Extracting Salient Curves from Images: An Analysis of the Saliency Network

T. D. Alter

MIT AI Laboratory
545 Technology Square
Cambridge, MA 02139, USA

Ronen Basri

Department of Applied Math.
The Weizmann Inst. of Science
Rehovot, 76100, Israel

Abstract

The Saliency Network proposed by Shashua and Ullman is a well-known approach to the problem of extracting salient curves from images while performing gap completion. This paper analyzes the Saliency Network. Although the network is attractive for a number reasons, our analysis reveals certain weaknesses with the method. In particular, we show cases in which the most salient element does not lie on the perceptually most salient curve. Furthermore, the saliency measure may change its preferences when curves are scaled uniformly. Also, for certain fragmented curves the measure prefers large gaps over a few small gaps of the same total size. We analyze the time complexity required by the method and discuss problems due to coarse sampling of the range of possible orientations. We show that with proper sampling the complexity of the network becomes cubic in the size of the network. Finally, we consider the possibility of using the Saliency Network for grouping. We show that the Saliency Network recovers the most salient curve efficiently, but it has problems with identifying any salient curve other than the most salient one.

1 Introduction

In line drawings, certain shapes attract our attention more than others. For example, these shapes may be the ones that are smooth, long, and closed (see for example Fig. 1). Shashua and Ullman [20] proposed a method, which attracted considerable attention, to extract such shapes from a line drawing. They defined a function that evaluates the "saliency" of a curve. Their function has the following properties. First, when all other parameters are held constant, it monotonically increases with the length of the evaluated curve. In addition, it decreases monotonically with the *energy* (the total squared curvature) of the curve. Thirdly, the function evaluates fragmented curves, in which case it penalizes according to the amount of fragmentation. Finally, the saliency measure is the sum of contributions from different sections of the curve, where these contributions decay with the sections' accumulated energy and gap length from the beginning of the curve. Using this saliency function, Shashua and Ullman defined the "saliency map" of an image to be an image in which the intensity value of each pixel is proportional to the score of the most salient curve emanating from that pixel.

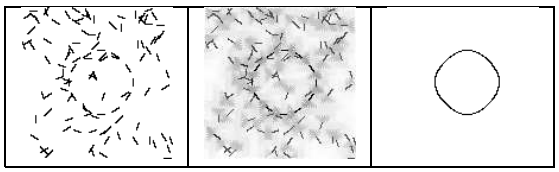

Figure 1: A fragmented circle in the middle of noise. The global shape of the circle is apparent. From left to right: input image, saliency map, and the most salient curve found.

A network of locally connected elements (the Saliency Network) was proposed for computing the saliency map. The Saliency Network's computation involves local interactions between image locations, and its size is proportional to the size of the image. The network implements a relaxation process that optimizes the saliency measure. As a consequence of the optimization, the network can identify the most salient location in the image, which could lie on either an open or closed curve. Additionally, the method attempts to fill in gaps smoothly while simultaneously overcoming noise.

The problem of marking salient locations in images is addressed differently in [8]. In addition, [21, 22] extend Shashua and Ullman's method to find skeletons of regions. Identifying salient structures in images is one of the objectives of *perceptual grouping*. By perceptual grouping, we refer to the (bottom-up) process of grouping together structures in the image that are likely to belong to a single object. Other tasks in perceptual grouping are image segmentation and gap completion. For instance, [10, 12, 13, 14, 15, 17, 18, 25, 26] extract contours from the image according to certain optimization criteria, [2, 4, 11, 19, 23] compute optimal curves for filling in gaps, and [3, 6, 7, 9, 16, 27, 24] identify occluded and subjective contours.

In this paper we provide an analysis of Shashua and Ullman's method. We examine both the measure of saliency and the computational performance of

the Saliency Network. Motivated by both perceptual and computational reasons we identify three properties which we believe a measure of saliency should satisfy. We then analyze Shashua and Ullman's measure with respect to these properties. The properties are:

Fidelity. The saliency map should highlight the locations in the image that lie on perceptually salient curves. In particular, the most salient location in the saliency map should lie on the most perceptually salient curve rather than on any of the surrounding line segments.

Invariance. In different images objects often appear in different positions and orientations or in different sizes. A saliency measure for curves should be insensitive to such variations. In particular, the measure should be invariant to 2D rigid transformations of the curve. In addition, the measure should be consistent over different scales. That is, given two curves Γ_1 and Γ_2, if Γ_1 is considered more salient than Γ_2, then Γ_1 should remain more salient when the curves are scaled uniformly.

Performance on gaps. In Fig. 1 as the size of gaps between edge elements is increased, our perception of the circle fades. We therefore expect the measure of saliency to degrade with gaps. Motivated by psychophysical studies [5], we require a saliency measure also to penalize large gaps more than few small gaps of the same total size.

The Saliency Network is an efficient and elegant method, well suited to locating salient structures in images. However, we found cases in which the network violates each of the above three properties. On the issue of fidelity, the network indeed locates the perceptually salient curves, so that long, smooth, closed curves are preferred over short, wiggly, open ones. Nonetheless, our analysis reveals cases in which the most salient location in the saliency map is not on the perceptually most salient curve. For example, if there are short line segments touching a salient curve, then often the short segments shall be judged more salient than the closed curve. In this situation, the most salient location in the network will not lie on the closed curve, but it will draw its saliency from the closed curve.

Since the saliency measure depends only on length and curvature, it is invariant to rigid transformations. We show, however, that at times the measure changes its preferences when the curves are scaled uniformly. For instance, consider a straight line and a circle of the same length. For lengths less than a certain value, the line is preferred over the circle, whereas for larger lengths this preference reverses.

Finally, the saliency measure can be applied to fragmented curves, in which case it will attenuate with gap length. However, our analysis indicates that, when circles of both the same size and gap length are compared, the measure prefers a circle with one long gap over a circle with few small gaps of the same total size.

In addition to studying properties of the saliency measure, we also examine the computational properties of the Saliency Network. In particular, we analyze the convergence rate of the network and show that the run-time complexity of the network in serial implementations is quadratic in the number of elements. We then discuss problems due to coarse sampling of the range of possible orientations. We show that, when this range is sampled too coarsely, undesirable effects may occur in which corners are preferred over straight lines. With proper sampling the complexity of the network becomes cubic in the size of the image.

Finally, we consider the possibility of using the Saliency Network for grouping. We note that, in contrast to other existing methods for grouping that search over the exponentially large space of all possible image curves (e.g., [10, 12, 17]), the Saliency Network recovers the most salient curve in time complexity that is polynomial in the size of the image. However, the network must take a single choice at every junction, and as a consequence has problems with identifying salient curves other than the most salient one.

The paper proceeds as follows. Section 2 contains definitions. Section 3 includes an analysis of the different properties of the saliency measure. Section 4 analyzes the time complexity of the network computation. Section 5 analyzes the effects of sampling on the computation. Finally, Section 6 discusses the issue of using the output of the network for grouping.

2 Definitions

Shashua and Ullman defined their saliency measure as follows. For every pixel in the image, there is a fixed set of "orientation elements" connecting the pixel to neighboring pixels. Each orientation element is called "actual" if it lies on an edge in the underlying image, and otherwise it is called "virtual" or "gap." Given a curve Γ composed of the $N+1$ orientation elements $p_i, p_{i+1}, ..., p_{i+N}$ the saliency of Γ is defined by

$$\Phi(\Gamma) = \sum_{j=i}^{i+N} \sigma_j \, \rho_{ij} \, C_{ij} \,, \tag{1}$$

with

$$\sigma_j = \begin{cases} 1, & \text{if } p_j \text{ is actual} \\ 0, & \text{if } p_j \text{ is virtual} \end{cases} \quad \text{and } \rho_{ij} = \prod_{k=i}^{j} \rho_k = \rho^{g_{ij}} \,,$$

where $\rho_{ii} = 1$ and where $\rho = 0.7$.[1] σ_j ensures that only actual elements will contribute to the saliency measure. g_{ij} is the number of gap elements between p_i and p_j, and ρ_{ij} reduces the contribution of an element according to the total length of the gaps up to that element. Further,

$$C_{ij} = e^{-K_{ij}}, \quad \text{with} \quad K_{ij} = \int_{p_i}^{p_j} \kappa^2(s) ds,$$

[1] The formula for ρ_{ij} appeared in [20] as $\rho_{ij} = \prod_{k=i+1}^{j} \rho_k$, but the computation actually performed by the network (which is given by Eq. 4) implements the modified formula given here.

where $\kappa(s)$ is the curvature at position s. K_{ij} reduces the contribution of elements according to the accumulated squared curvature from the beginning of the curve.

The saliency $\Phi(i)$ of an element p_i is defined to be the maximum saliency over all curves emanating from p_i. Shashua and Ullman showed how to compute $\Phi(i)$ on a network of locally connected elements. Denote by $\Phi_N(i)$ the saliency of the most salient curve of length $N+1$ or less emanating from p_i. Then,

$$\Phi_N(i) = \max_{p_j \in \mathcal{N}(i)} F(i, j, \Phi_{N-1}(j)), \qquad (2)$$

where $\mathcal{N}(i)$ is the set of all neighboring elements of p_i, and where $F()$ is a function of $\Phi_{N-1}()$ and constants stored at elements p_i and p_j. Shashua and Ullman referred to this type of measure as "extensible."[2] In the Saliency Network,

$$F(i, j, \Phi_{N-1}(j)) = \sigma_i + \rho_i C_{ij} \Phi_{N-1}(j), \qquad (3)$$

which gives

$$\Phi_N(i) = \sigma_i + \rho_i \max_{p_j \in \mathcal{N}(i)} C_{ij} \Phi_{N-1}(j). \qquad (4)$$

Note that because the saliency measure in Eq. 1 is monotonically non-decreasing with the number of elements N on a curve $\Phi_N(i)$ in fact represents the saliency of the most salient curve of length $\leq N$ emanating from p_i.

To make the saliency measure Φ independent of the particular implementation, we introduce a continuous version of Φ. Given a curve $\Gamma(s)$ of length l ($0 \leq s \leq l$, s denotes arc length), we define Φ by

$$\Phi(\Gamma) = \int_0^l \sigma(s)\rho(0,s)C(0,s)ds, \qquad (5)$$

$$\text{where} \quad \sigma(s) = \begin{cases} 1, & \text{if } \Gamma(s) \text{ is actual} \\ 0, & \text{if } \Gamma(s) \text{ is virtual} \end{cases}$$

$$\rho(s_1, s_2) = \rho^{g(s_1,s_2)}$$

$$C(s_1, s_2) = e^{-K(s_1,s_2)},$$

where $g(s_1, s_2)$ is the total gap length of Γ between s_1 and s_2 and $K(s_1, s_2)$ is the energy of the curve between s_1 and s_2, which are defined by

$$g(s_1, s_2) = \int_{s_1}^{s_2} (1 - \sigma(t))dt, \quad K(s_1, s_2) = \int_{s_1}^{s_2} \kappa^2(t)dt.$$

A useful tool in computing saliencies is the following rule. Given a curve Γ which is composed of two smoothly concatenated sections, Γ_1 and Γ_2, the saliency of Γ is given by

$$\Phi(\Gamma) = \Phi(\Gamma_1) + \rho^{g(\Gamma_1)} e^{-K(\Gamma_1)} \Phi(\Gamma_2), \qquad (6)$$

where $g(\Gamma_1)$ is the total gap length and $K(\Gamma_1)$ is the energy of Γ_1.

[2]Note that this definition of extensibility is different from that used by Brady et al. [2].

Over the remainder of the paper, we will present examples of the Saliency Network on simulated and real images. Our implementation replicates Shashua and Ullman's original implementation, except that we increased the number of orientation elements per pixel to obtain greater accuracy. We used twenty-four orientation elements per pixel, whereas Shashua and Ullman used sixteen elements per pixel. Also we set $\rho = .7$ as in the original implementation.

3 Properties of the Saliency Measure

We begin our analysis by examining the saliency measure proposed by Shashua and Ullman. Section 3.1 below discusses the treatment of cycles. Section 3.2 analyzes the behavior of the measure when applied to simple curves. Lastly, Section 3.3 analyzes the behavior of the measure when applied to curves that include gaps.

3.1 Cycles

The measure of saliency proposed by Shashua and Ullman is a positive function that increases monotonically with the lengths of the curves in the image. Closed curves (cycles) are considered to have infinite length, even though they form finite structures in the image. Shashua and Ullman showed that their network is guaranteed to converge when applied to closed curves. The reason it converges is that the contribution to the saliency from remote elements attenuates geometrically with the curvature accumulated from the beginning of the curve. In cycles this generates a geometric series that converges to a finite value.

Formally, given a closed curve Γ, denote by Φ the saliency of an element of Γ that is obtained by starting from that element and then proceeding once around the curve. Denote by K the total squared curvature of the cycle and by g the cycle's total gap length. Then by repeatedly applying Eq. 6 we obtain

$$\Phi(\Gamma) = \Phi + \rho^g e^{-K}\Phi + ... = \frac{\Phi}{1 - \rho^g e^{-K}}. \qquad (7)$$

When the network is applied to an open curve, after going once along the curve it is possible for the network to then take a 180° turn and walk back along the curve. The saliency of the resulting closed curve could be considered to be the saliency of the open curve at convergence. In [1] we show that the attenuation due to the 180° is so high that the additional score is negligible.

3.2 Straight lines and circles

In this section we compute the saliencies of a few simple curves. We then use these simple curves to examine the issues of fidelity and invariance. In general, we will only be interested in the measure of saliency obtained for the most salient element of the curve. Throughout this section we shall use the continuous definition of the saliency measure (Eq. 5). We consider only curves with no gaps (we will analyze curves with gaps in Section 3.3); hence $\sigma(s) = 1$ and $\rho(0, s) = 1$ for all s. Eq. 5 therefore becomes

$$\Phi(\Gamma) = \int_0^l C(0,s)ds, \text{ where } C = e^{-\int_0^s \kappa^2(t)dt}. \qquad (8)$$

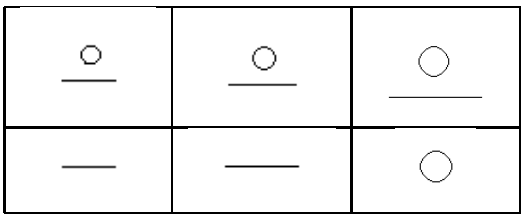

Figure 2: Lack of scale invariance in the Saliency Network. Top: three images that contain a straight line and a circle of roughly the same length. Bottom: the most salient curves that were found in these images. Lengths are 27, 39, and 84 . The saliency values obtained for the circles are 15.39, 33.60, and 132.05, and for the lines are 27.06, 39.00, and 84.00.

We begin by deriving explicit formulas for the saliency of straight lines and curves. For straight lines $C(0,s) = 1$ for all s. Therefore, ignoring the possibility that a line can wrap around itself (see Section 3.1), a straight line of length l will obtain the score $\Phi(\Gamma) = l$. The saliency of a straight line, therefore, grows linearly with the length of the line.

For a circle of radius r, the curvature is constant, $\kappa = 1/r$, and so for a circular arc of length s,

$$C(0,s) = e^{-\int_0^s \frac{1}{r^2} dt} = e^{-\frac{s}{r^2}}. \tag{9}$$

The saliency attributed for the circular arc is

$$\Phi(0,s) = \int_0^s C(0,t)dt = r^2 \left(1 - e^{-\frac{s}{r^2}}\right). \tag{10}$$

At convergence the saliency of the circle is given by

$$\Phi(\Gamma) = \lim_{s \to \infty} r^2 \left(1 - e^{-\frac{s}{r^2}}\right) = r^2. \tag{11}$$

The score of a circle, therefore, grows quadratically with the radius (and thus also with the perimeter) of the circle.

The fact that the saliency of a straight line grows linearly with its length, whereas the saliency of a circle grows quadratically with its perimeter, suggests that the network may treat the two differently when they are scaled. Consider a straight line of length l and a circle of perimeter $l = 2\pi r$. These two entities will have exactly the same saliency when $l = 0$ and when $l = 4\pi^2 \approx 39.48$. (The saliencies in the two cases are 0 and $4\pi^2$, respectively.) When $0 \leq l \leq 4\pi^2$ the line will be more salient than the circle, whereas when $l > 4\pi^2$ the circle will be more salient. Fig. 2 shows an example of three images, each of which contains a straight line and a circle of the same length. Consistent with our analysis, the Saliency Network found the straight line to be more salient than the circle at shorter lengths, and found the circle to be more salient at longer lengths.

A different problem is encountered in the case of a circle connected to short line segments. Consider the picture in Fig. 3-left. The circle seems to be the most perceptually salient curve in this image. Counterintuitively, the most salient element computed by the network is on one of the line segments connected to the circle, thus violating the fidelity requirement. The reason is that a neighboring line segment may increase its saliency by connecting to the circle, without affecting the saliency of the circle. Consider, for example, a circular arc of length 1 and curvature κ connected smoothly to a circle of radius r (which corresponds to a single element connected smoothly to the circle via curvature κ). Using Eq. 6 we obtain that the saliency of the first element on the arc is

$$\Phi_e = \Phi(\Gamma) + e^{-\kappa^2} \Phi_c, \tag{12}$$

where Γ represents the circular arc and Φ_c is the saliency of the circle. Now, using Eq. 9,

$$\Phi(\Gamma) = \int_0^1 C(0,s)ds = \frac{1 - e^{-\kappa^2}}{\kappa^2}. \tag{13}$$

Combining Eqs. 11-13, we obtain that

$$\Phi_e = \frac{1 - e^{-\kappa^2}}{\kappa^2} + e^{-\kappa^2} r^2. \tag{14}$$

If we now compare the saliency of the element, Φ_e, to that of the circle $\Phi_c = r^2$ (Eq. 11), we obtain that $\Phi_e > \Phi_c$ when

$$|\kappa| < \kappa_c, \tag{15}$$

where $\kappa_c = 1/r$. That is, the element will be more salient than the circle if and only if it connects to the circle at a curvature that is less than the curvature of the circle. This is consistent with the network's preference for straight curves. Notice that if the element is a line tangential to the curve ($\kappa = 0$) the element will be more salient than the circle regardless of the radius of the circle. This phenomenon, that curves connecting to a circle may increase their saliencies due to these connections and actually beat the circle, is more likely to occur for longer curves [1]. An example to this problem is given in Fig. 3.

3.3 Curves with gaps

One of the most important properties of Shashua and Ullman's saliency network is its ability to fill in gaps while computing the saliencies. The network handles gaps by using virtual elements, which compute the saliencies of curves emanating from their locations and transfer these saliencies to their neighboring elements. Via these transfers, actual elements evaluate the saliencies of curves that emanate from their locations and contain any number of gaps. The network avoids curves with large gaps by attenuating the scores of curves exponentially with gap size.

In [1] we analyze the performance of the saliency network in the presence of gaps. Due to the saliency measure attenuating exponentially with gap size, the network is capable of overcoming small gaps, but is unlikely to overcome large ones. As an example, consider the problem mentioned in Section 3.2, that a

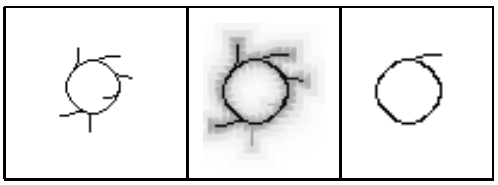

Figure 3: An example of a circle with a few short curves connecting to it. The most salient element (for which $\Phi = 136.63$) was not on the circle, although its saliency came mostly from the circle (the saliency of the circle is 130.74). If short gaps were added between the curves and the circle, the circle would become the most salient curve in the image. From left to right: input image, saliency map, and the most salient curve.

p	0.5	0.7	0.9	0.95	0.99	1
Φ	2.80	6.54	25.23	53.27	277.56	∞

Table 1: Φ for a straight infinite line with uniform gaps as a function of p ($\rho = 0.7$). Note that the score for infinite lines gives an upper bound for the score of finite ones.

short line segment in the neighborhood of a circle may increase its saliency by connecting to the circle. One consequence of the fast attenuation is that this problem almost disappears when the segment is not physically connected to the circle. On the other hand, we show that, due to the exponential decay, very long structures (straight lines and circles) obtain very low scores even when only a small fraction of the curves are gaps. Examples for the maximal saliency values obtained for straight lines and circles with gaps are given in Tables 1 and 2.

Finally, we explore the question of whether the network prefers fragmented curves (dashed lines) over curves with single gaps of the same total size. At first glance Shashua and Ullman's saliency measure appears indifferent to this property, because the total size of gaps is taken into account, irrespective of the fragmentation. In fact, for open curves there is no clear preference between a curve having many small gaps or a few long gaps. For closed curves, however, we show that a curve with a single large gap is preferred over the same curve with several small gaps of the same total size; this preference is inconsistent with the psychophysical experiments of Elder and Zucker [5].

Figure 4: Three circles of the same radius with the same total gap size. Using Shashua and Ullman's network the saliency values are 46.85 (left), 27.93 (middle), and 23.27 (right).

$p \setminus r$	1	2	4	8	16
0.5	0.42	1.17	2.07	2.58	2.74
0.9	0.87	3.15	9.17	17.55	22.74
0.95	0.93	3.55	11.82	28.39	43.70
0.99	0.99	3.90	14.98	51.58	132.48
1	1	4	16	64	256

Table 2: The saliency values of circles with uniform gaps as a function of p and r (for $\rho = 0.7$).

4 Complexity and Convergence

In this section we analyze the complexity of Shashua and Ullman's saliency network. Denote the total number of pixels in the image by p and the number of discrete orientation elements at every pixel by b. The network has pb elements. At each iteration every element has to evaluate all the saliencies obtained from elements connected to it. The complexity of each iteration therefore is pb^2. The question then is how many iterations are required before the network converges. Clearly, if we did not allow cycles the longest curve may be of length p, and so the total complexity of the computation would be at most p^2b^2. But when cycles are considered, we show below that the network converges in a linear number of iterations, and so the total complexity is indeed $O(p^2b^2)$.

Given a cycle Γ, denote by Φ_n the score obtained after going n times around the cycle, by K the energy of Γ, and by g the total gap size. Then from Eq. 7 the saliency of Γ is $\Phi = \Phi_1/(1 - \rho^g e^{-K})$. After going n times around the cycle, the accumulated score becomes (this is the finite sum of the geometric series in Eq. 7)

$$\Phi_n = \frac{1 - \rho^{ng} e^{-nK}}{1 - \rho^g e^{-K}} \Phi_1 . \tag{16}$$

Define the relative error by

$$E = \frac{\Phi - \Phi_n}{\Phi} = \rho^{ng} e^{-nK} . \tag{17}$$

We can now compute the number of cycles, n, needed to achieve an $E = \epsilon$ error:

$$\ln \epsilon = n(g \ln \rho - K) \quad \Longrightarrow \quad n = \frac{\ln \epsilon}{g \ln \rho - K} . \tag{18}$$

Assume Γ is a circle of radius r with no gaps. Then $K = \frac{2\pi}{r}$ and

$$n = -\frac{r \ln \epsilon}{2\pi} . \tag{19}$$

The number of cycles around the circle is $O(r)$. Assuming one iteration covers one unit of arc length, the number of iterations for each cycle is $2\pi r$. Thus, from Eq. 19 the total number of iterations needed to achieve an ϵ error is

$$N = 2\pi r n = -r^2 \ln \epsilon . \tag{20}$$

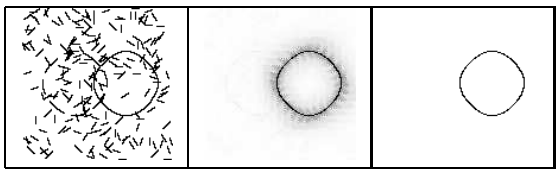

Figure 5: Running the Saliency Network on an image with gap-free and fragmented circles and a background of 200 random line segments. After 200 iterations, the maximum saliency was 130.8. The time to convergence and the maximum saliency are independent of the number of background elements. From left to right: input image, saliency map, and the most salient curve.

Consequently, the total number of iterations required is $O(r^2) = O(p)$, where p is the size of the image. As an example, the number of cycles required to achieve 1% error ($\epsilon = 0.01$, $\ln \epsilon \approx -4.6$) is $n \approx 2.3\, r/\pi \approx 0.7r$, and therefore the number of iterations is $N \approx 4.6r^2$.

By taking the maximal possible circle in the image, we account for the worst case complexity of the network. This is because any larger closed curve must accumulate comparable energy in order not to exceed the boundaries of the image. We can therefore conclude that the worst case complexity of the network is $O(p^2b^2)$, which is the squared number of elements in the network.

Fig. 5 shows an image of a gap-free and a fragmented circle on a noisy background. As expected, the Saliency Network chooses the gap-free circle as the most salient curve. Using Eq. 20, we could predict that the number of iterations for the network to converge on the gap-free circle is 200.

5 Discrete Implementations

Our analysis of Shashua and Ullman's method has concentrated on the theoretical, continuous version of their saliency measure. Shashua and Ullman proposed to compute this measure using a network of finitely many, locally connected elements. In this section we analyze the effect of computing the saliency measure on discrete networks. We show in particular that the network is extremely sensitive to the number of discrete orientation elements allocated per pixel.

Consider for example a nearly horizontal straight line segment. Due to aliasing, the line may be cut in the middle so that one part of the line is raised up by one pixel. This will induce two corners to the line. Consequently, we obtain the paradoxical result that two lines meeting at a corner may be more salient than a nearly straight continuation. Hence straight lines oriented such that they deviate slightly from horizontal will often be less salient than corners.

The discretization problem is carried over to other, more complicated examples. Consider a circle of radius r. When r is sufficiently small, the circle can be approximated by a regular polygon where each side includes a single orientation element. Let K be the total squared curvature corresponding to a turn $\alpha = 2\pi/n$, where n is the number of sides of the polygon. The discrete saliency of such a regular polygon is given by

$$\Phi \;=\; \frac{1}{1-e^{-K}} \;=\; \frac{1}{1-e^{-\frac{\Delta s}{r^2}}}, \qquad (21)$$

where Δs is the arc length of the circle that is covered in one iteration. Returning again to the gap-free circle in Fig. 5, for this circle $r = 11.39$ and $\Delta s = 2.983$ (see above), and so under discretization its saliency is 132.1. When $\Delta s/r^2$ is small,

$$\frac{1}{1-e^{-\frac{\Delta s}{r^2}}} \;\approx\; r^2, \qquad (22)$$

The approximation in this equation improves as r increases; this happens when the number of sides in the polygonal approximation increases and as a result fits a circle more closely. When r is big so that a good approximation by a regular polygon would require finer orientation changes (less than $2\pi/b$), a faithful discretization of the circle would involve many inflections (that is, clockwise turns balanced by counter-clockwise turns). These inflections would be penalized unduly by the network.

We could improve the saliency of the discretization if we instead represent the circle by a regular polygon with b sides; each side now contains more than one element. This, however, will still not result in a reasonable approximation to the continuous saliency of the circle. This can be seen by the following observation. Eq. 21 gives the saliency of a regular polygon with b sides, each of unit length, in terms of K, the total squared curvature assigned for a turn of $\frac{2\pi}{b}$. The saliency of a similar regular polygon in which every side is of length l is given by

$$\Phi_l = \frac{l}{1-e^{-K}} = l\Phi_1 . \qquad (23)$$

The saliency of a regular b-sided polygon, therefore, increases linearly with the length of each side, l. Since l is directly related to the radius of the circumscribed circle, the saliency of the polygon also increases linearly with the radius of that circle. Since the continuous saliency of a circle grows quadratically with the radius of the circle (Eq. 11), we obtain that, as r grows, the saliency of the polygon will considerably underestimate the saliency of the circle.

The results shown in this section establish that the Saliency Network faces serious difficulties due to discretization of the range of orientations. A faithful implementation of the continuous saliency measure would require a very fine discretization. The number of orientation elements needed to completely avoid the problems mentioned in this section is of the order of $\sqrt{p}$, where p is the total number of pixels in the image. With this number of orientation elements the overall time complexity of the network (see Section 4) becomes $O(p^2b^2) = O(p^3)$.

6 Applications to Grouping

The Saliency Network is viewed by many people not only as a mechanism for shifting attention to salient

structures, but also as a method for the initial grouping of curves. The problems of identifying salient structures and the grouping of curves are not identical. Saliency can be viewed as the problem of identifying the "odd man out," whereas grouping is the problem of identifying image structures that are likely to belong to a single object. The criteria of length and straightness can separate a smooth object from a background of short, broken curves (e.g., a disc on a background of grass), but they may be inappropriate for segmenting equally-smooth objects in cluttered scenes, since long smooth curves often will traverse a few objects. For example, Fig. 6 shows a case where the Saliency Network successively finds a curve that belongs to an object of interest (top images), and another case where the most salient curve traverses more than one object (bottom). Nevertheless, in many cases the salient curves may lie on objects of interest, and so may be useful for grouping.

In [1] we show that because the saliency measure is extensible if two different objects are touching, then always the best curve through one of the objects will merge into the other; this situation is illustrated by the real image example in Fig. 6, where the boundary curves of two objects (a flashlight and a telephone) merge together. Also due to extensiblity elements next to the most salient curve will tend to become salient themselves. This creates a problem when two objects are touching since to obtain the contour of the least salient object one has to trace the contour from the least salient element on the two objects. Finally, we discuss problems occur when the network is run for a small number of iterations. We show that in this case closure is not encouraged by the network. In addition, we show that, although at convergence the most salient curve can be traced with a simple procedure, the tracing procedure is not guaranteed to extract the best curve.

7 Conclusion

The Saliency Network is a mechanism for identifying salient curves in images based on length and straightness. The method is attractive for several reasons. First, the measure of saliency generally prefers long and smooth curves over short or wiggly ones. In addition, the network is guaranteed to find the most salient structure according to the measure. While so doing, the network fills in gaps with smooth completions and tolerates noise. The network itself is locally connected and its size is proportional to the size of the image. The locality is further emphasized since the contribution of remote elements to the score of a given element attenuates with the curvature and gap length separating the remote elements from the given element.

Our analysis revealed, however, certain weaknesses with the method. We found cases in which the most salient element does not lie on the perceptually most salient curve. Furthermore, we showed cases in which the saliency measure changes its preferences when curves are scaled uniformly. Finally, we found that for certain fragmented curves the measure prefers large gaps over a few small gaps of the same total size.

We believe that the weaknesses of the Saliency Network are due largely to two important properties of the saliency measure which are imposed by the Network's computation. The two properties are (1) extensibility and (2) geometric convergence for cycles. Extensibility implies that an optimal curve must be composed of sub-curves that are themselves optimal. Due to extensibility, saliencies can be computed efficiently using a procedure of recursive optimization (dynamic programming). One of the benefits of extensibility is that, although the Saliency Network finds the element from which the best curve emanates rather than extracting the best curve itself, the best curve can be extracted through a simple tracing procedure. Also due to extensibility, however, the method has difficulties at junctions; this leaves unclear how one could use the method for grouping in cluttered images.

The second property exhibited by the saliency measure is that the measure decreases in a geometric series when evaluated along a cycle. This property, which is essential for convergence, was used in this paper to compute the network's time complexity. In particular, we showed that the number of iterations is linear in the size of the image, and as a consequence the overall complexity in serial implementations is quadratic in the size of the network. This complexity result is based on the assumption that the number of discrete orientations per pixel is independent of the size of the image. On the other hand, we also showed the network's rankings of curves can be significantly altered when the range of possible orientations is coarsely sampled. With proper sampling, however, the complexity of the network becomes cubic in the size of the image.

In sum, extensibility and geometric convergence enable the saliency measure to be optimized and the optimal curves to be recovered efficiently (in polynomial time), but at the same time they restrict the set of possible functions that can be used as measures of saliency. Partly due to this restriction, the chosen measure has certain properties which counter that which we believe is expected in a measure of saliency. It remains to be seen whether variations of the current measure can be defined that remedy some of its weaknesses while still allowing the saliency map and most salient curves to be computed efficiently.

Code availability

We have made available our C-code implementation of the Saliency Network. To retrieve the code, ftp to "ftp.ai.mit.edu," then log in as "anonymous," then cd to "pub/users/tda/," and then get and uncompress "susal-code.tar.Z."

Acknowledgment

We thank David Jacobs for many fruitful discussions.

References

[1] Alter, T. D., and Basri, R, "Extracting salient curves from images: an analysis of the Saliency Network," *MIT AI Memo 1150*, 1995.

[2] Brady, M., W. E. L. Grimson, and D. J. Langridge, "Shape Encoding and Subjective Contours," *Proc. First Annual Conf. Artif. Int.*, 1980.

[3] Brady, M., and W. E. L. Grimson, "The Perception of Subjective Surfaces," *MIT A. I. Memo No. 666*, November 1981.

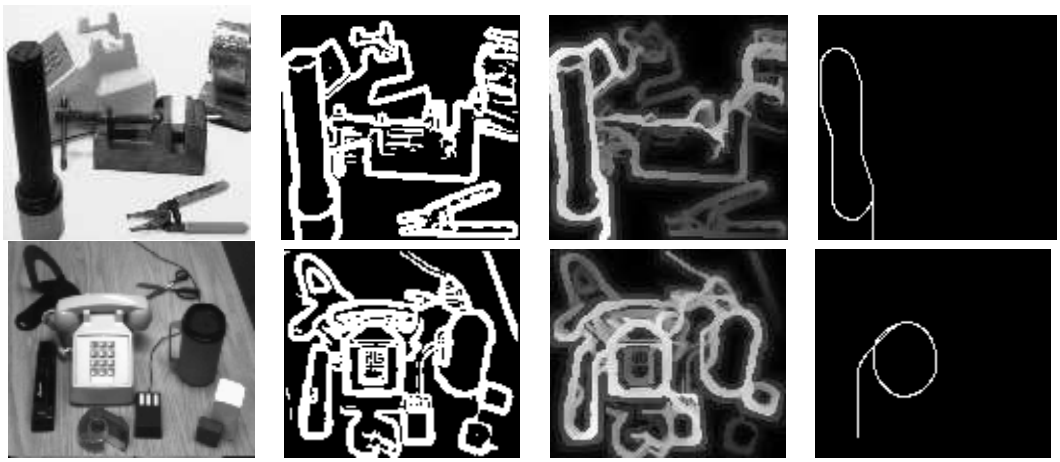

Figure 6: Shashua and Ullman's saliency map for a cluttered scene. Scene images (on the left) were smoothed and then the gradient magnitude was thresholded to get a binary image (second pictures from the left). These edge images were the input to the network. The third pictures display the saliency map, and the fourth show the curve emanating from the most salient element, for which $\Phi = 263.5$ and 210.1.

[4] Bruckstein, A. M., and A. N. Netravali, "On Minimal Energy Trajectories," *CVGIP*, 49:283–296, 1990.

[5] Elder, J., and S. Zucker, "The Integration of Figure Fragments into Representations of Planar Shape," McGill Res. Centre for Int. Machines TR-93-2, 1993.

[6] Finkel, L. H., and P. Sajda, "Object Discrimination Based on Depth-from-Occlusion," *Neural Computation*, 4:901–921, 1992.

[7] Grossberg, S., and M. Mingolla, "Neural Dynamics of Surface Perception: Boundary Webs, Illuminants, and Shape-from-Shading," *CVGIP*, 37:116–165, 1987.

[8] Guy, G., and G. Medioni, "Inferring Global Perceptual Contours from Local Features," *Proc. DARPA IUW*, Washington, D. C.:881–892, 1993.

[9] Heitger, F., and Rüdiger von der Heydt, "A Computational Model of Neural Contour Processing: Figure-Ground Segregation and Illusory Contours," *Proc. of ICCV*, pp. 32–40, 1993.

[10] Herault, L., and R. Horaud, "Figure-Ground Discrimination: A Combinatorial Optimization Approach," *IEEE Trans. on PAMI*, 15(9), 1993.

[11] Horn, B. K. P., "The Curve of Least Energy," *ACM Trans. on Math. Soft.*, 9(4):441–460, 1983.

[12] Jacobs, D. W., "Recognizing 3-D Objects Using 2-D Images," MIT AI Memo 1416, 1993.

[13] Martelli, A., "An Application of Heuristic Search Methods to Edge and Contour Detection," *Comm. ACM*, 19(2), February 1976.

[14] Mohan, R., and R. Nevatia, "Perceptual Grouping for Scene Description and Segmentation," *Trans. on PAMI*, 14(6), 1992.

[15] Montanari, U., "On the Optimal Detection of Curves in Noisy Pictures," *Comm. ACM*, 14(5), 1971.

[16] Mumford, D., "Elastica and Computer Vision," *Algebraic Geometry and Its Applications*, Chandrajit Bajaj (ed.), Springer Verlag:491–506, 1994.

[17] Parent, P., and S. W. Zucker. "Trace Inference, Curvature Consistency, and Curve Detection," *Trans. on PAMI*, 11(8), 1989.

[18] Pavlidis, T., and Y.-T. Liow, "Integrating Region Growing and Edge Detection," *Trans. on PAMI*, 12(3):225–233, 1990.

[19] Rutkowski, W. S., "Shape Completion," *CVGIP*, 9:89–101, 1979.

[20] Sha'ashua, A., and S. Ullman, "Structural Saliency: The Detection of Globally Salient Structures Using a Locally Connected Network," *Proc. of ICCV*:321–327, 1988.

[21] Subirana-Vilanova, J., "Curved Inertia Frames: Visual Attention and Perceptual Organization Using Convexity and Symmetry," *MIT AI Memo 1137*, Dec. 1991.

[22] Subirana-Vilanova, J., and K. K. Sung, "Multi-Scale Vector-Ridge-Detection for Perceptual Organization Without Edges," *MIT AI Memo 1318*, 1992.

[23] Ullman, S., "Filling-In the Gaps: The Shape of Subjective Contours and a Model for Their Generation," *Biological Cyber.*, 25:1–6, 1976.

[24] Webb, J. A., and E. Pervin, "The Shape of Subjective Contours," *Proc. National Conf. on Art. Int.*: 340–343, Aug. 1984.

[25] Weiss, I., "3D Shape Representation by Contours," *CGVIP*, 41, pp. 80–100, 1988.

[26] Williams, L. R., and A. R. Hanson, "Perceptual Completion of Occluded Surfaces," *Proc. of CVPR*, WA, 1994.

[27] Williams, L. R., and D. W. Jacobs, "Stochastic Completion Fields: A Neural Model of Illusory Contour Shape and Salience," *Proc. of ICCV*:408–415, 1995.