In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 1998, Santa Barbara, CA.

Salient and Multiple Illusory Surfaces *

Davi Geiger
Courant Institute
New York University
New York, NY 10012
geiger@cs.nyu.edu

Hsingkuo Pao
Courant Institute
New York University
New York, NY 10012
hsingkuo@cs.nyu.edu

Nava Rubin
Center for Neural Science
New York University
New York, NY 10012
nava@cns.nyu.edu

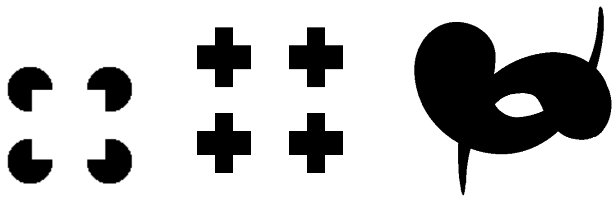
Abstract

All illusory surface figures yield a perception of a surface occluding another one or the background. Occluded surfaces yield completion, a phenomena known as amodal completions. It is intriguing that for some images illusory surfaces are perceived, but not for other images (see figure 1). Also, illusory surfaces may have portions occluded. We aim to understand these phenomena.

Our approach detects intensity edges and junctions. From the junctions we seek to find an optimal image organization, i.e., multiple ordered surfaces with the ordering accounting for salience. The most salient being the figure, while the other surfaces are classified as background. A decision of which surface is the top one, is made locally, at each pixel, allowing the salient surface (figure) to have portions occluded, i.e., with amodal completions. We account for a variety of imagery not explained before.

1. Introduction

When the Kanizsa Square [8] image and the Four Crosses image are shown (see figure 1) various possible visual organizations, figure-ground separations, are plausible. In the Kanizsa Square image the dominant organization is of a vivid (bright white) illusory square in front of four black discs with a white background. In the Four Crosses image the figure of four crosses (or, more precisely, of eight rectangles) in a white background is seen as the dominant organization. Why do we not see the illusory white square in this case ? Why do we see an illusory white square for the Kanizsa square ? We are also intrigued by salient illusory surfaces that exhibit amodal completions (see figure 1). We follow Ringach and Shapley [13] to argue for one model that account for both phenomena, illusory completions and amodal completions. Once an organization is selected, how do we make a decision of what is in front or behind ?



(a) (b) (c)

Figure 1. a. The Kanizsa Square image. b. The Four Crosses image. There is no strong perception of a white square. c. The Two Fish image. In this case the salient surface, either one, has portion of it occluded (amodal completion).

Our model provides a theory of illusory surfaces that explains and interconnects both, a reconstruction of the surfaces and an explanation for when they are or are not perceived. Our approach to find the optimal organization is by (i) detecting edges and junctions ; (ii) assigning, at each junction, a set of hypotheses of the local salient surface configuration and their likelihood; (iii) diffusing these hypotheses (blocked by the edges); and (iv) selecting the best image organization (set of hypotheses); (v) deciding which surface pixels are visible, and on top or occluded.

Our main contributions are

1. Phenomenologically, we offer an explanation for the presence of the vivid illusory square in the Kanisza Square and the absence of the square on the Four Crosses (see figure 1). We also account for amodal completions, even for salient surfaces.

*This work was supported by NSF CAREER award and the Sloan Foundation

2. Theoretically, we argue for an organization selection criteria based on the "entropy" of the diffusion of a set of hypotheses plus a bias to illusory surfaces that have smooth shape, i.e., a bias to remove L-junctions (to see L-junctions as T-junctions.)

3. Visible surfaces include the ones on top and the ones amodally completed. Which surface is on top or occluded is determined at each pixel, locally, based on the local surface probabilities.

Comparison to previous approaches Many computational models have been proposed to describe the formation of illusory contours. The first approach to propose a propagation of edges is due to Ullman[15]. Others include Grossberg & Mingolla[6] Guy & Medioni[5], Kellman & Shipley[9] . Approaches seeking regions are given by Brady and Grimson[2], Nitzberg and Mumford and Shiota [11] , Kumaran, Geiger and Gurvits[10].

The discussion on visual organization dates back at least to the labeling scheme by Huffman [7] . These approaches are too weak to constraint in the sense that they allow for far to many possible organizations (all of them equally likely). Williams and Hanson [16] and Williams and Rubin [17] propose a linear programming method based on local information at junctions, that is partially considered here. However, they do not take into account neither region information nor local properties of the reconstructed illusory contours. Shashua and Ullman [14] do have a line integration/selection approach that address organizational issues, and within the "line extension" ideas, it is the closest work to ours, but ours take a surface reconstruction view of the problem.

Our approach follows previous work of Kumaran, Geiger and Gurvits[10]. However, their model could not distinguish between the Kanizsa Square and the Four Crosses, i.e., could not account for why the illusory square is salient in one figure but not the other. We also note that their minimum cost criteria gaves the four inducers as the most salient surface in the Kanizsa Square figure, and not the illusory square. Our new model resolves this problem and the global minima are the desired salient surface (in agreement with human perception).

Moreover, different from any previous work we account for saliency with amodal completions (see figure 1 c.)

2. Junctions and surface hypotheses

It is well known that when occlusions are detected, an immediate sensation of depth change occurs. Local occlusion cues are provided by T-junctions, Y-junctions, L-junctions (corners) and line endings. Each of these cues could suggest various local occlusion scenarios, e.g., corners have multiple occlusion interpretations (see figure 3). We focus this work on L-junctions (corners).

2.1. Hypotheses: Multiple Surfaces

Say we have detected a total of P junctions, indexed by $p = 1,\ldots,P$. We define a binary parameter $\lambda_k^p = 0,1$ indicating whether a junction p is present or not at pixel k. We are indexing the pixels according to raster scanning convention, i.e., $k = 0,1,\ldots,N^2-1$ where N is the width of the square image. To represent multiple surfaces, say S surfaces, we have each one indexed by $s = 1,\ldots,S$. We define a set of M hypotheses $\{H_a;\ a = 1,\ldots,M\}$, each one assigning a possible configuration to a detected junction (see figure 2). The map $p \to a(p) \in \{1,2,\ldots,M\}$ assigns to each junction p a hypotheses H_a.

In order to represent a junction hypothesis assignment, $a(p)$, at the pixel level we introduce a binary data field $\{\sigma_0^{s,a(p),p};\ s = 1,\ldots,S\}$, where $(\sigma_0^{s,a(p),p})_k = 1$ represents a pixel k assignment to surface s by hypothesis a at junction p.

Note that given a hypothesis (at a junction), a pixel may be assigned to multiple surfaces, to account for overlaps (amodal completions), i.e., we can have $(\sigma_0^{s,a(p),p})_k = 1$ and $(\sigma_0^{s',a(p),p})_k = 1$ for $s \neq s'$ (see figure 3.)

Next step is to obtain a saliency map, a probability of a pixel k to be assigned to a surface s.

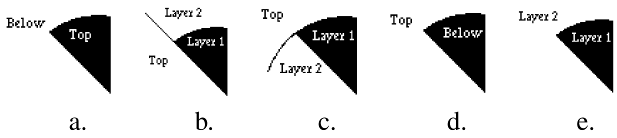


Figure 2. Each L-junction (corner) suggests five (5) possible local salient surface interpretations. Note that (b) and (c) suggest three surface layers, with the hat of the T as the salient one, while (e) removes the junction from the salient surface.

3. Salient Surfaces

Let us assume we are given $\{a(p)\}$ for all p. We now show how to reconstruct multiple surfaces. Our view is to diffuse the hypotheses, as if we are diffusing a set of

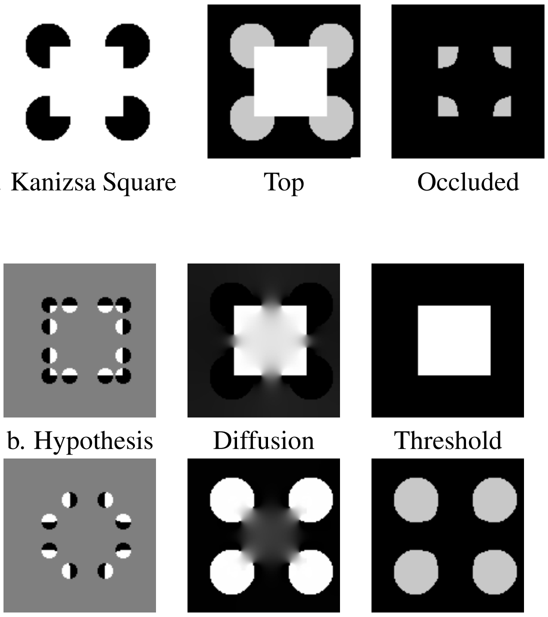


Figure 3. (a) Kanizsa Square image with 128×128 pixels. The optimal organization of the top surfaces (white is salient, grey is second layer, and black is the third layer); and occluded layers (amodal completions). (b) Hypothesis (the winner) consist of two layers of junction hypotheses. For each junction, white (or black) represents $(\sigma_0^{s,a(p),p})_k = 1(or0)$; The result of diffusions. $O = 0.47$, with $\lambda = 0.06$.

probability distributions about surfaces. Let us be more precise.

We define $P_s(k)$ to be a probability of a pixel k to be at surface s. Thus, $P_k^s = 1$ ($P_s(k) = 0$) indicates that the pixel k is with probability one (zero), assigned to a surface s. Note that a pixel may be assigned to multiple surfaces, i.e., we can have $P_s(k) = 1$ and $P_{s'}(k) = 1$ for $s \neq s'$. Our views is that this surface probability will take the hypothesis value (1 or 0) at the junction locations and be diffused from the junctions to every pixel in the lattice. Our aim is to recover the probability $P_s(k)$ for all s and all k.

It is simpler, notation and representation wise, to refer to the image lattice as a continuous space. In this case $\lambda_k^p \to \lambda^p(x,y)$, $P_s(k) \to P_s(x,y)$ and so on. Note that the set of probabilities $\{P_s(k)\}$ become a distribution $P_s(x,y)$. For completeness and clarity we offer both formulations, continuous and discrete. The computations are based on the discrete formulation.

3.1. Fitting the surfaces to the hypotheses

We require P_s to fit one of the hypotheses $\sigma_0^{s,a(p),p}$ at each junction p, i.e.,

$$\lambda_k^p(\sigma_0^{s,a(p),p})_k = \lambda_k^p P_s(k) \qquad \forall k, p, s\,,$$

or, more flexibly, we can assign an error function

$$Error(\{P_s\}\,|\,\{a(p)\}) = \sum_{k=0}^{N^2-1} \sum_{p=1}^{P} [\lambda_k^p(\sigma_0^{s,a(p),p}(k) - P_s(k))^2]\,,$$

or in the continuous

$$Error = ||\sum_{p=1}^{P} \lambda^p(x,y)(\sigma_0^{s,a(p),p}(x,y) - P_s(x,y))^2||\,,$$

where we have considered the distance measure $||v(x,y)|| = \int\int\, v(x,y)\, dxdy$, assuming $v(x,y) \geq 0$.

3.2. Smoothness assumption

In order to diffuse the information, the probabilities P_s, from the sparse junction locations we consider the cost

$$Smooth(\{P_s\}) = \sum_{k=0}^{N^2-1} \sum_{j \in N_k} [\mu_{k,j}(P_s(k) - P_s(j))^2]\,,$$

where $N_k = \{k+1, k-1, k-N, k+N\}$ is the four neighborhood set of pixel k, and the diffusion coefficients

$$\mu_{k,j} = \mu\,(1 - e_{k,j})\,,$$

are zero when there is an edge between pixels k and j, $e_{kj} = 1$, and μ in the absense of it, $e_{kj} = 0$. The edge detection is assumed to take place at the input level. The smoothing/diffusion coefficients $\mu_{jk} = \mu\,(1 - e_{jk})$ prevent smoothing along intensity boundaries. In the continuous, this cost gives

$$Smooth = ||\mu(x,y)\,(\nabla P_s(x,y))^2||\,,$$

where $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})$ is the gradient operator.

3.3. Diffusion from sparse data

The final cost function, for each surface s and set of hypotheses $\{a(p); p = 1, \ldots P\}$, is then the sum of both criteria, error to fit the hypotheses and smoothness of the saliency map, i.e.,

$$E_s(\{P_s\}\,|\,\{a(p)\}) = Error + Smooth\,. \qquad (1)$$

This energy has its minimum $P_s(x,y)$ satisfying the Euler- Lagrange equation

$$\sum_{p=1}^{P} \lambda^p (\sigma_0^{s,a(p),p} - P_s)(x,y) + \nabla.[\mu(x,y)\,\nabla P_s(x,y)] = 0\,. \tag{2}$$

This is analogous to a Fokker-Plank equation for probability distributions in statistical physics [12] . The optimal solution can be obtained by the Green's function method, and yields

$$P_s(x,y) =$$

$$\int\!\!\int G(x-x',y-y') \sum_{p=1}^{P} \lambda^p(x',y')\,\sigma_0^{s,a(p),p}(x',y')\; dx'\,dy'\,,$$

where $G(x,y)$ is the Green's function, i.e., it is essentially the inverse of the operator induced by equation 2. This is a linear operation on the data (a convolution), depending on the coefficients $\mu(x,y)$. In the discrete case equation 2 has the linear form

$$\mathbf{D}\,P_s = \lambda^p\,\sigma_0^{s,a(p),p}\,, \tag{3}$$

where the $N^2 \times N^2$ diagonal matrix λ^p has entries λ_k^p. The symmetric and band limited matrix $\mathbf{D}$ has five non zero diagonals. More precisely, $\mathbf{D}$ has the following structure: (i) Diagonal elements $D_{kk} = (\sum_{p=1}^{P} \lambda_k^p) + \mu_{k,k-N} + \mu_{k,k+N} + \mu_{k,k-1} + \mu_{k,k+1}$; (ii) There are four off-diagonals with non zero elements. Along row k, $D_{k,k-1} = -\mu_{k,k-1}$, $D_{k,k+1} = -\mu_{k,k+1}$, $D_{k,k+N} = -\mu_{k,k+N}$, and $D_{k,k-N} = -\mu_{k,k-N}$, representing the west, east, south, and north first neighbor interactions, respectively.

Determining the Green's function is equivalent to inverting the matrix $\mathbf{D}$. To solve equation (3) we use Cholesky decomposition, since it is well suited given that the matrix $\mathbf{D}$ is a block tridiagonal, symmetric, positive matrix. On a 2-dimensional lattice with N^2 sites this method takes $O(N^4)$ operations. Results are shown in Figures 3, 4, 8, 6.

The optimal solution $P_s^*(k)$ is bounded by the maximum and minimun values of $\sigma_0^{s,a(p),p}$, i.e., by 0 and 1. This property allows us to interpret the optimal solution $P_s^*(k)$ as the probabilities of pixel k being or not assigned to the surface s.

3.4. Illusory surfaces and their boundaries (contours)

Given the probabilities $\{P_s(k)\}$ we devise a pixel decision where if $P_s(k) \geq 0.5$ we say pixel k is at surface s and if $P_s(k) < 0.5$ it is not. We can then find the boundaries of a surface s (illusory contours) as the locations where the probability changes from below 0.5 to above it (see results in figures 3).

4. Visual Organization

We have devised a scheme to reconstruct multiple surfaces. However, we have assumed an organization to be provided, i.e., $a(p)$ was prefixed. We now discuss the main theme of this paper, how to select the optimal $a(p)$. Our criteria will be based on the result of the reconstructions P_s.

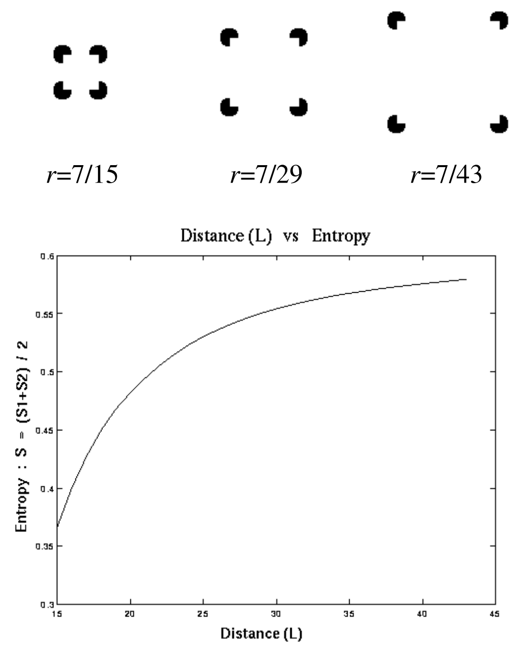


Figure 4. A sequence of Kanizsa Square images with different ratios, $r = R/L$, where R is the radius of the circles, and L is the distance between inducers. The graph of the entropy S versus r for the Kanizsa square. We crudely estimated that the change in perception from the organization where the square is on top to the one where the four inducers on top occurs for $r = 7/20 = 0.35$.

4.1. Prior distribution for the hypotheses

It is plausible that each junction-hypothesis configuration $(\sigma_0^{a(p),p})_k$ has a different probability to occur. Salient surfaces, including illusory surfaces, encountered in nature tend to have smooth shapes. We conjecture such a bias for smooth surface shapes when selecting figure from ground. This bias will prefer to see an L-junction (corner) as a T-junction, thus eliminating an infinity curvature (corner) hypotheses. We propose a cost criteria

$$C_1(\{a(p)\}) = -\#(L - junction \rightarrow T - junction)\,.$$

The more L-junctions are "eliminated", hypothesized as T-junctions, the more likely is the organization.

We note that a similar measure could be provided by the total number of L-junctions of the final thresholded solution, for all surfaces. The less the number of L-junctions the more likely is the organization to be perceived. In this case, we could write the criteria as

$$C(\{a(p)\}) = \sum_{s=1}^{S} \#L - junctions(P_s)\,.$$

The main computational disadvantage of the second criteria is that it requires the result of the reconstruction. However, we can not rule out C, since there is one main difference between the two criteria: L-Junctions that are created only due to the reconstruction are only evaluated by C. These are very rare situations, but figure 5 suggest that C may be more appropriate. C favor organizations yielding surfaces with shapes that have less "turns" . We have adopted in this paper the criteria C and not C_1. A somewhat similar criteria is offered in [17], based on a contour driven view of the problem.

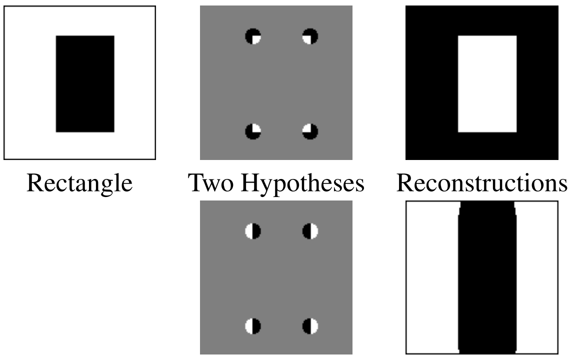


Figure 5. The Rectangle image with 128×128 pixels. The perceived organization is of a rectangle in front of a background. The hypothesis that eliminates all $L - junctions$ would see two curtains covering a much longer rectangle. However, other $L - junctions$ are created from the reconstruction, the ones that are formed with the frame of the image.

4.2. Entropy of the visible surfaces

The entropy criteria, per pixel, for each hypothesis map $\{a(p), p = 1, ..., P\}$, is given by

$$S(\{a(p)\}) = -\frac{1}{(\sum_{s=1}^{S} N_s^*)} \sum_{s=1}^{S} \sum_{k=1}^{N^2} P_s^*(k) log P_s^*(k)\,,$$

where $P_s^*(k) = P_s(k)$ if $P_s(k) \geq 0.5$, but $P_s^*(k) = 0$ otherwise. Only the region that can be visible contributes to the entropy. Then, N_s^* are the number of pixels where surface s can be visible, i.e., where $P_s(k) \geq 0.5$. This criteria encourages "sharp" diffusions for the possible visible surface, i.e., it bias for $P_s(k)$ that are closer to 1.

To identify the need for this criteria we consider figure 4, a sequence of Kanizsa square images with its inducers varying the separation distance. As the separation increases the perception of the square diminishes and fade away. The junctions are unchanged and so this affect is only due to the entropy criteria.

4.3. Organization Criteria

Thus, our final organization criteria is the sum of both, entropy per pixel and bias towards smooth shapes, i.e.,

$$O(a(p)) = \lambda C(\{a(p)\}) - \frac{1}{(\sum_{s=1}^{S} N_s^*)} \sum_{s=1}^{S} \sum_{k=1}^{N^2} P_s^*(k) log P_s^*(k)\,.$$

The parameter λ balances both criteria. To estimate λ we can use the experiments shown in figure 4 for the Kanizsa square image. As the ratio of the radius of the circles (inducers) and length of the square (distance between inducers) gets smaller, the perceived square, hypothesis 1 (H1), gets weak and eventually the illusory figure fade away yielding the four inducers figure, hypothesis 2 (H2). During this process, as the ratio of the radius of the circles and length of the square gets smaller, the entropy S gets larger, while $C(a(p))$ stays constant (see figure 4). For H1 four L-junctions are counted while the entropy varies as the distance between inducers vary according to the graph shown in figure 4. For H2, twelve (12) L-junctions are present, while the entropy is zero (all $P_s(k) = 0, 1$). Thus, at the configuration where the perception changes from H1 to H2, we have

$$4\lambda + S^* = 12\lambda \rightarrow \lambda = S^*/8\,,$$

where S^* is the entropy of the configuration at the transition point (perception wise). We have crudely estimated $\lambda = 0.06$ and adopted in all our experiments.

We have studied a variety of illusory figures. We report results that focus on the Kanizsa Square (see figures 3 and

4) and the Four Crosses (see figures 7 and 8). The organization criteria explains why the illusory square is perceived in the Kanisza square image while not well perceived in the Four Crosses one.

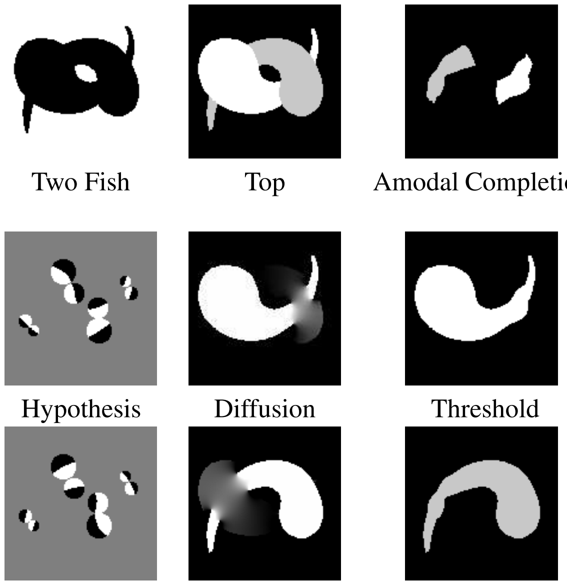


Figure 6. The Two Fish image and its results. In this case pixels where the salient surface is visible (probabilities above 0.5) may still be amodal completed because other surfaces may have locally higher probabilities (come on top).

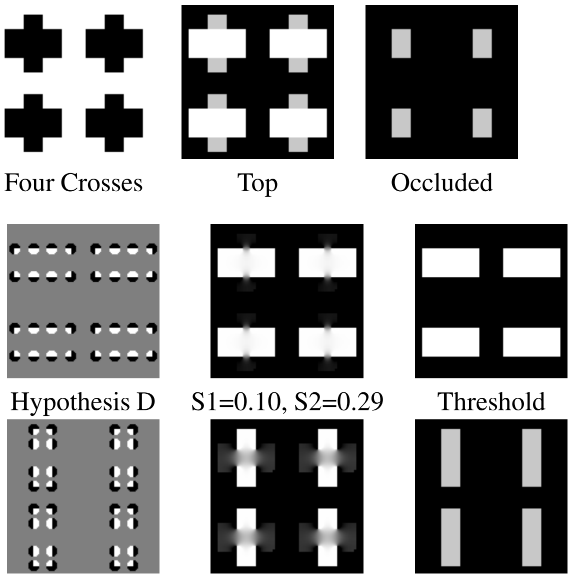


Figure 7. The Four Crosses image, 128×128 pixels; Top surfaces for the winner hypothesis with brightness representing saliency; and Occluded surfaces (amodal completions). The winner hypothesis, D. Note that the larger (and not smaller) rectangles are salient because of the lower entropy.

5. Surfaces: Salient, visible, top and amodal completions

Given a set of winner hypotheses $\{a^*(p), p = 1, ..., P\}$ we obtain $\{P_s(k)\}$ for each surface s. We have argued that the surface s with largest entropy per pixel is considered the salient one. For example, in the Four Crosses image the rectangle with largest area was selected as the salient one (see figure 7). Still, are salient surfaces the same as top surfaces ? For the Two Fish images (see figure 6) we have the larger fish as salient, but yet, portions of it are amodal completed (not on top). Let us now account for this phenomena with the machinery we have just built.

We have defined a pixel k to belong to a visible surface s whenever $P_s(k) > 0.5$. The decision for a visible surface s to be or not salient depends on the entropy per pixel. To decide for each pixel k which surface is perceived, we invoke the order of the local probabilities $P_s(k)$. We order surfaces locally according to $P_s(k)$. More precisely, surface s is on the top of surface s' at pixel k if $P_s(k) > P_{s'}(k)$. Thus, we define top surface s at pixel k if

$$P_s(k) > P_{s'}(k) \qquad \forall s' \neq s\,.$$

Amodal completions are then defined at pixel k as surfaces s at pixels k where $P_s(k) > 0.5$ and they are not on top. It is clear that a visible surface, where $P_s(k) > 0.5$, can be amodal completed or on top. We examine for the Two Fish images (see figure 6) these distinctions: Saliency, visible, top and amodal completions.

References

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

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

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

[4] 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.

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

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

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

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

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

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

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

[12] G. Parisi. Statistical Field Theory. *Addison-Wesley*, 1988.

[13] D. L. Ringach and R. Shapley. Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Research*, 36, 3037–3050, 1996.

[14] A. Shashua and S. Ullman. Structural saliency: The detection of globally salient structures using a locally connected network. In *ICCV*, pages 321–327, 1988.

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

[16] L. R.Williams and A. Hanson. Perceptual completion of occluded surfaces. *CVPR*, 1994.

[17] L. R. Williams and N. Rubin, A surface-completion model explains Kanizsa's crosses demo and predicts a novel perceptual effect. *Invest. Ophth. and Vis. Sci. Suppl.* 39, 3959.

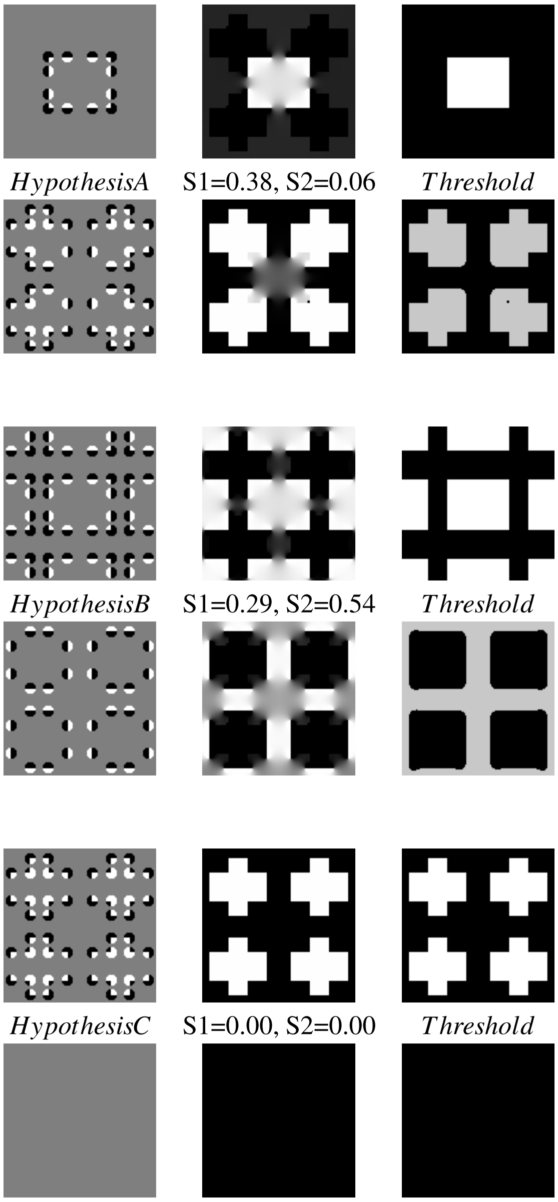


Figure 8. The Four Crosses image's hypotheses, A, B, and C; Hypothesis D is shown in figure 7. There are (4+36=40), (32+0=32), (48+0=48), and (16+16=32) corners (L-junctions) at hypothesis A, B, C and D respectively. The organization criteria $O(a(p))$ are 2.63, 2.34, 2.88, and 2.12 respectively, making the winner to be hypothesis D. The second voted hypothesis is for the four crosses on top, while the illusory square, hypothesis A, has a lower score.