Scott McCloskey COMP 766 Paper Critique
Lance Williams and David Jacobs, "Stochastic Completion Fields: A Neural Model of Illusory Contour Shape and Salience"

1 The Big Picture

In this paper the authors address the problem of illusory contours, and describe an algorithm for generating such contours. They describe a Markov process for the generation of random walks, each step of which relates to a cell within the visual cortex. The claim is that, when a Monte Carlo simulation of such walks is performed, the path formed by the most frequently-visited states corresponds to the contour most likely to be imagined by a human observer. They present several examples of the method on several images designed to illicit illusory contours.

2 The Gory Details

In the introduction, the authors outline the problem of shape completion and saliency. Their underlying point is that, like many problems in computer vision, it is under constrained; we try to complete shapes that could have arbitrary properties in the occluded regions. Consequentially, given any algorithm that attempts to solve the problem, it is possible to construct a counter-example. Since the human visual system can be fooled in the same way, the real problem is to develop a method that models the likely human response.

With this goal in mind - and absent a conclusive model of human vision in this area - it is necessary to make some assumptions. The authors propose that the statistics of the distribution of potential completions are computed by the human visual system when presented with illusory stimuli. This is certainly a debatable assertion, but it gives rise to the method that they describe.

Specifically, they model a potential completion shape by the stochastic motion of a particle. The particle's state at time t is given by (x_t, y_t, θ_t), its position and orientation. At time $t+1$, the particle moves to state $(x_t + cos\theta_t, y_t + sin\theta_t, \theta + \theta')$ where θ' is a zero-mean, normally distributed random value with some variance (a parameter).

As a sort of justification for their assumption, the authors note the correspondence between their particle's state and neurons in the primate visual cortex. Due to the elliptic shape of these neurons, their activity measures both the position and orientation of light in the plane. The activity of the collection of such neurons, then, can be modeled as a probability density function over position and orientation.

After a few pages of mathematical justification, the authors describe the way that they generate the probability density function. Given a initial position and orientation, the probability density function is generated by a Monte Carlo simulation.

For each (x, y, θ) point, we have two distributions. The $sourcefield$ of that point describes the likelihood that a walk starting there will visit a different state. The $sinkfield$ of that point describes the likelihood that a walk ending there will have passed through a different state. Given a source state A and sink state B, then, we can multiply these fields to determine the likelihood that a given state is visited on a walk from A to B.

Through several examples, the authors show that correctly placed sources and sinks in some images can be shown to generate probability density functions that correspond to the imagined contours. The examples are rather convincing, and look very much like the edges that I imagined from the original stimuli.

3 The Scott Opinion

One of the most interesting points presented in the article is one that isn't highlighted very well. At the end of page 5 and beginning of page 6, they suggest that the variance of the distribution indicates the sharpness of the imagined contour. When presenting their examples in section 5, they don't stress this point, and don't explain what this suggests.

The larger question, in my mind, is how you define the sharpness of an edge that isn't there. For example you can't rely on the contrast, which is 0. If viewers were asked to describe the sharpness of the edges that they see, then the best that you could hope for is that those descriptions correlate with the distribution. If this is what the authors mean then they should say as much.

While the use of the (x, y, θ) representation is explicitly motivated by cortical neurons, the authors ignore the issue of scale. It is understandable that they don't normalize the image so that the grid spacings correspond to the size of the neurons, since that would require knowing the viewing distance. It is strange, though, that they don't account for the fact that these neurons come in more then one size. It might be interesting to look at the results in a coarser grid, and how results at multiple scales could be combined to give a better result. The use of multiple scales would almost certainly have an impact on the distribution about the mode, which gets back to the previous point.

The authors do not discuss how these probability density functions behave in images where people do not imagine lines. If each of the pacmen in figure 9 were rotated, it seems clear that the non-circular probabilities (the ones connecting the corners of opposite pacmen) would drop as a function of the angle. At some point, however, the rotation becomes large enough that viewers no longer see the connecting lines. Do the probability values at that angle represent a threshold? If so, does that threshold work for all images, or does each stimulus have its own?

It also seems strange that, at the last stage of the algorithm, orientations are discarded. The final images that they present in the examples have intensity values related to the sum of the probabilities of all orientations at that point. If a position's probabilities were low for all orientations except for two that differed by 90^o, it's unclear that the sum is meaningful. The odds of this taking place are limited by the fact that orientation changes are drawn from a normal distribution, but the authors do not address the point.

The biggest hurdle to using this method in an automated system is the lack of a single method to places sources and sinks. Each of the examples presented in the paper comes with a different rationale for the placement of sources and sinks in such a way as to get the desired result. What's more is that they don't present cases that illustrate what happens when sources and sinks are positioned improperly. If such a case resulted in probability density functions with low magnitudes, we might be able to get away with placing a large number of each in the hopes that a few of them would work out. The authors make no claim to having an automated method to places the sources and sinks, which is fine. They mention that research into the location of L-, T-, Y- and X-junctions may produce such a method, though it seems less then certain.

On the first pass through the paper, the presentation of the mathematical model seems superfluous, as it's never used in any sort of implementation. In the interest of comparing the authors' methods to those referenced in the prior work, though, this treatment is indispensable. Moreover, the proof in the appendix provides the reader with a concrete connection to some of this previous work.