a shifted Gaussian:

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

using the set of discrete directions $m, n \in M$ and the parameters $\lambda = 1.7$, $\mu = 0.3$ and $\eta = 0.7$. 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:

$$\begin{aligned} f(\delta\Phi) &= \frac{\delta\Phi}{\pi}(\beta + cos(\delta\Phi)) && (4)\\ \delta\Phi &= \phi_{i+k,j+l} - \phi_{ij}; && (5)\\ -\pi &<= \delta\Phi < \pi. && (6) \end{aligned}$$

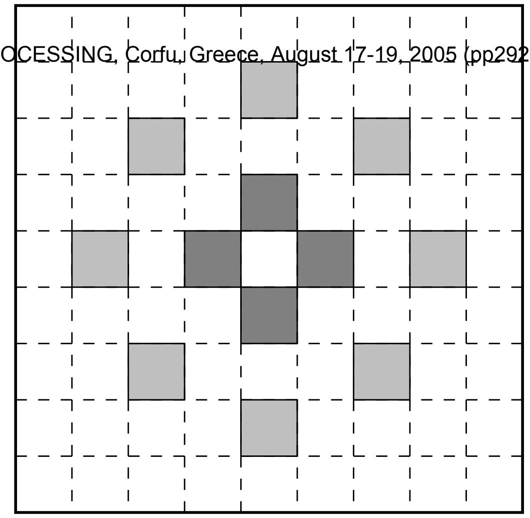

Figure 6. *The sparse connectivity scheme $h_{k,l}$ for the propagation of local phase and edge responses between neighboring columns.*

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 6) to speed up the synchronization between columns:

$$\begin{aligned} \dot{\phi}_{i,j} &= \omega_{i,j} + \sum_{m,n \in M} v_{m,n} E_{i,j}(m) Z_{i,j}(n) && (7)\\ Z_{i,j}(n) &= g\Big(\alpha \sum_{k,l \in N} h_{k,l} E_{i+k,j+l}(n) f(\delta\Phi)\Big) && (8) \end{aligned}$$

where $\omega_{i,j}$ is a zero mean random variable introducing noise into the phase process, thereby resolving ambiguous situations, and forcing the process to move from the 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 [21] 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.

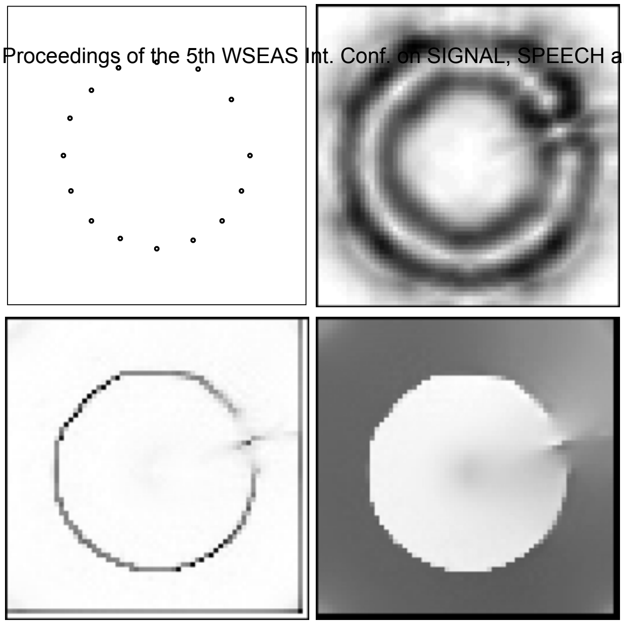

Figure 7. *Clockwise from top left: a) Scene with circular object; b) Sum of ON and OFF channels; c) Phase image; d) Phase gradient.*

5 Results

Figure 7 shows the result of applying the proposed perceptual grouping process to a circular object defined by differently spaced dots. Note the circular structure visible in the phase gradient and the discontinuity in the upper right of the grouping indicating a deviation from the homogeneous dot spacing in the stimulus. This feature of the grouping process shares similarities to human visual processing where salient configurational aspects of the stimulus seem to pop out preattentively.

The perceptual groupings detected in figure 8 resemble the visual perception of most of the dot groups in the image, although not all groupings are detected at the chosen scale. Note that due to circular phases a white phase label is similar to a black phase label; also note that the frequent black spots in the phase gradient are caused by phase singularities which usually disappear as the relaxation process continues.

A further aspect of the grouping process is visible in the upper left part of the phase gradient in figure 8d): due to the long range interactions within and between perceptual groups, illusory phase gradients appear in the middle of the elongated object and between the perceptual groups in the upper left of the image. This additional gradient can be compared to the result of applying a skeleton filter to the extracted objects and could be used to extract the middle line of objects or between objects in the image.

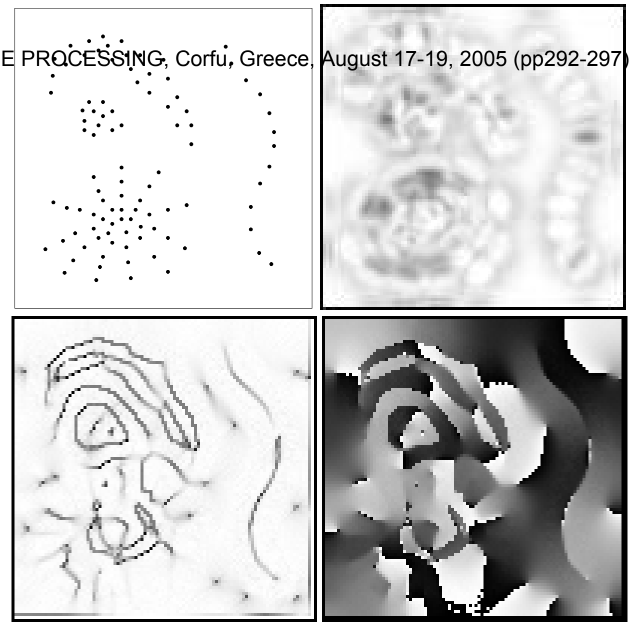

Figure 8. *Clockwise from top left: a) Scene with dot groupings; b) Sum of ON and OFF channels; c) Phase image (white correspond to black phase labels); d) Phase gradient.*

6 Conclusion

A hypercolumnar processing model for perceptual grouping has been proposed 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. Furthermore, 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 [4, 22] and the integration of parallel texture-, and color specific processing channels [23, 2]. A further extension on the feature level would be the integration of distinctive maps for two-dimensional features like direction of motion, curvature, endstoppings and junctions.

An extension on the conceptual level would be 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 would be 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.

References

[1] N. Ahuja and M. Tuceryan. Extraction of early perceptual structure in dot patterns: integrating region, boundary, and component Gestalt. *Computer Vision, Graphics, and Image Processing*, Vol.48, 1989, pp. 304–356.

[2] J. Aloimonos and D. Shulman. *Integration of Visual Modules: An Extension to the Marr Paradigm*. Academic Press, 1989.

[3] H. G. Barrow and J. M. Tenenbaum. Interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence*, Vol.17, 1991, pp. 75–116.

[4] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, Vol.31, No.4, 1983, pp. 532–540.

[5] B. J. Compton and G. D. Logan. Evaluating a computational model of perceptual grouping by proximity. *Perception and Psychophysics*, Vol.53, No.4, 1989, pp. 403–421.

[6] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Optical Society of America A*, Vol.2 1985, pp. 1160–1168.

[7] W. A. Fellenz. A neural network for preattentive perceptual grouping. In *Proc. 4th Irish Neural Network Conference*, 1994, pp. 67–72.

[8] W. A. Fellenz and G. Hartmann. Image segmentation by phase label diffusion. In *Proc. Int. Conference on Artificial Neural Networks, ICANN-95*, Vol. II, 1995, pp. 309–314.

[9] W. A. Fellenz and G. Hartmann. Preattentive grouping and attentive selection for early visual computation. In *Proc. Int. Conf. on Pattern Recognition*, IEEE, 1996, Vol. IV, pp. 340–345.

[10] L. Glass. Moiré effect from random dots. *Nature*, Vol.243, 1969, pp. 578–580.

[11] S. Grossberg and E. Mingolla. Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Perception and Psychophysics*, Vol.38, No.2, 1985, pp. 141–171.

[12] G. Guy and G. Medioni. Inferring global perceptual contours from local features. *International Journal of Computer Vision*, Vol.20, No.1, 1996, pp. 113–133.

[13] L. Hérault and R. Horaud. Figure-ground discrimination: A combinatorial optimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.15, No.9, 1993, pp. 899–914.

[14] R. A. Hummel and S. W. Zucker. On the foundations of relaxation labeling processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.5, 1983, pp. 267–287.

[15] K. Koffka. *Principles of Gestalt Psychology*. Hartcourt, Brace & World, New York, 1935.

[16] W. Köhler. *Gestalt Psychology: An introduction to new concepts in modern psychology (revised edition)*. Liveright, New York, 1947.

[17] D. G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer, Boston, 1985.

[18] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman, San Francisco, 1982.

[19] R. Mohan and R. Nevatia. Perceptual organization for scene segmentation and description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.14, No.6, 1992, pp. 616–634.

[20] M. C. Morrone and D. C. Burr. Feature detection in human vision: A phase-dependent energy model. *Proc. Royal Society London* B, Vol.235, 1988, pp. 221–245.

[21] M. A. Paradiso and K. Nakayama. Brightness perception and filling-in. *Vision Research*, Vol.31, No.7/8, 1991, pp. 1221–1236.

[22] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.12, No.7, 1990, pp. 629–639.

[23] T. Poggio, E. B. Gamble, and J. J. Little. Parallel integration of visual modules. *Science*, Vol.242, 1988, pp. 436–242.

[24] J. R. Pomerantz and E. A. Pristach. Emergent features, attention, and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception and Performance*, Vol.15, No.4, 1989, pp. 635–649.

[25] J. T. S. Smits and P. G. Vos. A model for the perception of curves in dot figures: The role of local salience of 'virtual lines'. *Biol. Cyb.*, Vol.54, 1986, pp. 407–416.

[26] K. A. Stevens. Computation of locally parallel structure. *Biological Cybernetics*, Vol.29, 1978, pp. 13–28.

[27] S. Ullman. Low-level aspects of segmentation and recognition. *Proceedings of the Royal Society of London* B, Vol.337, 1992, pp. 371–379.

[28] R. J. Watt. *Visual Processing: Computational, Psychophysical and Cognitive Research*. Erlbaum, 1988.

[29] M. Wertheimer. Untersuchungen zur Lehre von der Gestalt II. *Psych. Forsch.*, Vol.4, 1923, pp. 301–350.

[30] S. W. Zucker. Early orientation selection: Tangent fields and the dimensionality of their support. *Computer Vision, Graphics, and Image Processing*, Vol.32, 1985, pp. 74–103.