Visual Organization for Figure/Ground Separation

Davi Geiger
Courant Institute
251 Mercer Street
New York, NY 10012

Krishnan Kumaran
Smith Kettlewell Institute
2232 Webster Street
San Francisco, CA 94115

Laxmi Parida
Courant Institute
251 Mercer Street
New York, NY 10012

Abstract

A common factor in all illusory contour figures is the perception of a surface occluding part of a background. In our previous work, we have shown we could diffuse a proper set of junction hypothesis (what is salient or background) to obtain a surface where their boundaries represented illusory contours. Amodal completions emerge at the overlapping surfaces.

We address the problem of selecting the best image organization (set of hypothesis). We propose an optimization criteria based on a coherence measure between pairs of junctions (correlation between the diffusion of each pair). A statistical physics approach to select the best organization is applied. The experiments suggest that despite the large number of possible organizations our approach may take only a few steps (in organization space) to select the best one.

1 Introduction

When the Kanizsa square image is shown various possible visual organizations, figure-ground separations, are plausible (see figure 1).

From a computational view, two natural questions to ask are (a) how the shape of these illusory contours arise ? and (b) why these contours are seen and not others among all possible ones, i.e., how a visual organization is selected ?

These two questions are intriguing not just to explain these illusions, but because they address the problems of an itermediate visual process, grouping, needed to link low level/local feature extraction to recognition.

Our proposed approach: We take the view that this intermediate process task is to resolve the figure/ground separation problem. We note that a perception of an object separated from the ground emerges from these illusions.

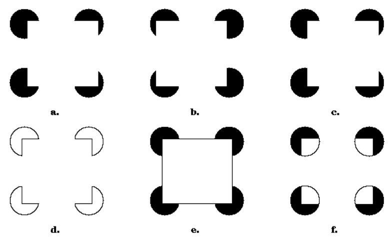


Figure 1: Kanizsa stereo pair. Different possible interpretations of the Kanizsa square figure. Cross-fusing (b.) and (c.) leads to percept (e.) while cross-fusing (a.) and (b.) leads to percept (f.). Percept (d.) is obtained from monocular images when interpreted "literally", i.e. without the illusory contours.

Our approach is to iteratively find the most salient surface by (i) detecting occlusions; (ii) assigning a set of hypothesis of the local salient surface configuration; (iii) applying a Bayesian model to diffuse these hypothesis; and (iv) efficiently selecting the best image organization (set of hypothesis) based on a coherence measure (correlation of the activities/diffusion of junctions pairs). We also consider an entropy measure.

Here we skip the discussion on (i), briefly review (iii) (see our work [9]). We then, focus on (ii) and (iv). In our approach the illusory contours arise from the surface boundaries and the amodal completions emerge at the overlapping surfaces.

Previous approaches: Many computational models have been proposed to describe the formation of illusory contours. The following ones take the primary view of extending the intensity edges, Ullman[29], Heitger & von der Heydt [14], Grossberg & Mingolla[12] and Grossberg [13], Guy & Medioni[11], Kellman & Shipley[20] , Mumford [23], Williams & Jacobs [32]. The approach of regions is presented by Nitzberg & Mumford [25] [24] and Brady & Grimson[3]. We have recently proposed a diffusion

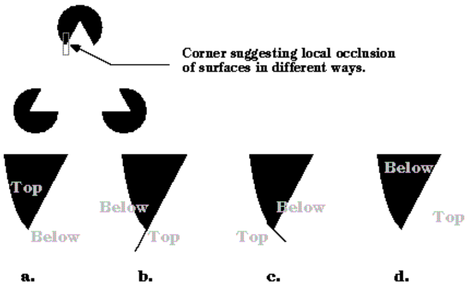


Figure 2: Each corner has five (5) local salient surface interpretations. (a.) Convex shape. Salient black (on top). (b.,c.) Two different T-junction interpretations. (Extension shown to denote depth edge) (d.) Concave shape. Salient white (on top). The last one is the null hypothesis, i.e., no salient surface.

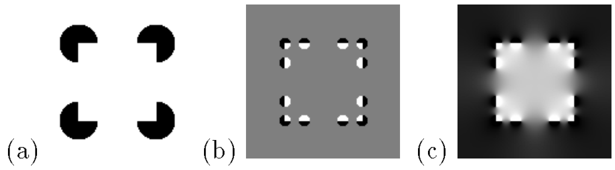


Figure 3: (a) The Kanizsa square and (b) Initial available data - $\lambda_k = 1$: wither white, $\sigma_k^0 = 1$ (salient) or black, $\sigma_k^0 = 0$ (not salient); when no information is availble - $\lambda_k = 0$: represented by grey. (c) diffusion, σ.

approach [9].

There is very little discussion on how a particular illusory contour is selected, but rather focus on the shape of the contour. The discussion on visual organization date back at least to the labeling scheme by Huffman [17], extended in Waltz [30] and it is followed in Williams and Hanson [31]. It is to weak constraint in the sense that allows for far to many possible organizations (all of them equally likely). Shashua and Ullman [28] have a line integration/selection approach that does not account for surface properties.

2 Junctions and surface hypothesis

It is well known that when occlusions are detected, an immediate sensation of depth change occurs. This can occur in monocular images due to features such as line endings and junctions. In stereo vision, occlusions along an epipolar line in one eye correspond to depth discontinuities in the other[8]. Occlusions are typically perceived where junctions or line stops occur.

Thus, we start by identifying the local occlusion cues such as T-junctions, Y-junctions, corners and line endings. Each of these cues could suggest various local occlusion scenarios, e.g. corners have multiple occlusion interpretations (see figure 2). Initial information is then provided in these regions in the form of salient-surface-states representing local surface configurations.

Salient-surface-state:

We define a binary field $\sigma_k \in [0,1]$ representing the salient-surface-state at pixel k on the image. More precisely, $\sigma_k = 1$ indicates that the pixel k is salient, and $\sigma_k = 0$ otherwise. Initially the information is only availble at occlusion. We are indexing the pixels according to raster scanning convention, i.e. $k = 1, 2, ..., N^2$ where N is the width of the square image.

We define $\{(\sigma_0^{a,p})_k, \lambda_k^p\}$ as the hypothesis a at each junction p. We may have M hypothesis, $a = 1, 2, ..., M$, for each of the P junctions, $p = 1, 2, ..., P$. The binary parameter $\lambda_k^p = 0, 1$ indicates whether a junction is present or not at pixel k.

Corners (as T-junctions): At a corner, five (5) different possible configurations (hypothesis) for $(\sigma_0^a)_k = 1, 0$ must be analyzed (see figure 2), i.e., $a = 1, 2, 3, 4, 5$. Different from previous work, we are conjecturing a mechanism where even at corners a T-junction is hypothesized. This mechanism will be of key importance to obtain the formation of illusory surfaces.

T- and Y- junctions: These 3-junctions admit any of the three surface patches on top, with the angles biasing the probability. A null hypothesis must be always added to consider the junction not part of the salient surface.

End-Stopping: At an end stop of a bar, a surface discontinuity is suggested, the bar is being occluded at that location. Thus, just one interpretation is provided and the null hypothesis, i.e., $a = 1, 2$.

It is plausible that each configuration $(\sigma_0^{a,p})_k$ has a different probability to occur and a statistical analysis of scenes would be necessary to estimate $P(I|\{(\sigma_0^{a,p})_k, \lambda_k^p\})$. For simplification, we will set all of them equal, i.e., we will consider the image to be conditionally independent on the parameters $\{(\sigma_0^{a,p})_k, \lambda_k^p\}$.

3 Diffusion process from sparse data

In order to diffuse the salient-surface-state values from the sparse junction locations, we have considered [9][22] a linear diffusion approach where the diffusion coefficients block the diffusion at intensity edges. More precisely, we apply

$$\sigma^{a,p} = (\mathbf{A} - \mathbf{I} + \sum_{\mathbf{p=1}}^{\mathbf{P}} \lambda^p)^{-1} \, \lambda^p \, \sigma_0^{a,p} \, . \qquad (1)$$

where $\sigma_0^{a,p}$ and $\sigma^{a,p}$ are N^2 dimensional vectors containing the values of $(\sigma_0^{a,p})_k$ and $\sigma_k^{a,p}$ respectively. The $N^2 \times N^2$ diagonal matrix λ^p has entries λ_k^p. The symmetric and block tri-diagonal matrix $\mathbf{A}$ have the following structure

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & -\mathbf{D}_1^\dagger & \mathbf{0} & \dots \\ -\mathbf{D}_1 & \mathbf{A}_2 & -\mathbf{D}_2^\dagger & \ddots \\ \mathbf{0} & -\mathbf{D}_2 & \mathbf{A}_3 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}$$

where the matrices $\mathbf{D}_i$ and $\mathbf{A}_i$ are given by

$$\mathbf{D}_i = \begin{pmatrix} \mu^h_{i.(N-1)} & 0 & \dots \\ 0 & \mu^h_{i.(N-1)+1} & 0 \\ \vdots & 0 & \ddots \end{pmatrix},$$

$$\mathbf{A}_i = \begin{pmatrix} \chi_{i.(N-1)} & -\mu^v_{i.(N-1)} & 0 \\ -\mu^v_{i.(N-1)} & \chi_{i.(N-1)+1} & \ddots \\ 0 & \ddots & \ddots \end{pmatrix}$$

where $\chi_k = 1 + \mu^h_{k-N} + \mu^h_k + \mu^v_{k-1} + \mu^v_k$ and $i = 1, 2, \ldots N-1$ is a row index. In order to prevent smoothing along intensity boundaries, we define the smoothing coefficients as

$$\mu_i^h = \mu \, (1 - e_i^h) \quad \text{and} \quad \mu_i^v = \mu \, (1 - e_i^v) \, ,$$

where $e_i^{h(v)} = 1$ indicates the presence of a horizontal edge (or vertical edge) at pixel i. The intensity edge map allow for neighbor pixels to interact with the same "color" ones and to block the ones of different "color".

We note that (1) minimizes

$$E(\sigma^{a,p}) = (\sigma^{a,p} - \mathbf{M}^{-1} \lambda^p \sigma_0^{a,p})^\dagger \, \mathbf{M} \, (\sigma^{a,p} - \mathbf{M}^{-1} \lambda^p \sigma_0^{a,p}) \, .$$

where $\mathbf{M} = \mathbf{A} - \mathbf{I} + \sum_{\mathbf{p}} \lambda^{\mathbf{p}}$.

Computation: To compute $\sigma^{a,p}$ can use a computationally expensive Cholevsky decomposition or an approximate very fast and parallel method, based on renormalization group (averaging process). See results in figures 4 and 3. We can also show the solution $\sigma_k^{a,p}$ is bounded by the initial data $(\sigma_0^{a,p})_k$. Thus, $\sigma_k^{a,p} \in [0, 1]$ if $(\sigma_0^{a,p})_k \in [0, 1]$ and the $\sigma_k^{a,p}$ can be interpreted as the degree of saliency (salient-surface-states)[22].

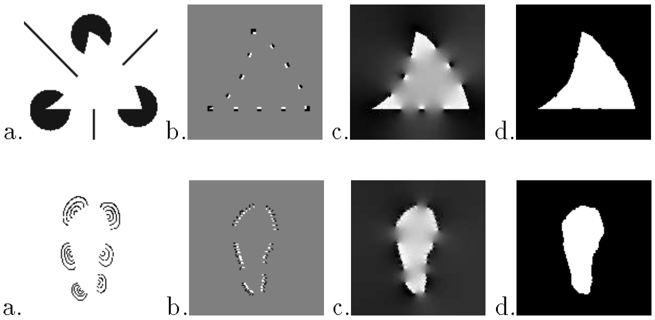
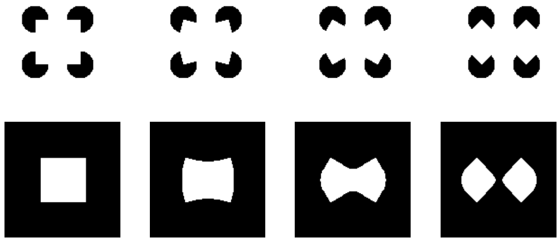
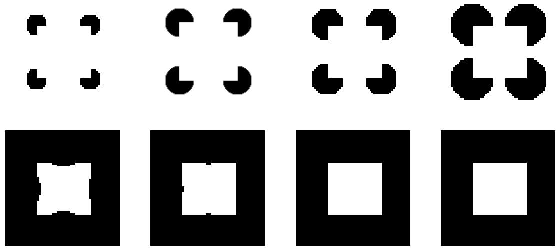


Figure 4: Two examples of illusory surface perception, stressing the surface aspect of the illusory figures. Each of the above figures depict a. The original figure. b. Input data - white=1; black=0; gray=0.5, $\sigma_0(\phi)$ c. The reconstruction (diffused data), $\sigma(\phi)$. d. Thresholding in 0.5.



$O(\sigma) = 0.16$ $O(\sigma) = 0.17$ $O(\sigma) = 0.20$ $O(\sigma) = 0.23$

Figure 5: The change in the perceived surface leads gradually to the changed grouping of features as the vertices of the illusory square are rotated, while the (optimal organization) entropy values show a degradation in the perception of the illusory contours with increasing curvature.



$O(\sigma) = 0.45$ $O(\sigma) = 0.39$ $O(\sigma) = 0.23$ $O(\sigma) = 0.04$

Figure 6: The perception of the illusory square grows stronger as the ratio γ=(circle radius)/(square side) is increased, as depicted by the (optimal organization) entropy values. The model only converges to a square if the ratio exceeds a threshold, which is in qualitative agreement with human perception.

4 Salient Organization and Grouping

We now define a measure to evaluate the saliency of a surface obtained from a set of hypothesis obtained at the occlusion cues.

We first introduce a binary field $\phi_{a,p} = 0, 1$ that selects one hypothesis for each junction, when $\phi_{a,p} = 1$ the $p-th$ junction is assigned to the $a-th$ hypothesis. Since only one hypothesis is admitted per junction we have

$$\sum_{a=1}^{M} \phi_{a,p} = 1 \,. \qquad (2)$$

This constraint yields the initial data

$$\sigma_0(\phi) = \sum_{p=1}^{P}\sum_{a=1}^{M} \phi_{a,p}\sigma_0^{a,p} \qquad \text{and} \qquad \lambda = \sum_{p=1}^{P} \lambda^p \,.$$

When this data is diffused by (1), we obtain

$$\begin{aligned}
\sigma(\phi) &= (\mathbf{A} - \mathbf{I} + \lambda)^{-1}\lambda\sigma_0(\phi) \\
&= (\mathbf{A} - \mathbf{I} + \lambda)^{-1}(\sum_{\mathbf{p=1}}^{\mathbf{P}}\lambda^p)(\sum_{p=1}^{P}\sum_{a=1}^{M}\phi_{a,p}\sigma_0^{a,p}) \\
&= (\mathbf{A} - \mathbf{I} + \lambda)^{-1}\sum_{p=1}^{P}\sum_{a=1}^{M}\phi_{a,p}\ \lambda^p\ \sigma_0^{a,p} \\
&= \sum_{a=1}^{M}\phi_{a,p}(\mathbf{A} - \mathbf{I} + \lambda)^{-1}\lambda^p\ \sigma_0^{a,p} \\
&= \sum_{p=1}^{P}\sum_{a=1}^{M}\phi_{a,p}\sigma^{a,p} \,, \qquad (3)
\end{aligned}$$

where we have used the non-overlapping property $\sum_{p'} \lambda^{p'}\phi_{a,p}\sigma_0^{a,p} = \delta_{p,p'}\lambda^p\phi_{a,p}\sigma_0^{a,p}$. This linear result allows us to obtain a diffusion of a whole organization by simply summing the diffusions due to each junction.

Coherence to bind an organization: We propose a selection criteria of a visual organization based on the coherence of each pair of junctions/hypothesis and the sum over all pairs to obtain the final score for a visual organization (a set of hypothesis). More precisely, we propose the coherence to be defined as

$$\begin{aligned}
O(\phi) &= \frac{1}{2\,N^2}\sum_{a,b=1}^{M}\sum_{p,q=1}^{P}\phi_{a,p}\phi_{b,q}\,[\sum_k(\sigma_k^{a,p} - \sigma_k^{b,q})^2] \\
&= \frac{1}{N}\sum_{a=1}^{M}\sum_{p=1}^{P}\phi_{a,p}\sum_k(\sigma_k^{a,p})^2 \\
&\quad -\frac{1}{N^2}\sum_{a,b=1}^{M}\sum_{p,q=1}^{P}\phi_{a,p}\phi_{b,q}\sum_k\sigma_k^{a,p}\,\sigma_k^{b,q} \\
&= \sum_{a=1}^{M}\sum_{p=1}^{P}A_{a,p}\phi_{a,p} - \sum_{a,b=1}^{M}\sum_{p,q=1}^{P}B_{a,p}^{b,q}\,\phi_{a,p}\,\phi_{b,q} \,,
\end{aligned}$$

where $A_{a,p} = \frac{1}{N}\sum_k(\sigma_k^{a,p})^2$ and $B_{a,p}^{b,q} = \frac{1}{N^2}\sum_k\sigma_k^{a,p}\,\sigma_k^{b,q}$. The optimal ϕ will select an organization.

Complexity: Let us say we have P junctions, each one having M possible configuration states. The total number of hypothesis is M^P, i.e., exponential complexity with the number of junctions. One may reduce M by only allowing salient hypothesis that are "color coherent", e.g., assigning saliency $(\sigma_0)_k = 1$ to white regions. However, it still leads to an exponential growth on the total number of hypothesis (since $M > 1$ with the null hypothesis).

Statistical Physics: Vast literature exists on approximately solving this problem, e.g., [18][2]. We use a statistical physics method [7] [10] where the average state $\bar{\phi}_{ap}$ is estimated by the Gibbs distribution. More precisely,

$$\bar{\phi}_{ap} = \sum_{\phi}\phi_{ap}\frac{e^{-\beta O(\phi)}}{Z} \,,$$

where Z is a normalization constant, β is an annealing parameter and the sum over all ϕ is restricted to obey (2). Using a mean field approximation, the method leads to a solution of the form

$$\bar{\phi}_{ap} = \frac{e^{-\beta(A_{a,p} - \sum_{b,q}B_{a,p}^{b,q}\bar{\phi}_{bq})}}{\sum_{a=1}^{M}e^{-\beta(A_{a,p} - \sum_{b,q}\bar{\phi}_{b,q}B_{a,p}^{b,q})}} \,.$$

For $\beta = 0$ the solution is $\bar{\phi}_{ap} = 1/M$ since all hypothesis are equally likely, and as $\beta \to \infty$ a solution to the problem is obtained. We have noted that for the Kanizsa square (see figure 7) the local minima of a square in front of four circles is always reached even for starting values $\beta = 0.25$ (high values) and random configurations, suggesting that the algorithm justifyies the stability of this solution.

"Entropy" to bind an organization: Another organization criteria is to select the organization that produces a diffusion that is most salient, i.e., that produces σ_k closest to 0's and 1's.

$$O(\phi) = \frac{1}{N^2}\sum_k[\sigma_k(\phi)\,(1 - \sigma_k(\phi))]$$

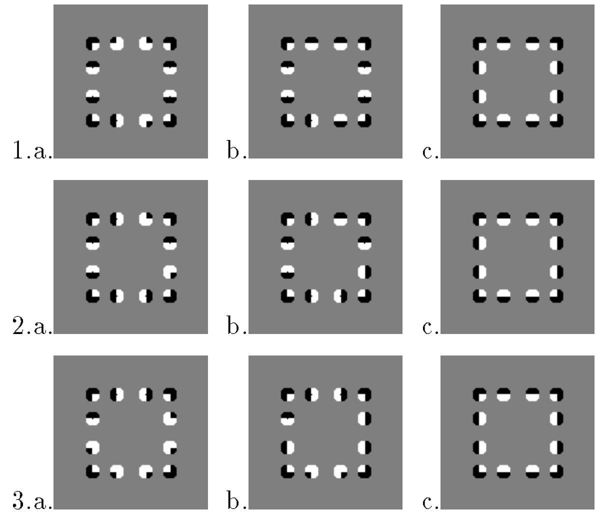


Figure 7: Three random initial data configurations are shown in (1,2,3)a. Note, the center corner, less ambiguous have been kept fixed, so we have a total of eight (8) corners that are "free". The sequence (a,b,c) from left to right depicts the convergence of the iterative algorithm to the correct solution in each case. We start with random values for $\phi_{a,p} = 0, 1$ and $\beta = 0.0001$. iteratively set $\beta^{t+1} = 1.1\beta^t$ and compute the total change in $\bar{\phi}_{a,p}$ (stop criteria if the total change is less than threshold 0.01). We display only thresholded ($Thr(\bar{\phi}_{a,p}) = 0, 1$) intermediate steps and after three (3) steps (out of 4^8 possible ones) the system is on the desired state.

$$= \frac{1}{N^2} \sum_{p=1}^{P} \sum_{a=1}^{M} \phi_{a,p} (\sigma^{a,p})^\dagger (\mathbf{1} - \sum_{q=1}^{P} \sum_{b=1}^{M} \phi_{b,q} \sigma^{b,q})$$

$$= \frac{1}{N^2} \sum_{a,p}^{M,P} \phi_{a,p} (\sum_k \sigma_k^{a,p}) - \sum_{a,p,b,q}^{M,P,M,P} \phi_{a,p} \phi_{b,q} (\sigma^{a,p})^\dagger \sigma^{b,q} .$$

This criterion favors organizations where regions have been more clearly segregated by the diffusion process, i.e., where the final states are closer to 1's and 0's, and penalizes for ambiguities $\sigma_k = 0.5$. It is very similar to the coherence criteria.

Results: For evaluation of perception degradation, see figures 5,6 and for the optimization process see figure 7.

5 Multiple surfaces and amodal completion

To extend the model to account for multiple surfaces and amodal completion we have considered a layered and stage wise model [25][16][1] [4]. At each stage s, the salient-surface is reconstructed and then, removed, and subsequently the same process continues to the next stage until all surfaces have been reconstructed. The overlapping regions correspond to amodal completions of partially occluded surfaces.

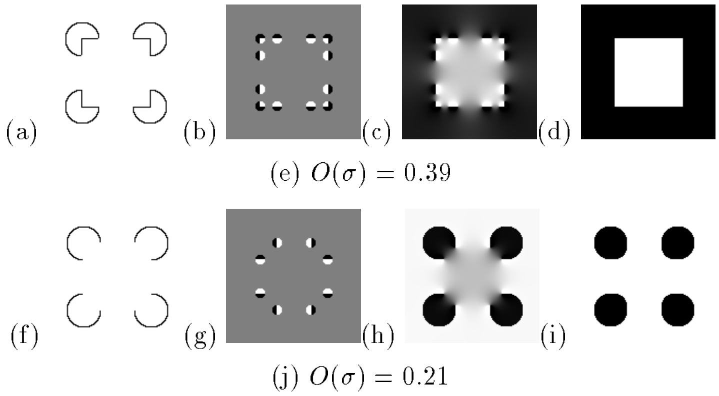


Figure 8: First layer: (a) Initial values of $\mu_k^{s=top}$; (b) Sparsity map - λ_k. (c) Reconstruction, σ_k; (d) thresholded at 0.5; (e) Entropy values. Removing the Kanizsa square, and reconstructing the Second Layer (Amodal completion): (f), (g), (h), (i), (j)

Updating μ_k^s and $\sigma_{0_k}^s$: Once the salient surface is reconstructed, at stage s, it is removed so that concealed surfaces can be reconstructed. The intensity edges associated to the first surface are also removed. One must update the salienct-surface-states (compute again an optimal organization) since the removed surface is no longer "occluding".

References

[1] E. Adelson and P. Anandan. Ordinal characteristics of transparency. In Proc. AAAI Workshop on Qualitative Vision, pages 552–573, CA, 1990.

[2] D. J. Amit, H. Gutfreund, and H. Sompolinsky. Spin-glass models of neural networks. *Physical Review A*, 32:1007–1018, 1985.

[3] M. Brady and W. E. L. Grimson. The perception of subjective surfaces. A.I. Memo No. 666, AI Lab., MIT, Nov. 1982.

[4] T. Darrell and A. Pentland. Cooperative robust estimation using layers of support. *IEEE Trans.*, PAMI-17(5):474–487, 1995.

[5] D. Field, A. Hayes and R. Hess. Contour Integration by the Human Visual System: Evidence for a Local 'Association field'. *Vis. Res.* 33:173-193, 1993.

[6] R. von der Heydt. E. Peterhans and G. Baumgartner. Neuronal responses to illusory contour stimuli reveal stages of visual cortical processing. In *Visual Neurosc.*, pages 343–351.1986.

[7] D. Geiger and F. Girosi. Parallel and deterministic algorithms for mrfs: surface reconstruction. *IEEE Trans. PAMI*, PAMI-13(5):401–412, May 1991.

[8] D. Geiger, B. Ladendorf, and A. Yuille. Binocular stereo with occlusion. In *2nd ECCV*, Santa Marguerita, Italy, May 1992. Springer-Verlag.

[9] D. Geiger and K. Kumaran. Illusory Contours and Visual Organization. *4th ECCV96*, Cambridge, England, April 1996.

[10] D. Geiger and A. Yuille. A common framework for image segmentation. *International Journal of Computer Vision*, 6:3:227–243, August 1991.

[11] G.Guy and G. Medioni. Inferring global perceptual contours from local features. In *Proc. IU Workshop DARPA*, Sept. 1992.

[12] S. Grossberg and E. Mingolla. Neural dynamics of perceptual grouping:textures, boundaries and emergent segmentations. *Perception & Psychophysics*, 38(2):141–170, 1985.

[13] S. Grossberg. 3-d vision and figure-ground separation by visual cortex. *Perception & Psychophysics*, 55(1):48–120, 1994.

[14] F. Heitger and R. von der Heydt. A computational model of neural contour processing: Figure-ground segregation and illusory contours. *Proceedings of the IEEE*, pages 32–40, 1993.

[15] R. von der Heydt, E. Peterhans, and G. Baumgartner. Illusory contours and cortical neuron responses. *Sci. Wash.*, 224:1260–1262, 1984.

[16] S. Hsu, P. Anandan, and S. Peleg. Accurate computation of optical flow by using layered motion representation. In *ICPR*, Oct 1994.

[17] D.A. Huffman. A duality concept for the analysis of polyhedral scenes. In *Machine Intelligence*, volume 6. Edinb. Univ. Press, Edinb., U.K., 1971.

[18] A. Kamath and N. Karmarkar. A continuous approach to compute upper bounds in quadratic minimization problems with integer constraints. In C. A. Floudas and P. M. Pardalos, ed., *Recent Adv. in Global Opt.*, volume 2, 229–241, 1991.

[19] G. Kanizsa. *Organization in Vision.* Praeger, New York, 1979.

[20] P.J. Kellman and T.F. Shipley. A theory of visual interpolation in object perception. *Cognitive Psychology*, 23:141–221, 1995.

[21] I. Kovacs and B. Julesz. Perceptual Sensitivity Maps within Globally Defined Visual Shapes. *Nature*, vol.370:644-646, 1994.

[22] K. Kumaran, D. Geiger, and L. Gurvits. Illusory surfaces and visual organization. *Network:Comput. in Neural Syst.*, 7(1), Feb. 1996.

[23] D. Mumford. Elastica and computer vision. In C. L. Bajaj, editor, *Algebraic Geometry and Its Applications.* Springer-Verlag, New York, 1993.

[24] Nitzberg, Mumford, and Shiota. *Filtering, Segmentation, and Depth.* Springer-Verlag, New York, 1993.

[25] M. Nitzberg and D. Mumford. The 2.1-d sketch. In *Proceedings of the International Conference on Computer Vision*, pages 138–144. 1990.

[26] D. L. Ringach and R. Shapley. The dynamics of illusory contour integration. *Investigative Ophthalmology and Visual Science*, 35,#2:4196–, 1994.

[27] D. L. Ringach and R. Shapley. Similar mechanisms for illusory contour and amodal completion. *Invest. Ophthalm. and Visual Sci.*, 35,#2:1089–, 1994.

[28] A. Shashua and S. Ullman. Structural saliency: The detection of globally salient structures using a locally connected network. In *Proceedings of the International Conference on Computer Vision*, pages 321–327, 1988.

[29] S. Ullman. Filling in the gaps: The shape of subjective contours and a model for their generation. *Biological Cybernetics*, 25:1–6, 1976.

[30] D. L. Waltz. Understanding Line Drawings of Scenes with Shadows. *The Psychology of Computer Vision*, ed. P. Winston, McGraw-Hill, 1972.

[31] L. R. Williams and A.R. Hanson. Perceptual completion of occluded surfaces. *Proc. of IEEE Computer Vision and Pattern Recognition*, 1994.

[32] L. R. Williams and D.W. Jacobs. Stochastic completion fields: A neural model of illusory contour shape and salience. *Proc. of 5th Intl. Conf. on Comp. Vision*, 1995.