Proc Int'l Conf Image Processing, 1999
Invited

The Shape of Illusory Figures *

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

Krishnan Kumaran
Bell Labs, Lucent Technologies
700 Mountain Ave
Murray Hill, NJ 07974
kumaran@lucent.com

Hsing-Kuo 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

We have been developing a stochastic model for figure-ground separation[9][3][12]. The model selects/constructs the foreground with preference for figures with "more convex" shapes. When these models are applied to illusory figures ([7]) they yield perceptually accurate selection of figure and background. The approach is based on an "entropy" measure of a region diffusion Markov model from a set of local figure/ground hypothesis. The contour boundaries are implicitly represented, via the thresholding of the diffusion result.

What optimal properties do the illusory contours satisfies ? We show that the entropy criteria selects contours such as to minimize a Taylor series of the even derivatives with respect to the length of the contour. The coefficients are positive and they get exponentially smaller as the derivatives increase. The zeroth order term suggest that small length contours are preferred, the second order terms suggests that curvature-like term is minimized (with less strength compared to the zero order one), and higher order derivatives give additional contour smoothness constraints.

1 Introduction

The selection of salient illusory surfaces such as those shown in Figure 3 ([7]) reveals important properties of the human visual system. Two problems are being solved, the computation of an illusory boundary and the decision of what is a figure and what is background. One may put together both problems by asking to each pixel if it belongs to the foreground or to the background. The detection of the boundary and the *border ownership* problem are then automatic: the contour and its border ownership is given by the set of pixels assigned to the figure (or foreground) with a neighboring pixels assigned to the background.

We present the stochastic model [] that respond at each pixel the question: does it belong to the foreground or background ? The model is based on local (pixel-to-pixel) computations in a stochastic two-dimensional network. The model can be readily adapted to work on richer representations, such as wavelets. In this case the network would set interactions between wavelet coefficients. Effectively, the diffusion process group intensity edges to produce closed regions, without any restriction on the number of regions and with topological freedom for admitting any number of holes in each region. The preference for global shape properties (specifically, convexity and larger regions) is realized by an "entropy" criteria on the output of the diffusion-like process.

Essentially this is all there is, an "entropy" criteria to select figure/ground. One still may ask for a characterization of the selection criteria in terms of the boundary contour. In particular, what optimal criteria does the contours satisfies that is equivalent to the entropy one ? Contours tend to fit the intensity edges, but what properties they satisfy at the illusory area, where no intensity edges exist ?

We show that the entropy criteria selects contours such as to minimize a Taylor series of the even derivatives with respect to the length of the contour. The coeffi-

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

0-7803-5467-2/99/ $ 10.00 © 1999 IEEE

cients are positive and they get exponentially smaller as the derivatives increase. The zeroth order term suggest that small length contours are preferred, the second order terms suggests that curvature-like term is minimized (with less strength compared to the zero order one), and higher order derivatives give additional contour smoothness constraints.

1.1 Background

The discussion to characterize the shape of contours start with Ullman [15] and a more elaborated models are developed by Parent and Zucker [13] and by Mumford [10], the so called elastica. Our approach does not minimize shape contour properties and is based on a region model. Regions models start with Brady and Grimson [1] and later Nitzberg and Mumford[11]. In particular we follow the models of Kumaran et al. [9], Geiger et al.[3] and Pao et al. [12]. These region models have not develop a characterization of the contours they select. This is our effort here.

1.2 Organization

We first, on section 2 review the stochastic model we have been developing for addressing the figure-ground problem. We show how an "entropy" criteria to select a set of local figure-ground hypothesis is in agreement with human perception for this class of illusory figures. Then, on section 3, we show how the entropy criteria can be recast as a criteria on the shapes of the contours.

2 Figure-Ground and Entropy: Overview

In images the partition of the image in regions is not known a priori and this problem is best illustrated with illusory figures (see Figures 3. Moreover, given a partition of the image into regions one still needs to ask which ones are foreground regions and which ones are background regions. These two problems, separating a regions into images and deciding which ones are foreground constitutecd st the segmentation problem.

To approach this problem we follow our previous work: Kumaran et al. [9], Geiger et al. [3] Pao et al [12].

A set of sparse features, so called *inducers*, such as corners, T-junctions, end lines, and all available intensity edges.

2.1 Naming the Variables

The input image is defined in a discrete lattice of size N by N.

$$\mathbf{I}^2 \equiv \{i = 1,...,N^2\} \quad \text{is the set of pixels.}$$

Let us indicate a shape-figure $\mathbf{S}$ of size S pixels in the image either by

$$\mathbf{S} \equiv \{s = 1,...,S\} \quad \text{region}$$

with area $\mathcal{A}(\mathbf{S}) = \#\mathbf{S} = S$ pixels, or simply by

$$C(\mathbf{S}) \quad \text{closed boundary}.$$

Image wise, closed boundary like to "sit" on intensity edges, but boundary points may not have an intensity edge. We then define the edge-perimeter of $\mathbf{S}$ as the size of the boundary set weighted by the normalized intensity edges, i.e.,

$$\mathcal{P}(\mathbf{S},I) = \sum_{k \in C(\mathbf{S})} e(k) \quad \text{edge-perimeter},$$

where $e(k) \in [0,1]$ is the intensity edge (magnitude of intensity gradient) at pixel k, normalized by the largest intensity edge.

2.2 Variational Model

Let us define $P(k)$ to be the value at pixel k of a steady state solution of the diffusion inside and outside $\mathbf{S}$. We create the diffusion process by formulating it as a variational problem.

Local Hypothesis and Data Fitting: At the inducers we assign the source value $\sigma_0 = 1$ for foreground and $\sigma_0 = -1$ for background, both at $\mathcal{P}(\mathbf{S},I)$ (see Figure 1b.) At every inducer/source a local choice of figure/background is thus required. These are the local hypothesis and an entropy measure will select the best hypothesis (best set of local hypothesis).

We also consider a "decay" process outside the inducers. One way to present it is to push non-source pixels k to a non-commitment between figure and ground, i.e., to push (or decay) $P(k)$ to be zero ("neutral").

This idea can be implemented in the variational approach by minimizing an energy with respect to the set of variables $P(k)$, i.e.,

$$E(P) = \sum_{k=1}^{N^2} \lambda_k (P(k) - \sigma_0(k))^2, \qquad (1)$$

where if pixel k is not a source we have $\lambda_k = \nu(1 - e(k))$ and $\sigma_0(k) = 0$; For source pixels we have $\lambda_k = e(k)$ and $\sigma_0(k) = \pm 1$. A perhaps good criteria to decide which pixels are sources or not, is based on comparing $e(k)$ to $1 - e(k)$ whichever is larger, i.e., if $e(k) \geq 0.5$ then it is a source pixel. In our experiments $e(k) = 0, 1$ so the decision was easy to make.

Smoothness: In order to obtain a diffusion process, from a minimization stand point, we insert a smoothness constraint on $P(k)$. A simple one minimizes the square of the

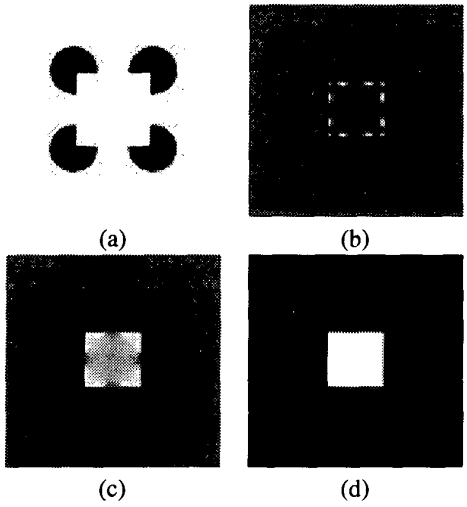

Figure 1. (a) The Kanizsa square and (b), (c), (d) its hypothesis set, diffussion and threshold. In (b), we use white and black color to indicate the foreground ($\sigma_0 = +1$) and background ($\sigma_0 = -1$) hypothesis respectively. The gray level indicates the neutral hypothesis ($\sigma_0 = 0$).

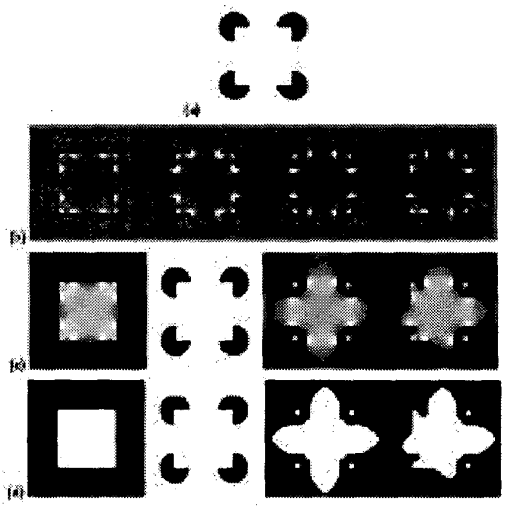

Figure 2. (a) The Kanizsa square and the (d) shape solution, (c) diffusion, for (b) different organizations. The entropy criteria chooses the Kaniza square. How can this formulated in terms of contour properties is addressed in this paper.

length of the gradient vector $(\frac{\partial P(x,y)}{\partial x}, \frac{\partial P(s,y)}{\partial y})$, or in the discrete setting we write

$$Smooth(\{P\}) = \mu \sum_{k=1}^{N^2} \sum_{j \in N_k} (1 - e(k,j)) (P(k) - P(j))^2,$$

where $N_k = \{k+1, k-1, k-N, k+N\}$ is the subset of the four neighbors of pixel k that are either inside the object shape **S** or outside it and $e(k,j)$ is the magnitude of the intensity change from pixel k to pixel j, normalized to the largest $e(k,j)$. Note that we can define $e(k) = \max_{j \in N_k} e(k,j)$.

Energy Model: With the smoothing criteria the total cost function becomes

$$\begin{aligned} E(P) &= \sum_{k=1}^{N^2} [\lambda_k (P(k) - \sigma_0(k))^2 \\ &\quad + \sum_{j \in N_k} \mu (1 - e(k,j))(P(k) - P(j))^2]. \quad (2) \end{aligned}$$

Thus, the optimal solution $P^*(k)$ balances fitting the local hypothesis and smoothing. It is clear (e.g., [9]) that $P^*(k)$ is bounded by the maximum and minimum values of σ_0, i.e., ± 1. This gives a diffusion property to this process.

Closed Contour and Level Sets: In order to obtain the figure **S** we consider all pixels $k \in \mathbf{I}$ such that $P^*(k) \geq 0$. The background pixels are obtained as pixels $k \in \mathbf{I}$ such that $P^*(k) < 0$. The level set $k \in \mathbf{I}$ such that $P^*(k) = 0$, represent the closed contours $C(\mathbf{S})$. The result of Figure 1 d. (and [9], [3]) suggest that the shapes obtained are "roughly" in agreement to perception.

2.3 Shape-Figure and Entropy

In order to select the hypothesis that produce the "best figure" (best shape) we consider an entropy measure. After all, given a source, we do not know which side takes $\sigma_0(k) = 1$ or -1. At junctions the multiplicity of hypothesis grows (see [9], [3]). We first, for simplicity, convert $-1 \leq P^*(k) \leq 1$ into a probability distribution at each pixel, via the linear map

$$p(k) = \frac{1}{2}(1 + P^*(k)).$$

Thus, the entropy criteria becomes

$$S = -\frac{1}{N^2} \sum_{k \in \mathbf{I}^2} p(k) log p(k) + (1 - p(k)) log(1 - p(k)).$$

The sharper is the diffusion, the closer to 1 is $P^*(k)$ (and $p(k)$) inside **S**, the better is the figure perception, i.e., the

lower the entropy the more salient is the region. Note that the entropy is a per pixel entropy or the total entropy normalized by the number of pixels. In such model convex regions have lower entropy than concave ones [12].

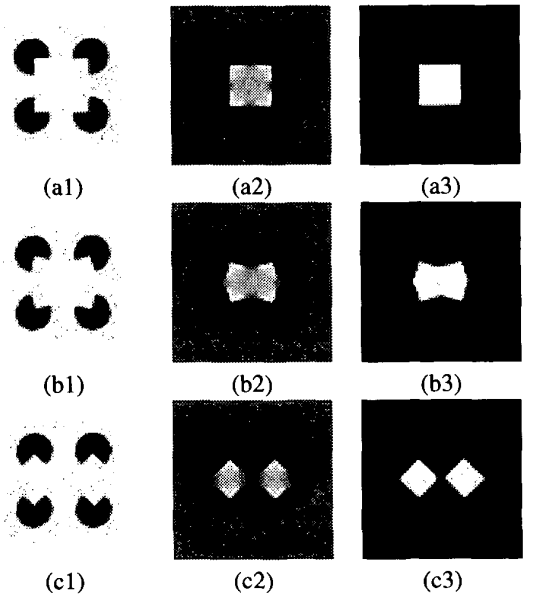

Figure 3. (a1), (b1) and (c1) The 'Kanizsa Square' with pac men of different orientations. (a2), (b2) and (c2) are the diffusion result. (a3), (b3) and (c3) are the threshold. The shape of objects changed when the orientation is changed. Moreover, when the change is abrupt, the whole meaning could also be changed. In (c3), the object is broken into two parts, Perceptually, we said we have two objects other than one.

3 Entropy and Contour Criteria

To see how the entropy criteria relates to contour criteria such as length, curvature and etc., we rewrite the entropy as follows.

$$S = -\sum_{p\geq 0}^{1} N(p)[plogp + (1-p)log(1-p)] \quad (3)$$

where $N(p)$ is the number of image locations bearing the reconstructed value p, i.e. it is the histogram or marginal distribution function for the p values. To better understand the relation to properties of contours, we extend the above formulation to a continuous 2-D image. On a continuum, the entropy function would take the form

$$S = -\int_0^1 n(p)[plogp + (1-p)log(1-p)]dp, \quad (4)$$

where the function $n(p)$, which is now an area/length per value p, is potentially discontinuous at the values $p = 0 or 1$, since the diffusion is blocked at intensity boundaries. However, since the illusory edges are obtained by thresholding the reconstruction $\mathbf{p}$ at the value 0.5, $n(p)$ is the total perimeter of the continuous ISO-contours contained entirely in the smooth image domains. It can hence be Taylor-expanded about this value as follows.

$$\begin{aligned} S &= \int_0^1 (\sum_{k=0}^{\infty} \frac{(p-0.5)^k}{k!} \frac{d^k n(p)}{dp^k}|_{p=0.5}) \\ &\quad plogp + (1-p)log(1-p)dp \\ &= \sum_{k=0}^{\infty} C_{2k} \frac{d^{2k} n(p)}{dp^{2k}}|_{p=0.5} \end{aligned} \quad (5)$$

where

$$C_{2k} = \frac{-1}{(2k)!} \int_0^1 (p-0.5)^{2k} plogp + (1-p)log(1-p)dp$$

are the set of coefficients independent of contour shape information. The odd coefficients have vanished. The even coefficients are all positive and diminish exponentially. The latter property makes the expansion convergent if the derivatives are not too badly behaved, as is likely to be the case in a diffusion process. This form can now be compared with the contour models.

The zero order term gives a term proportional to the length of the contour

$$S_0 = C_0 n(p)|_{p=0.5}.$$

The second term (and second order term) is given by

$$S_2 = C_2 \frac{d^2 n(p)}{dp^2}|_{p=0.5}.$$

To better understand this term we write the density function $n(p)$ as

$$n(p) = \int_{image} \delta(p(x,y) - p)dxdy$$

where $\delta()$ is the Dirac-delta function. Then we have

$$\frac{dn(p)}{dp} = \int_{image} \delta'(p(x,y) - p)dxdy$$

where the $'$ indicates derivative with respect to the argument. We now rewrite this derivative using the product rule as

$$\delta_I(p(x,y)-p)=\frac{\nabla p(x,y)v\cdot\nabla}{|\nabla p(x,y)|^2}\delta(p(x,y)-p)$$

and similarly

$$\delta_{II}(p(x,y)-p)=\frac{\nabla p(x,y)\cdot\nabla}{|\nabla p(x,y)|^2}\{\frac{\nabla p(x,y)\cdot\nabla}{|\nabla p(x,y)|^2}\delta(p(x,y)-p)\}.$$

This expression can then be used to evaluate the above integral by parts, which would lead to higher derivative contributions evaluated at the 0.5 ISO-contours.

Thus, the optimization of the entropy function might be performed in an approximate manner by truncating the Taylor expansion and optimizing the resulting cost functional of the curve parameters. Note, however, that this renders a highly non-linear problem that would be cumbersome to solve even if only a small number of leading terms are retained. This complexity is entirely avoided in our surface model while the reconstructions obtained are at least of the quality of the most sophisticated contour models.

References

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

[2] J. Elder and S. W. Zucker. A measure of Closure. *Vision Research*, Vol. 34 (24), pp.3361-3369, 1994.

[3] D. Geiger, H. Pao and N. Rubin. Salient and multiple illusory surfaces. *Computer Vision and Pattern recognition.*, June. 1998.

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

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

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

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

[8] B. Kimia, A. Tannenbaum, S. Zucker, "Shapes, Shocks, and Deformations I: The components of two-dimensional shape and the reaction-diffusion space", *Int. J. Comp. Vis.***1**: 189-224, 1995.

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

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

[11] M. Nitzberg and D. Mumford. The 2.1-d sketch. In *ICCV*, pages 138–144. 1990.

[12] H. K. Pao, D. Geiger and N. Rubin, "Measuring Convexity for Figure Ground-Separation", *International Conference on Computer Vision*, Sep., 1999.

[13] S. Parent and S. W. Zucker, "Trace inference, curvature consistency and curve detection", *IEEE PAMI*, Vol. 11, No. 8, pp. 823-839, 1989.

[14] 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.

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