Completion Energies and Scale

Eitan Sharon, Achi Brandt*, Ronen Basri†
Dept. of Applied Math
The Weizmann Inst. of Science
Rehovot, 76100, Israel

November 6, 1997

Abstract

The detection of smooth curves in images and their completion over gaps are two important problems in perceptual grouping. In this study we examine the notion of completion energy of curve elements, showing and exploiting its intrinsic dependence on length and width scales. We introduce a fast method for computing the most likely completion between two elements, by developing novel analytic approximations and a fast numerical procedure for computing the curve of least energy. We then use our newly developed energies to find the most likely completions in images through a generalized summation of induction fields. This is done through multiscale procedures, i.e., separate processing at different scales with some interscale interactions. Such procedures allow the summation of all induction fields to be done in a total of only $O(N \log N)$ operations, where N is the number of pixels in the image. More important, such procedures yield a more realistic dependence of the induction field on the length and width scales: the field of a long element is very different from the sum of the fields of its composing short segments.

*Research supported in part by Israel Ministry of Science Grant 4135-1-93 and by the Gauss Minerva Center for Scientific Computation.

†Research supported in part by the Unites States-Israel Binational Science Foundation, Grant No. 94-00100. The vision group at the Weizmann Inst. is supported in part by the Israeli Ministry of Science, Grant No. 8504. Ronen Basri is an incumbent of Arye Dissentshik Career Development Chair at the Weizmann Institute.

1 Introduction

The smooth completion of fragmented curve segments is a skill of the human visual system that has been demonstrated through many compelling examples. Due to this skill people often are able to perceive the boundaries of objects even in the lack of sufficient contrast or in the presence of occlusions. A number of computational studies have addressed the problem of curve completion in an attempt to both provide a computational theory of the problem and as part of a process of extracting the smooth curves from images. These studies commonly obtain two or more edge elements (also referred to as *edgels*) and find either the most likely completions that connect the elements or the smoothest curves traveling through them. The methods proposed for this problem generally require massive computations, and their results strongly depend on the energy function used to evaluate the curves in the image. In addition, these methods ignore the size of the edge elements, and consequently often give inconsistent (and undesired) results at different scales (see, e.g., [1]). It is therefore important to develop methods which simplify the computation involved in these methods while providing results competitive with the existing approaches. Below we present such a method that directly relates to a number of recent studies of completion and curve salience [13, 28, 5, 10, 17, 30] (see also [23, 33, 2, 9, 25, 34, 20, 18, 11, 12, 24]). Along with simplifying the computations proposed in these studies our method also takes into account the size of edge elements, allowing for a proper computation of completion and saliency at different scales.

A number of studies have addressed the problem of determining the smoothest completion between pairs of edge elements [27, 23, 2, 13, 28, 5]. These studies seek to define a functional that, given two edge elements defined by their location and orientation in the image, selects the smoothest curve that connects the two as its minimizing curve. The most common functional is based on the notion of *elastica*, that is, minimizing the total squared curvature of the curve [13]. Scale invariant variations of this functional were introduced in [28, 5]. While the definition of scale-invariant elastica is intuitive, there exists no simple analytic expression to calculate its shape or its energy, and existing numerical computations are orders-of-magnitude too expensive, as will be shown below.

In the first part of this paper we revisit the problem of determining the smoothest completion between pairs of edges and introduce two new analytic approximations to the curve of least energy. These approximations are obtained by assuming that the deviation of the two input edgels from the straight line connecting them is relatively small. This assumption is valid in most of the examples used to demonstrate perceptual completions in humans and monkeys [14, 16, 8, 15, 32]. We show that under this simplifying assumption the Hermite spline (see, e.g., [19]) provides a good approximation to the curve of least energy and a very good approximation to the least energy itself. We further develop a second expression, which directly involves the angles formed by the edgels and the straight line connecting them. The second expression

is shown to give extremely accurate approximations to the curve of least energy even when the input edgels deviate significantly from the line connecting them. We then introduce a new, fast numerical method to compute the curve of least energy and show that our analytic approximations are obtained at early stages of this numerical computation.

Several recent studies view the problems of curve completion and salience as follows. Given M edge elements, the space of all curves connecting pairs of elements is examined in an attempt to determine which of these completions is most likely using smoothness and length considerations. For this purpose [10, 30] define an affinity measure between two edge elements that grows with the likelihood of these elements being connected by a curve. By fixing one of the elements and allowing the other element to vary over the entire image an *induction field* representing the affinity values induced by the fixed element on the rest of the image is obtained. The system finds the most likely completions for the M elements by applying a process that includes a summation of the induction fields for all M elements.

In the second part of this paper we use our newly developed completion energies to define an affinity measure that encourages smoothness and penalizes for gap length. We then use the induction fields defined by this affinity measure to solve the problem of finding the most likely completions for M elements. Since in practice edge elements are never dimensionless, because they are usually obtained by applying filters of a certain width w and length l to the image, we adjust our affinity measure to take these parameters into account. We do so by relating the scale of these filters to the range of curvature radii that they detect, and by determining the orientational resolution required for representing properly all significantly different edge elements of every scale. Finally, we show that due to the *smoothness* and *decay* properties of our affinity measure it can be implemented in all significantly different (l, w) scales, using multigrid methods, and run efficiently in time complexity $O(N \log N)$, where N is the number of pixels in the image.

Several existing studies acknowledge the importance of scale in curve detection and perceptual grouping (e.g., [34, 6]). These studies apply filters of different size to the image to determine the local properties of the measured curves (such as the local curvature), but they do not incorporate scale in the global stage of connecting the edgels to form curves. In contrast, we introduce a method for completion and curve extraction that incorporates scale in all stages of the computation. We apply filters of different lengths and widths to the image to detect edgels of different size and curvature. We then use these edgels to complete over gaps according to the size of the edgels. The long and thin elements are allowed to reach farther within a fairly specific orientation, while the short and fat elements are allowed to reach closer within a wide range of orientations. Moreover, none of the mentioned studies provide a comprehensive completion process in all different scales in a total complexity which is practically linear. ($O(N \log N)$, where N is the number of pixels in the image.) It should be also pointed out that some of the psychophysical studies of curve detection and perceptual grouping indicate (even

if indirectly) a dependence of curve completion on scale (see [21, 22]).

The paper is divided as follows. In Section 2 we review the notion of elastica and its scale invariant variation. In Section 3 we introduce the two analytic approximations to the curve of least energy. Then, in Section 4 we develop a fast numerical method to compute the curve of least energy and compare it to our analytic approximations. Finally, in Section 5 we construct an affinity measure taking into account the length and width of the edge filters applied to the image. We then discuss a multiscale (multigrid) method for fast summation of induction fields.

2 Elastica

Consider two edge elements e_1, e_2 positioned at $P_1, P_2 \in \mathcal{R}^2$ with directed orientations Ψ_1 and Ψ_2 respectively measured from the right-hand side of the line passing through P_1 and P_2. Below we shall confine ourselves to the case that $\Psi_1, \Psi_2 \in (-\frac{\pi}{2}, \frac{\pi}{2})$. Denote by $r = \|P_2 - P_1\|$, we may conveniently assume that $P_1 = (0,0)$ and $P_2 = (r,0)$. This is illustrated in Fig. 1. Let C_{12} denote the set of all smooth curves through e_1 and e_2. Denote such a curve by its orientation representation $\Psi(s)$, where $0 \leq s \leq L$ is the arclength along the curve. That is, $x(s) = \int_0^s \cos(\Psi(\hat{s}))d\hat{s}$ and $y(s) = \int_0^s \sin(\Psi(\hat{s}))d\hat{s}$. Also denote the curvature of the curve at s by $\kappa(s) = d\Psi(s)/ds$.

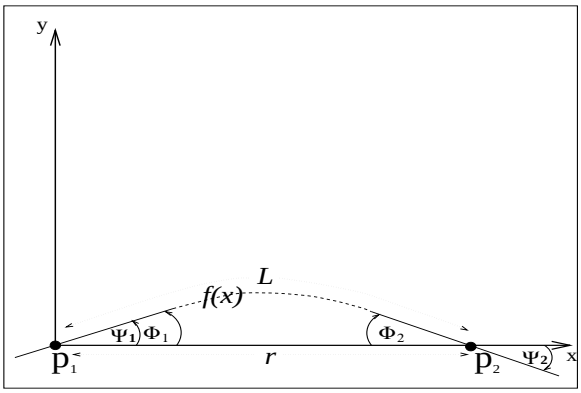

Figure 1: The planar relation between two edge elements, (P_1, Ψ_1) and (P_2, Ψ_2). This relation is governed by $\Phi_1 \in (-\frac{\pi}{2}, \frac{\pi}{2})$, $\Phi_2 \in (-\frac{\pi}{2}, \frac{\pi}{2})$, and r, where Φ_1 and Φ_2 are measured from the line P_1P_2, hence $\Phi_1 = \Psi_1$ and $\Phi_2 = -\Psi_2$. Note the more general relation between Φ_i and Ψ_i in Fig. 4.

The most common functional used to determine the smoothest curve traveling through P_1 and P_2 with respective orientations Ψ_1 and Ψ_2 is the *elastica* functional. Namely, the smoothest curve through e_1 and e_2 is the curve $\Psi(s)$ which minimizes the functional

$$\Gamma_{el}(\Psi) \stackrel{def}{=} \int_0^L \kappa^2(s)ds. \tag{1}$$

Elastica was already introduced by Euler. It was first applied to completion by Ullman [27], and its properties were further investigated by Horn [13].

One of the problems with the classical elastica model is that it changes its behavior with a uniform scaling of the image. In fact, according to this model if we increase r, the distance between the two input elements, the energy of the curve connecting them proportionately *decreases*. This can be easily seen by rescaling s, that is setting $\hat{s} = \alpha s$ where $0 < \alpha \in \mathcal{R}$, and letting $\hat{\Psi}(\hat{s}) = \Psi(\hat{s}/\alpha)$, so that

$$\Gamma_{el}(\hat{\Psi}) = \int_0^L (\frac{d\hat{\Psi}(\alpha s)}{d(\alpha s)})^2 ds = \frac{1}{\alpha}\Gamma_{el}(\Psi). \tag{2}$$

This is somewhat counter-intuitive since psychophysical and neurobiological evidence suggests that the affinity between a pair of straight elements drops rapidly with the distance between them [15]. Also, the classical elastica does not yield circular arcs to complete cocircular elements. To solve these problems Weiss [28, 5] proposed to modify the elastica model to make it scale invariant. His functional is defined as

$$\Gamma_{inv}(\Psi) \stackrel{def}{=} L \int_0^L \kappa^2(s) ds \ . \tag{3}$$

We believe that a proper adjustment of the completion energy to scale must take into account not only the length of the curve (or equivalently the distance between the input elements), but also the dimensions of the input edge elements. Both the elastica functional and its scale invariant version assume that the input elements have no dimensions. In practice, however, edge elements are frequently obtained by convolving the image with filters of some specified width and length. A proper adjustment of the completion energy as a result of scaling the distance between the elements should also consider whether a corresponding scaling in the width and length of the elements has taken place. Below we first develop useful approximations to the scale invariant functional. (These approximations can also be used with slight modifications to the classical elastica measure.) Later, in Section 5, we develop an affinity measure between elements that also takes into account both the distance between the elements and their dimensions.

3 Analytic simplification of Γ_{inv}

Although the definition of both the classical and the scale invariant elastica functionals is fairly intuitive, there is no simple closed-form expression that specifies the energy or the curve shape obtained with these functionals. In this section we introduce two simple, closed-form approximations to these functionals. Our approximations are valid when the sum of angles $|\Phi_1| + |\Phi_2|$ is relatively small. This assumption represents the intuition that in most psychophysical demonstrations gap completion is perceived when the orientations of the curve portions to be

completed are nearly *collinear*. With this assumption we may also restrict for now the range of applicable orientations to $\Psi_1, \Psi_2 \in (-\frac{\pi}{2}, \frac{\pi}{2})$.

Since the curve of least energy is supposed to be very smooth, it is reasonable to assume that within the chosen range of Ψ_1, Ψ_2 the smoothest curve will not wind much. Consequently, it can be described as a function $y = f(x)$, as in Fig. 1. Expressing the curvature in terms of x and y we obtain that

$$\Gamma_{inv}(\Psi) = L \int_0^L \kappa^2(s) ds = L \int_0^r \frac{(f'')^2}{(1 + (f')^2)^{\frac{5}{2}}} dx \ . \tag{4}$$

For small $|\Phi_1| + |\Phi_2|$ we have that due to the smoothness of the minimizing curve, both f' and f'' are of the order of magnitude of $\max\{|t_1|, |t_2|\}$, where $t_i = \tan\Phi_i$, $i = 1, 2$. Therefore we get that $L \simeq r$, and that the variation of f' becomes unimportant for the comparison of $\Gamma_{inv}(\Psi)$ over different curves $\Psi \in C_{12}$ that are relevant for the minimization. Thus for small $|\Phi_1| + |\Phi_2|$,

$$\Gamma_{inv}(\Psi) \simeq r \int_0^r (f'')^2 dx \ . \tag{5}$$

Hence

$$E_{inv} \stackrel{def}{=} \min_{\Psi \in C_{12}} \Gamma_{inv}(\Psi) \simeq r \min_{\Psi \in C_{12}} \int_0^r (f'')^2 dx \ . \tag{6}$$

The minimizing curve is the appropriate cubic Hermite spline (see Appendix A)

$$f(x) = x(x - r)\left(\frac{1}{r^2}(t_1 - t_2)x - \frac{t_1}{r}\right) \ , \tag{7}$$

where $t_1 = \tan\Phi_1$ and $t_2 = \tan\Phi_2$, so that

$$E_{inv} \simeq 4(t_1^2 + t_2^2 - t_1 t_2) \ . \tag{8}$$

Evidently, this simple approximation to E_{inv} is scale-independent. This leads us to define the scale-invariant *spline* completion energy as:

$$E_{spln}(\Phi_1, \Phi_2) \stackrel{def}{=} 4(t_1^2 + t_2^2 - t_1 t_2) \ . \tag{9}$$

Although the spline energy provides a good approximation to the scale invariant elastica measure for small values of $|\Phi_1| + |\Phi_2|$, the measure diverges for large values. An alternative approximation to E_{inv} can be constructed by noticing that for such small values $\tan\Phi_1 \simeq \Phi_1$ and $\tan\Phi_2 \simeq \Phi_2$. Thus, we may define:

$$E_{ang}(\Phi_1, \Phi_2) \stackrel{def}{=} 4(\Phi_1^2 + \Phi_2^2 - \Phi_1\Phi_2) \ . \tag{10}$$

We refer to this functional as the scale-invariant *angular* completion energy. This measure does not diverge for large values of $|\Phi_1| + |\Phi_2|$. In fact, when $\Phi_1 = \Phi_2 = \pi/2$ we obtain $E_{ang} = E_{inv} = \pi^2$. In Section 4 below we show that this angular energy is obtained in an early stage of the numeric computation of E_{inv}, and that it provides extremely accurate approximations to the scale invariant least energy functional even for relatively large values of $|\Phi_1|+|\Phi_2|$. In fact, particularly good approximations are obtained for small $|\Psi_1 + \Psi_2|$, i.e., for the range of nearly cocircular elements. Using the numeric computation we can also derive the smoothest curve according to E_{ang}:

$$\bar{\Psi}(s) = 3(\Psi_1 + \Psi_2)s^2 - (4\Psi_1 + 2\Psi_2)s + \Psi_1 \,. \tag{11}$$

The *angular* completion energy can be generalized as follows:

$$E_{gang}(\Phi_1, \Phi_2) = a(\Phi_1^2 + \Phi_2^2) + b(\Phi_1 - \Phi_2)^2 \,, \tag{12}$$

where Eq. (10) is identical to Eq. (12) with $a = b = 2$. That is, the *angular* completion energy is made of an equal sum of two penalties. One is for the squared difference between Φ_1 and Φ_2, and the other is for the growth in each of them. This suggests a possible generalization of E_{ang} to other weights $a \geq 0$ and $b \geq 0$.

In defining the curve of least energy and its approximations we seek the smoothest completion between two given edge elements. An interesting variation of this problem is the following. Given a single edge element $e_1 = (P_1, \Phi_1)$, and given a location in the image P_2, suppose we pass the smoothest curve from e_1 through P_2, what orientation would this curve take at $P_2\Gamma$ This question is relevant, for instance, in studies, such as [10, 30], which compare, given an edge element, several different candidate elements for completion.

Consider the angular completion energy (Eq. (10)). Notice first that when a scale invariant energy is used the answer will not depend on the position of P_2, but only on the orientation Φ_1. According to E_{ang} the minimal energy is obtained when the orientation of the element at P_2 is given by $\Phi_2 = \frac{1}{2}\Phi_1$. (A similar result is also approximately true for the spline, the scale-invariant elastica and the classical elastica energies.) For the generalized angular completion energy (Eq. (12)) the preferred orientation is $\Phi_2 = \frac{a}{a+b}\Phi_1$. The energy field induced by one edge element, therefore, does not prefer circular completions ($\Phi_1 = \Phi_2$). A circular completion can be preferred if we modify the angular energy to be:

$$E_{circ}(\Phi_1, \Phi_2) \stackrel{def}{=} a\Phi_1^2 + b(\Phi_1 - \Phi_2)^2 \,,$$

for which, if we fix Φ_1, the lowest energy is obtained when $\Phi_2 = \Phi_1$. This energy, however, is non-symmetric for Φ_1 and Φ_2.

Finally, we note that the new approximations at small angles can also be used to approximate the classical elastica energy, since

$$E_{el} \stackrel{def}{=} \min_{\Psi \in C_{12}} \Gamma_{el}(\Psi) \simeq \frac{1}{r} E_{ang}(\Phi_1, \Phi_2) \simeq \frac{1}{r} E_{spln}(\Phi_1, \Phi_2) \,. \tag{13}$$

4 Computation of E_{inv}

We now introduce a fast numeric method to compute E_{inv}. We use the scale-invariance property of Γ_{inv} (as in [5]) in order to reformulate the minimization problem. That is, every curve Ψ corresponds to a rescaled version of it, $\widetilde{\Psi}$, which satisfies $L = 1$ and $\Gamma_{inv}(\widetilde{\Psi}) = \Gamma_{inv}(\Psi)$. We can see this by setting (scaling) $\tilde{s} = s/L$ and defining $\widetilde{\Psi}(\tilde{s}) \stackrel{def}{=} \Psi(L\tilde{s})$. Then, the minimization problem takes the form:

$$\min_{\widetilde{\Psi}} \left(\int_0^1 (\widetilde{\Psi}'(\tilde{s}))^2 d\tilde{s}\right) \quad s.t. \tag{14}$$
$$\int_0^1 \sin(\widetilde{\Psi}(\tilde{s}))d\tilde{s} = 0, \quad \widetilde{\Psi}(0) = \Psi_1, \quad \widetilde{\Psi}(1) = \Psi_2 \ .$$

Denoting the extremal curve found by $\widetilde{\Psi}_{min}(\tilde{s})$, set $L = r/(\int_0^1 \cos(\widetilde{\Psi}_{min}(\tilde{s}))d\tilde{s})$, so that $\int_0^L \cos(\widetilde{\Psi}_{min}(s/L))ds = r$. Thus, $\Psi_{min}(s) \stackrel{def}{=} \widetilde{\Psi}_{min}(s/L)$ will be the "minimal curve". In fact, to calculate E_{inv} it is not necessary to calculate L, since $E_{inv} = \Gamma_{inv}(\widetilde{\Psi}_{min})$. Now, Eq. (14) can be transformed by the Euler-Lagrange equations (see, e.g., [7]) into an ODE problem. That is, a necessary condition for $\widetilde{\Psi}(\tilde{s})$ to be an extremal curve is that it should satisfy for some λ:

$$2\widetilde{\Psi}'' = \lambda \cos\widetilde{\Psi} \quad s.t. \tag{15}$$
$$\int_0^1 \sin(\widetilde{\Psi}(\tilde{s}))d\tilde{s} = 0, \quad \widetilde{\Psi}(0) = \Psi_1, \quad \widetilde{\Psi}(1) = \Psi_2 \ .$$

Considering the very nature of the original minimization problem, and also by repeatedly differentiating both sides of the ODE equation, it can be shown that its solution must be very smooth. Hence, we can well approximate the solution by a polynomial of the form

$$\widetilde{\Psi}_n(s) = (1-s)\Psi_1 + s\Psi_2 + s(1-s)\sum_{k=0}^{n} a_k s^k \ , \tag{16}$$

where n is small. (By comparison, the discretization of the same problem presented in [5] is far less efficient, since it does not exploit the infinite smoothness of the solution on the full interval (0,1). As a result the accuracy in [5] is only second order, while here it is "∞-order", i.e., the error decreases *exponentially* in the number of discrete variables n+2 (i.e., $\lambda, a_0, ..., a_n$ from Eq. (15) and (16)).) Fixing n, as well as two other integers $\bar{n}$ and p, we will build the following system of $n+2$ equations for the $n+2$ unknowns $a_0, a_1, ..., a_n, \ and \ \lambda$

$$\widetilde{\Psi}_n''(\frac{i+1}{n+2}) + \lambda_n \cos\widetilde{\Psi}_n(\frac{i+1}{n+2}) = 0 \ , \ (0 \le i \le n)$$

collocating the ODE, and

$$\sum_{j=0}^{\bar{n}} w_j \sin\widetilde{\Psi}_n(\frac{j}{\bar{n}}) = 0 \ ,$$

where w_j $(0 \leq j \leq \bar{n})$ are the weights of a p-order numerical integration. Generally, we increase n gradually and increase $\bar{n}$ and p as functions of n in such a way that the discretization error will not be governed by the discretization error of the integration. The nonlinear system of $n+2$ equations is solved by Newton iterations (also called Newton-Raphson; see, e.g. [19].) We start the Newton iterations from a solution previously obtained for a system with a lower n. Actually, only *one* Newton iteration is needed for each value of n if n is not incremented too fast. In this way convergence is extremely fast. At each step, in just several dozen computer operations, the error in solving the differential equation can be *squared*. In fact, due to the smoothness of the solution for the ODE, already for the simple $(n = 0, \bar{n} = 2)$-system and the Simpson integration rule $(p = 3)$, a very good approximation to the accurate solution $\widetilde{\Psi}(s), \lambda$ and also to $E_{inv} = \int_0^1 (\widetilde{\Psi}'(\tilde{s}))^2 d\tilde{s}$ is obtained, as can be seen in Table 1 below. The good approximations obtained already for small values of $(n,\bar{n})$ suggest that E_{inv} can be well approximated by simple analytic expressions, as indeed we see next.

Table 1 compares the result of applying the numerical computation of E_{inv} to several simple approximations, each of which can be used as the scale-invariant completion energy. These approximations are:

- The "$(n = 0, \bar{n} = 2)$-system" using the Simpson $(p = 3)$ integration method is solved numerically, yielding the "Simpson method" energy. Note that we can also well approximate the solution of this system by the *analytic* expression $\bar{\Psi}(s)$ presented in Eq. (11), which arises from approximating the system to its first order assuming small $|\Psi_1| + |\Psi_2|$ (see Appendix B). The "Simpson method" energy can be approximated analytically by

$$\Gamma_{inv}(\bar{\Psi}(s)) = 4(\Phi_1^2 + \Phi_2^2 - \Phi_1\Phi_2) \ ,$$

  and thus is expected to yield results similar to $E_{ang}(\Phi_1, \Phi_2)$.

- The "$(n = 0, \bar{n} = 2)$-system" using the Trapezoidal $(p = 2)$ integration method can be *solved analytically* for all values of Φ_1 and Φ_2 in $(-\frac{\pi}{2}, \frac{\pi}{2})$, yielding the "Trapezoidal method" energy

$$E_{inv}^{(1)} = (\Phi_1^2 + \Phi_2^2) + \frac{4}{3}(\Phi_1 - \Phi_2)^2 \ .$$

  This energy is of the type of Eq. (12)

- The angular completion energy $E_{ang}(\Phi_1, \Phi_2)$ derived in Sec. 3. This method is referred to as the "Angular method."

- The spline completion energy $E_{spln}(\Phi_1, \Phi_2)$ derived in Sec. 3. This method is referred to as the "Spline method."

The results in Table 1 demonstrate that in the range of angles in which perceptual completions are anticipated there is hardly any difference between the four different approximations

and the accurate solution of E_{inv}. An even closer agreement is obtained when derivatives of these energies are compared (see Table 2.)

Fig. 2 illustrates some of the completions obtained using E_{inv} and the two analytic approximations E_{ang} and E_{spln}. It can be seen that the differences between the three curves is barely noticeable, except in large angles where E_{spln} diverges. Notice especially the close agreement between the curve obtained with the angular energy (Eq. (11)) and that obtained with the scale-invariant elastica measure even in large angles and when the angles deviate significantly from cocircularity.

Note that although the spline curve does not approximate the scale invariant elastica curve for large angles $|\Phi_1|$ and $|\Phi_2|$ it still produces a reasonable completion for the elements. In fact, when the two elements deviate from cocircularity the elastica accumulates high curvature at one of its ends, whereas the spline curve continues to roughly follow the tangent to the two elements at both ends (see, e.g, Fig. 2(c) and (d)). This behavior is desirable especially when the elements represent long curve segments (see Section 5.2).

5 Completion field summation

Until now we have considered the problem of finding the smoothest completion between pairs of edge elements. A natural generalization of this problem is, given an image from which M edge elements are extracted, find the most likely completions connecting pairs of elements in the image and rank them according to their likelihoods. This problem has recently been investigated in [10, 30]. In these studies affinity measures relating pairs of elements were defined. The measures encourage proximity and smoothness of completion. Using the affinity measures the affinities induced by an element over all other elements in the image (referred to as the *induction field* of the element) are derived. The likelihoods of all possible completions are then computed simultaneously by a process which includes summation of the induction fields for all M elements.

An important issue that was overlooked in previous approaches, however, is the issue of size of the edge elements. Most studies of curve completion assume that the edge elements are dimensionless. In practice, however, edge elements are usually obtained by convolving the image with filters of certain width and length. A proper handling of scale must take these parameters into account. Thus, for example, one may expect that scaling the distance between two elements would not result in a change in the affinity of the two elements if the elements themselves are scaled by the same proportion. Below we first present the general type of *non-scaled* induction underlying previous works. We then modify that induction to properly account for the width and length of the edge elements.

Finally, the process of summing the induction fields may be computationally intensive. Nevertheless, in the third part of this section we show that the summation kernel obtained with

Φ_1	Φ_2	Accur. meth.	Simps. meth.	Trapz. meth.	Angul. meth.	Splin. meth.
0	0	0	0	0	0	0
.01	0	.00039	.00039	.00023	.00040	.00040
	.01	.00040	.00040	.00040	.00040	.00040
	.02	.0011	.0011	.0010	.0012	.0012
	.03	.0027	.0027	.0021	.0028	.0028
.1	0	.0399	.0399	.0233	.0400	.0403
	.1	.0399	.0400	.0400	.0400	.0403
	.2	.1199	.1198	.1033	.1200	.1233
	.3	.2793	.2792	.2133	.2800	.2989
.3	0	.3590	.3588	.2100	.3600	.3828
	.3	.3599	.3600	.3600	.3600	.3828
	.6	1.0725	1.0718	.9300	1.0800	1.4084
	.9	2.4660	2.4600	1.9200	2.5200	5.1755
.5	0	.9928	.9909	.5833	1.0000	1.1938
	.5	1.0000	1.0000	1.0000	1.0000	1.1938
	1.0	2.9425	2.9379	2.5833	3.0000	7.4926
	1.5	6.5819	6.5570	5.3333	7.0000	765.77
1	0	3.8851	3.8597	2.3333	4.0000	9.7021
	1	3.9999	4.0000	4.0000	4.0000	9.7021

Table 1: Comparison table for the various $E_{inv}(\Phi_1, \Phi_2)$ approximations. Φ_1 and Φ_2 are given in radians.

Φ_1	Φ_2	Accur. meth.	Simps. meth.	Trapz. meth.	Angul. meth.	Splin. meth.
.01	0	266.66	266.66	200.14	200	199.98
	.01	200	200	200	200	199.98
	.02	218.18	218.18	206.70	200	66.648
	.03	207.40	207.40	203.19	200	114.35
.1	0	20.037	20.035	20.030	20	20.069
	.1	20.050	20	20	20	20.069
	.2	19.985	19.999	20.006	20	7.1266
	.3	19.955	19.949	20.003	20	12.741
.3	0	6.6507	6.6455	6.6667	6.6667	7.0836
	.3	6.6685	6.6667	6.6667	6.6667	7.0836
	.6	6.6209	6.6163	6.6667	6.6667	3.9182
	.9	6.5207	6.5084	6.6667	6.6667	9.8004
.5	0	3.9713	3.9643	4.0002	4.0000	4.7544
	.5	4.0000	4.0000	4.0000	4.0000	4.7544
	1	3.9216	3.9180	4.0000	4.0000	5.9946
	1.5	3.7444	3.7479	4.0000	4.0000	84.718
1	0	1.9406	1.9300	2.0000	2.0000	4.3991
	1	2.0001	2.0000	2.0000	2.0000	4.3991

Table 2: Comparison table for the values of the ratios $\frac{dE(\Phi_1,\Phi_2)}{d\Phi_1}\frac{1}{E(\Phi_1,\Phi_2)}$, where $\Phi_2 = k\Phi_1$ for $k = 0, 1, 2, 3$, and E stands for the various E_{inv} approximations. Φ_1 and Φ_2 are given in radians.

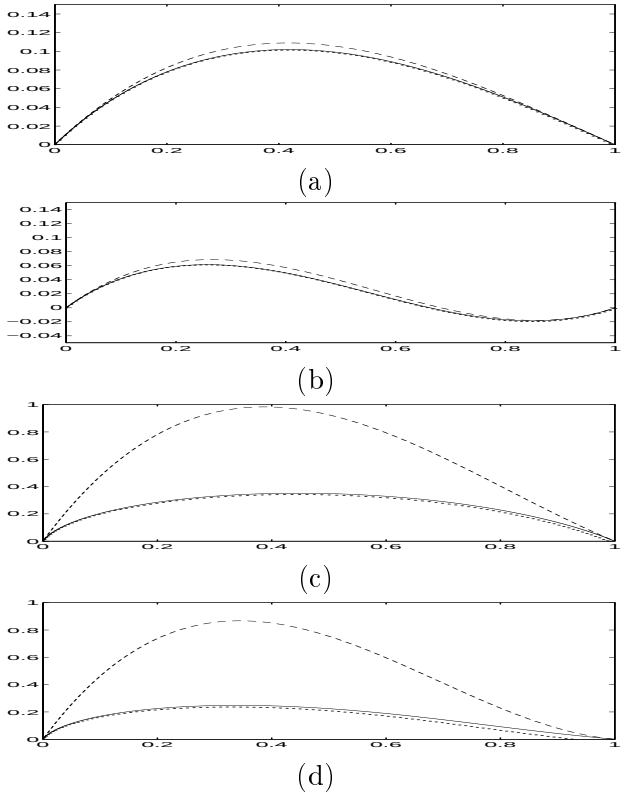

Figure 2: Completion curves: elastica in solid line, $\bar{\Psi}(s)$ (Eq. (11)) in dotted line, and the cubic Hermite spline (Eq. (7)) in dashed line. (a) $\Phi_1 = 30°, \Phi_2 = 15°$, (b) $\Phi_1 = 30°, \Phi_2 = -15°$, (c) $\Phi_1 = 80°, \Phi_2 = 60°$, (d) $\Phi_1 = 80°, \Phi_2 = 20°$.

our method is very smooth. Thus, the summation of our induction fields can be speeded up considerably using a multigrid algorithm. This result also applies to the summation kernels in [30, 26, 10], and so an efficient implementation of these methods can be obtained with a similar multigrid algorithm.

5.1 Non-scaled induction

In [17, 30] a model for computing the likelihoods of curve completions, referred to as *Stochastic Completion Fields*, was proposed. According to this model, the edge elements in the image emit particles which follow the trajectories of a Brownian motion. It was shown that the most likely path that a particle may take between a *source* element and a *sink* element is the curve of least energy according to the Elastica energy function[1]. To compute the stochastic completion fields a process of summing the affinity measures representing the source and sink fields was used. In

[1]Actually, the path minimizes the energy functional $\int_0^L \kappa^2(s)ds + \lambda L$ for some predetermined constant λ.

Appendix C we show, by further analyzing the results in [26], that the affinity measure used for the induction in [30, 26] is of the general type:

$$A(e_1, e_2) \stackrel{def}{=} e^{-r/r_0} e^{-E_{ang}/(r\sigma_0)} \ ,$$

where r_0 and σ_0 are strictly positive a-priori set parameters. These parameters need to be adjusted properly according to the scale involved (see Sec. 5.2). Note that for small values of $(|\Phi_1|,|\Phi_2|)$: $E_{ang}/r \simeq E_{el}$. Hence,

$$A(e_1, e_2) \simeq e^{-r/r_0} e^{-E_{el}/\sigma_0} \ .$$

Another method which uses summation of induction fields to compute the salience of curves was presented in [10]. In this method the affinity between two edge elements which are cocircular has the form: $e^{-\gamma r} e^{-\delta\kappa}$, where γ and δ are strictly positive constants, κ is the curvature of the circle connecting e_1 and e_2, and r is the distance between e_1 and e_2. A reasonable and straightforward definition in that spirit is

$$\widetilde{A}(e_1, e_2) \stackrel{def}{=} e^{-\gamma r} e^{-\delta E_{spln}}$$

where E_{spln} serves as an approximation for E_{inv} according to Eq. (8). Fig. 3 shows an example of computing the "stochastic completion field," suggested by Williams and Jacobs in [30], while replacing their affinity measure with the simple expression $\widetilde{A}(e_1, e_2)$. It can be verified by comparing the fields obtained with our affinity measure with the fields presented in [30] that the results are very similar although a much simpler affinity measure was employed.

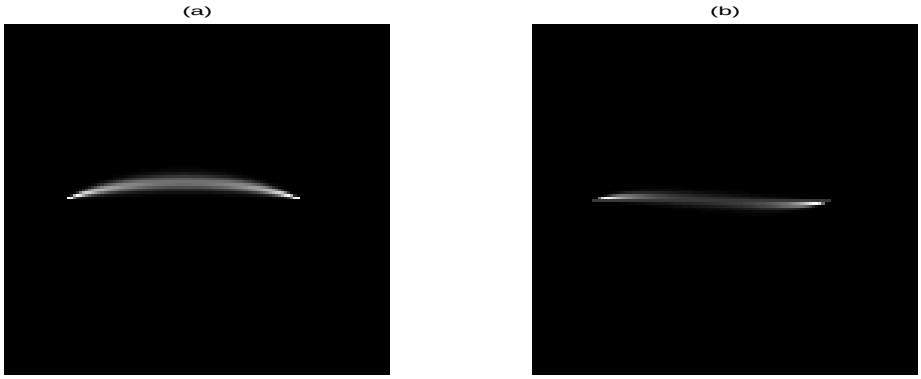

Figure 3: Stochastic completion fields (128×128 pixels, 36 orientations) with the induction $e^{-2r} e^{-20E_{spln}}$. (a) $\Phi_1 = 30^\circ$ and $\Phi_2 = 30^\circ$, (b) $\Phi_1 = 30^\circ$ and $\Phi_2 = -30^\circ$. The results closely resemble those obtained in [30].

5.2 Induction and scale

Given an image, an edge element is produced by selecting a filter of a certain length l and width w (e.g., rectangular filters) and convolving the filter with the image at a certain position and orientation. The result of this convolution is a scalar value, referred to as the *response* of the filter. An edge filter may, for example, measure the contrast along its primary axis, in which case its response represents the "edgeness level", or the likelihood of the relevant subarea of the image to contain an edge of (l, w) scale. Similarly, a filter may indicate the existence of fiber-like shapes in the image, in which case its response represents the "fiberness level" of the relevant subarea of the image. Below we use the term "straight responses" to refer to the responses obtained by convolving the image with either an edge or a fiber filter.

Consider now the edge elements obtained by convolving the image with a filter of some fixed length l and width w. Every edge element now is positioned at a certain pixel P and is oriented in two opposite directed orientations Ψ and $\Psi + \pi$, where Ψ is measured from the right-hand side of the horizontal axis (see Fig. 4).

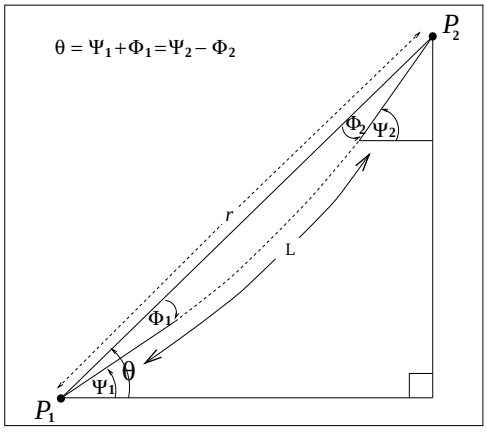

Figure 4: The relation between the two straight responses (x_1, y_1, Ψ_1) and (x_2, y_2, Ψ_2). This relation is governed by Φ_1,Φ_2, and r.

The number of edge elements required to faithfully represent the image at this scale depends on l and w. Thus, long and thin elements require finer resolution in orientation than square elements. In fact, the orientational resolution required to sample significantly different orientations increases linearly with l/w (see [4]). Similarly, elements of larger size require less spatial resolution than elements of smaller size. Brandt and Dym ([4]) use these observations in order to introduce a very efficient computation ($O(N \log N)$, where N is the number of pixels in the image) of *all* significantly different edge elements.

Given a particular scale determined by the length l and width w of edge elements, we would like to compute a completion field for this scale. Note that only curves within a relevant range of curvature radii can arouse significant responses for our $l \times w$ elements. Denote the

smallest curvature radius that will arouse a still significant response by $\rho = \rho(l, w)$. (Larger curvature radii will arouse significant responses also in larger l/w scales, implying there for a farther-reaching and more orientation-specific continuation.)

By Fig. 5 we see that

$$\frac{l}{2} = \rho \sin \alpha \approx \rho\alpha$$

and

$$w = \rho - \rho \cos \alpha \approx \frac{\rho\alpha^2}{2} \ .$$

Consequently, we have

$$\left(\frac{l}{2\rho}\right)^2 \approx \alpha^2 \approx \frac{2w}{\rho} \ ,$$

implying that

$$\rho \approx \frac{l^2}{8w} \ .$$

Next, consider a pair of straight responses. Assuming these elements are roughly cocircular, then, using the relations defined in Fig. 6, the differential relation

$$\Psi'(s) = \frac{1}{\rho(s)}$$

can be approximated by

$$\frac{\Psi_1 - \Psi_2}{r} \approx \frac{2}{\rho_1 + \rho_2} \ ,$$

so that

$$\Omega \approx \frac{r}{\rho} \ .$$

Hence, for completion at a particular scale (l, w), it is reasonable to define for every pair of points P_1 and P_2 a *scale* for the turning angle Ω given by $r/\rho(l, w)$. That is, in the scale (l,w) we define the completion energy between the pair of straight responses so as to depend on the *scaled* turning angle $\Omega\rho/r$. Since $\Omega = \Phi_1 + \Phi_2$, it is straightforward to show that

$$0.5\Omega \leq \sqrt{E_{ang}(\Phi_1, \Phi_2)} \leq \Omega \ .$$

A reasonable definition for the *scaled* angular energy, therefore, is a monotonically decreasing function of

$$\frac{\rho}{r}\sqrt{\Phi_1^2 + \Phi_2^2 - \Phi_1\Phi_2} \ .$$

Obviously, in any given scale of straight responses, (l, w), for every Φ_1 and Φ_2, the induction of P_1 upon P_2 should decrease with an increase of r/ρ. Hence, we define the field induced by

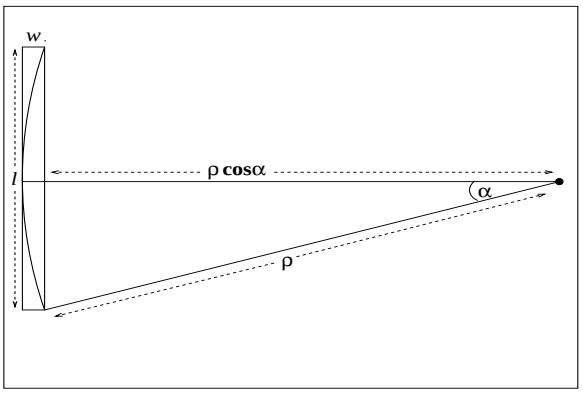

Figure 5: The relation between l,w, and the curvature radius ρ.

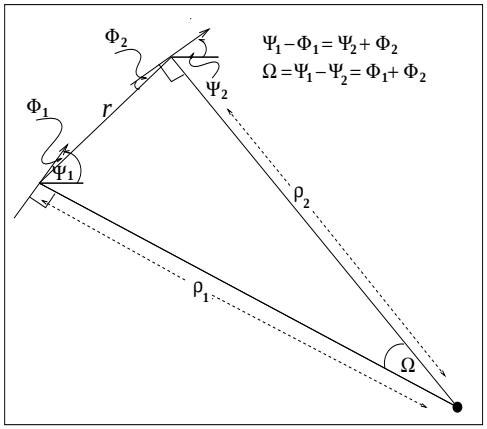

Figure 6: The turn Ω that a moving particle takes in its way between two straight responses, characterized each by a planar location and an orientation.

an element e_1 of length l and width w at location P_1 and directed orientation Ψ_1 on a similar element e_2 at (P_2, Ψ_2) by

$$G_{(l,w)}(e_1; e_2)\tilde{f}(u_1)\,, \tag{17}$$

where u_1 denotes the strength of response at e_1, $\tilde{f}(u_1)$ is some appropriate function of this response, and

$$G_{(l,w)}(e_1; e_2) = F_d\left(\frac{r}{\rho}\right) F_t\left(\frac{\rho}{r}\sqrt{\Phi_1^2 + \Phi_2^2 - \Phi_1\Phi_2}\right)\,. \tag{18}$$

F_d and F_t (the distance and turning attenuation functions, respectively) are smoothly decreasing dimensionless functions that should be determined by further considerations and experience. Thus, our summation kernel is a product of the orientational and the spatial components involved in completing a curve between e_1 and e_2. As we shall see below, this definition has many computational advantages.

Let $\{u_i\}$ denote the set of straight responses for a given scale (l, w), where each u_i is associated with two directed edges $e_i = (P_i, \Psi_i)$ and $\bar{e}_i = (P_i, \Psi_i + \pi)$. The total field induced at any element $e_j = (P_j, \Psi_j)$ by all elements $\{e_i, \bar{e}_i\}$ is expressed by

$$v_j \stackrel{def}{=} \sum_i \left(G_{(l,w)}[e_i; e_j] + G_{(l,w)}[\bar{e}_i; e_j] \right) f(u_i) \ . \tag{19}$$

The total field induced at $\bar{e}_j$ by $\{e_i, \bar{e}_i\}$ is given by

$$\bar{v}_j \stackrel{def}{=} \sum_i \left(G_{(l,w)}[e_i; \bar{e}_j] + G_{(l,w)}[\bar{e}_i; \bar{e}_j] \right) f(u_i) \ . \tag{20}$$

Since in general the responses obtained by convolving the image with edge filters are bi-directional we may want to combine these two fields into one. This can be done in various ways. The simplest way is to take the sum $\{v_j + \bar{v}_j\}$ as the completion field. Another possibility, in the spirit of [30], is to take the product $\{v_j \bar{v}_j\}$ as the completion field.

Note that the field of a long straight response should be very different (farther-reaching and more orientation-specific) than the sum of the fields of shorter elements composing it, and should strongly depend on its width (see Fig. 7). This suggests that for a comprehensive completion process one must practice a multiscale process, performing a *separate* completion within each scale. The *scaled* induction field (17)-(18), avoids a fundamental difficulty of non-scaled fields like [10, 30, 26]. The latter exhibit so weak a completion for far elements, that it would be completely masked out by local noise and foreign local features.

The fact that filters of different lengths respond differently depending on the curvature radius of the measured curve was noticed also by Zucker et al. [34], who used this fact to obtain an estimate of the local curvature at every point. In contrast to this work we do not estimate the curvatures locally, but only determine locally the range of curvatures that is compatible with the size of the filters. We then allow the global process of summation to detect the smoothest curves whose curvatures are within this range. The disadvantage of estimating the curvatures locally is that gaps may severely affect the curvature estimation. For example, the response of filters of different lengths when applied to a segment of a dashed, straight line will be identical to their response if the segment was part of a curved contour.

5.3 Fast multigrid summation of induction-fields

Let $n = n(l, w)$ be the number of sites (P), and $m = m(l, w)$ the number of orientations (Ψ) at each site, that are required in order to describe all the $l \times w$ straight responses that are significantly different from each other. It can be shown (see [4]) that if l and w are measured in pixel units then, for any N-pixel picture, $n = O(N/(lw))$ and $m = O(l/w)$, so the total number of $l \times w$ elements is $O(N/w^2)$. Hence, for any geometric sequence of scales (e.g., l=1,2,4,..., and w=1,3,9,...) the total number of straight elements is $O(N \log N)$. It has been shown (in

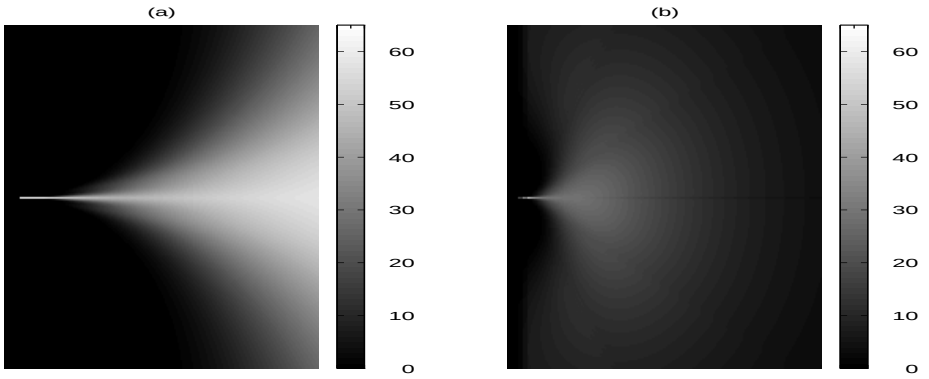

Figure 7: Induction fields (200 × 200 pixels) in different scales using $F_d(\xi) = e^{-\alpha_1 \xi^{\beta_1}}$, and $F_t(\eta) = e^{-\alpha_2 \eta^{\beta_2}}$, where $\alpha_1 = 0.5$, β_1=0.5, $\alpha_2 = 128$, and $\beta_2 = 1.5$. (a) The induction field of one long element: $l = 9$, $w = 1/2$, 25 directed orientations. (b) The sum of induction fields of the three shorter elements composing this long element, each consist of: $l = 3$, $w = 1/2$, 12 directed orientations.

[4]) that all the responses at all these elements can be calculated in only $O(N \log N)$ computer operations, using a multiscale algorithm that constructs longer-element responses from shorter ones.

At any given scale $l \times w$, it seems that the summations (Eqs. (19) and (20)), summing over $i = 1, 2, ..., nm$ for each value of $j = 1, 2, ..., nm$, would require a total of $O(n^2 m^2)$ operations (even though some of them can be performed in parallel, as in [31]). However, using the smoothness properties of the particular kernel (18), the summation can be reorganized in a multiscale algorithm that totals only $O(nm)$ operations (and the number of unparallelizable steps grows only logarithmically in nm.) To see this, note first that the functions in (18) would usually take on the typical form

$$F_d(\xi) = e^{-\alpha_1 \xi^{\beta_1}} \text{ and } F_t(\eta) = e^{-\alpha_2 \eta^{\beta_2}} \tag{21}$$

(as is explained in Sec. 5.1). For the purpose of the multiscale algorithm we call any function F *smooth on scale* s (in some region) *iff* it can (throughout that region) be interpolated, to any desired accuracy, from its values on a grid with meshsize s (the accuracy being increased by increasing the interpolation *order*). We call F *local on scale* s if F is negligibly small outside a disk of radius $O(s)$. One can then easily see that, as function of e_1 (similarly e_2), F_d in (18) is everywhere *smooth* on any scale s for which $s < O(\rho)$, and *local* on all other scales. Also (Cf. Appendix D), F_t is *smooth* on any scale s, in the region $r > O(s)$. Consequently, we get that $G = F_d F_t$ is *smooth* on any scale s for which $s < O(\rho)$, in the region $r > O(s)$, and *local* on all other scales.

Due to these smoothness properties of $G = G(e_i; e_j) = G(x_i, y_i, \Psi_i; x_j, y_j, \Psi_j)$, the total contribution to v_j (and $\bar{v}_j$) of all elements *far* (on scale s) from P_j is a smooth (on scale

$$\begin{bmatrix} -0.125 & 0 & 0.25 & 0 & -0.125 \\ -0.25 & 0 & 0.5 & 0 & -0.25 \\ -0.125 & 0 & 0.25 & 0 & -0.125 \end{bmatrix} \qquad \begin{bmatrix} -0.0625 & 0 & 0.125 & 0 & -0.0625 \\ -0.125 & 0 & 0.25 & 0 & -0.125 \\ -0.125 & 0 & 0.25 & 0 & -0.125 \\ -0.125 & 0 & 0.25 & 0 & -0.125 \\ -0.0625 & 0 & 0.125 & 0 & -0.0625 \end{bmatrix}$$

Figure 8: The filters used to detect vertical fibers of lengths 3 (left) and 5 (right) and width 1. Fibers of other directions (total of 8 and 16 for each length scale respectively) are obtained in a similar way with proper interpolations to account for discretization misalignments.

s) function of (x_j, y_j, Ψ_j), hence it need not be computed separately for each j, but can be interpolated (q-order interpolation, with as small an error as desired by using sufficiently high q) from its values on a grid with meshsize s. For this and similar reasons, multiscale algorithms, which split the summations into various scales of *farness* (see details in [3]) can perform all the summations in merely $O(nm)$ operations. Indeed, for any scale for which $s < O(\rho)$, we can use the *smoothness* properties of G in order to aggregate the summation onto a coarser grid (for which $s = O(\rho)$) in a cost of only $O(nm)$ operations. On such a grid we can already use the *localness* of $F_d(\xi)$ in order to get a cost of summation of $O(nm)$ operations.

However, the use of $F_d(\xi)$ which decays this fast (see Eq. 21) is well justified only in the case of employing a *multiscale* completion process. The reason for this is that a *single scale* completion process (as in [30, 10]) must have the following important property: the total influence of several elements of this scale that compose a longer element of a certain length $\tilde{l}$, should extend to a distance which is at least of the order of magnitude of $\tilde{l}$. This total influence is well approximated by $\int_r^{r+\tilde{l}} F_d(\xi)d\xi$ (since when considering distant smooth continuations $F_t \simeq 1$,) by which it can be readily shown that in the case of a *single scale* completion process F_d should not decay any faster than $F_d(\xi) = 1/\xi$. Nevertheless, for such choices of the functions, the kernel G still has the property of "asymptotic smoothness" (see Appendix D). By this we mean that any q-order derivative of G with respect to any of its six arguments decays fast with $r_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2]^{\frac{1}{2}}$, and the higher q is, the faster is the decay. Also, for any fixed r_{ij} (even the smallest, i.e., $r_{ij} = O(l)$,) G is a very smooth function of Ψ_i and of Ψ_j. Thus, even when practicing such a *single scale* completion process, employing, e.g., $F_d(\xi) = 1/\xi$, the summation can still be reorganized in a *multigrid* algorithm that totals only $O(nm)$ operations.

Figures 9-11 demonstrate the completion field $\{v_j \bar{v}_j\}$ as defined by Eq. (19)-(20), using F_d and F_t as defined by Eq. (21) with the free parameters $\alpha_1 = 1$, $\beta_1 = 1.5$, $\alpha_2 = 1$, and $\beta_2 = 1.5$. To obtain these completions we first detected fibers by using differences of straight intensity integrals of length 3 and 5 (width 1) in all significantly different directions (the actual filters used are shown in Figure 8). Finally, we produced a map of contour completion that combines the two scales.

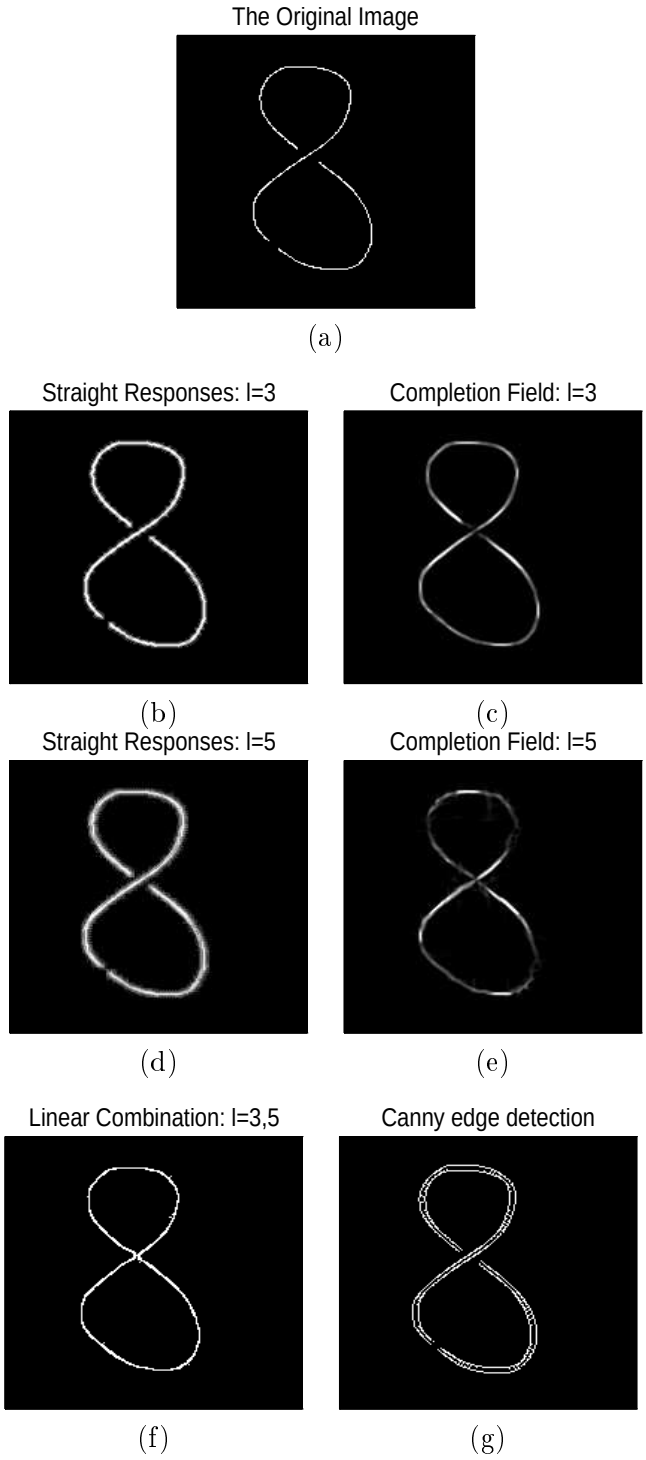

Figure 9: (a) The original image. (b)-(e) Completion fields in length scales 3 and 5 (width 1). (f) This picture shows the application of a skeleton detector to a linear combination of the two completion fields (after thresholding). (g) Canny edge detection with SD=2 and T=0.2. Images are of size 200 x 200 pixels.

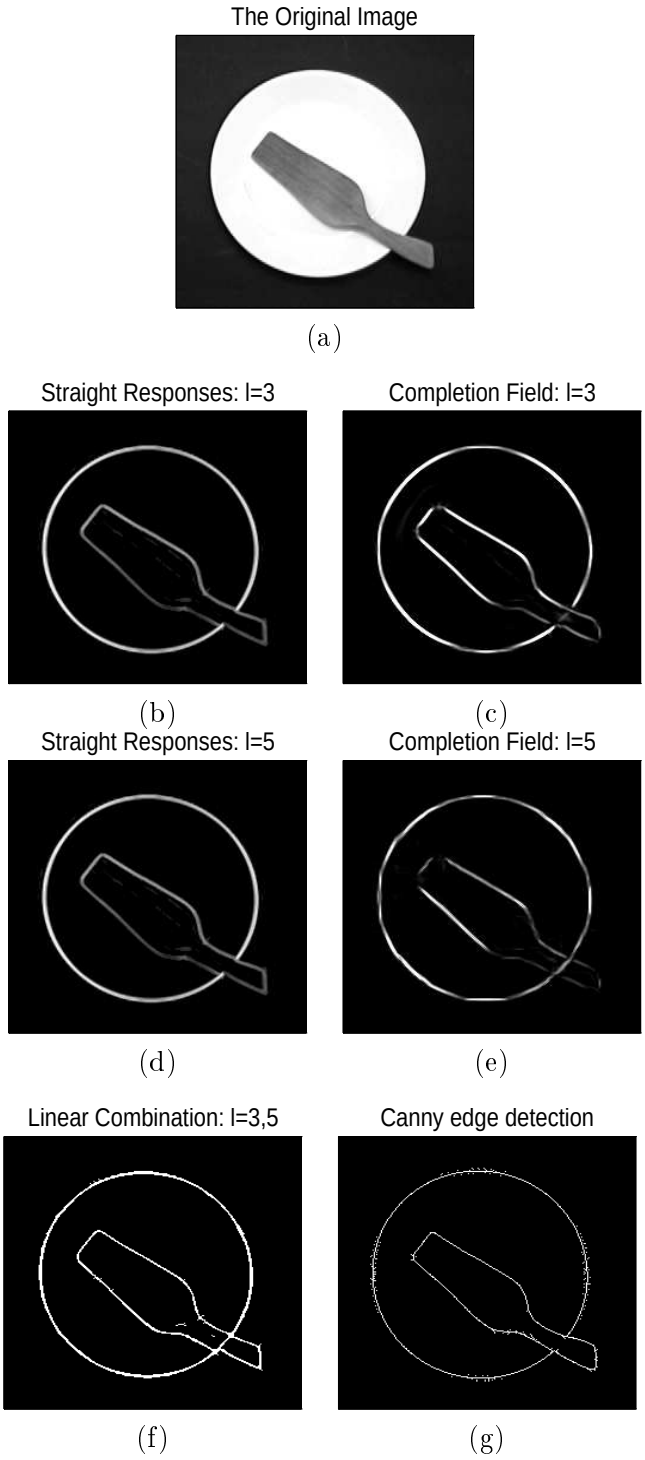

Figure 10: (a) The original image. (b)-(e) Completion fields in length scales 3 and 5 (width 1). (f) This picture shows the application of a skeleton detector to a linear combination of the two completion fields (after thresholding). (g) Canny edge detection with SD=2 and T=0.2. Images are of size 250 x 250 pixels.

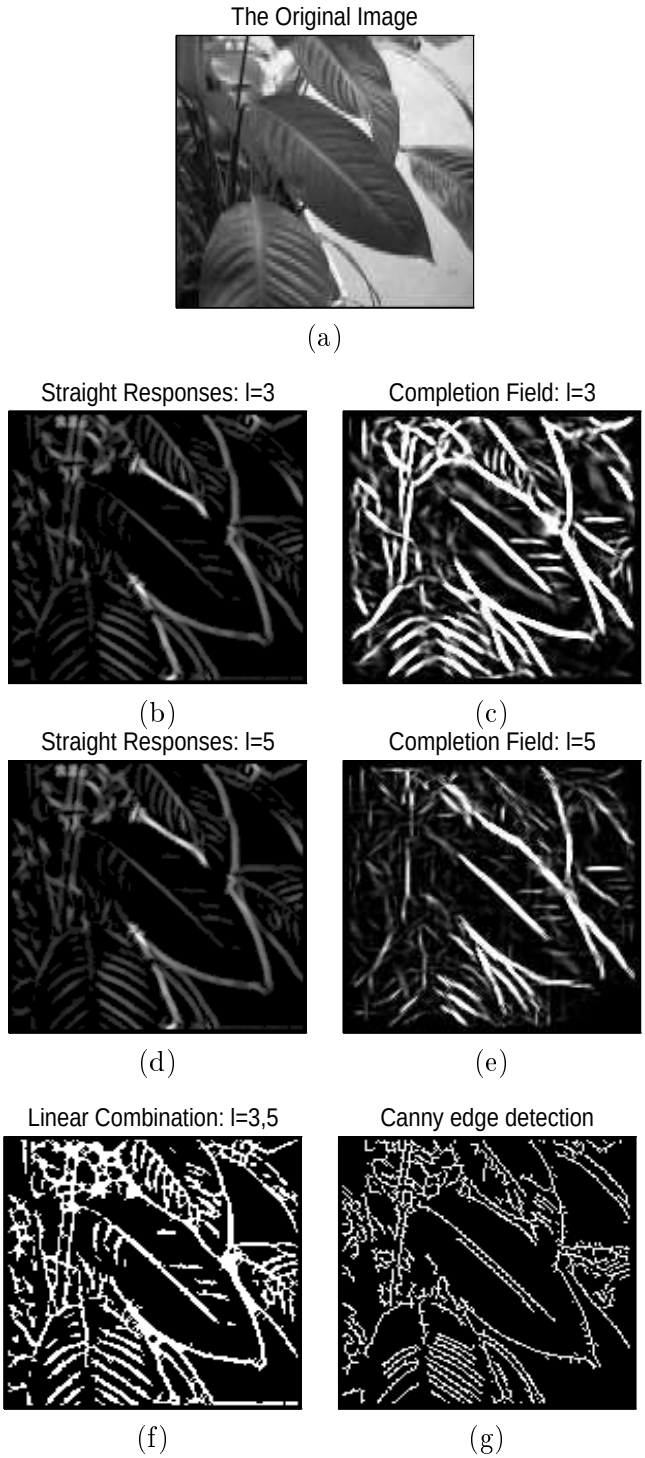

Figure 11: (a) The original image. (b)-(e) Completion fields in length scales 3 and 5 (width 1). (f) This picture shows the application of a skeleton detector to a linear combination of the two completion fields (after thresholding). (g) Canny edge detection with SD=2 and T=0.2. Images are of size 144 x 144 pixels.

6 Conclusion

Important problems in perceptual grouping are the detection of smooth curves in images and their completion over gaps. In this paper we have simplified the computation involved in the process of completion, exploiting the smoothness of the solution to the problem, and have defined affinity measures for completion that take into a proper account the scale of edge elements. In particular, we have introduced new, closed-form approximations for the elastica energy functional and presented a fast numeric method to compute the curve of least energy. In this method the error decreases exponentially with the number of discrete elements. We then have used our approximations to define an affinity measure which takes into account the width and length of the edge elements by considering the range of curvatures that can be detected with corresponding filters of the same scale. Finally, we have shown that solutions to the problem of finding the most likely completions in an image can be implemented using a multigrid algorithm in time that is linear in the number of discrete edge elements in the image. This last observation applies also to recent methods for completion and salience [10, 30]. In the future we intend to use the multigrid algorithm to simultaneously detect completions at different scales in order to combine these completions into a single saliency map.

Appendix

A Minimization over C_{12}

Below we compute the minimum of Eq. (6). It will be shown, using calculus of variations (see, e.g., [7]), that :

$$\min_{g \in C_{12}} \int_0^r (g'')^2 dx = \frac{4}{r}(t_1^2 + t_2^2 - t_1 t_2) \ ,$$
$$\text{where } t_i = \tan \Phi_i \ ; \ i = 1, 2 \ .$$

- Proof :

  Assume first that there exists a smooth enough function f for which the minimum is attained. Consider functions h for which $\exists g, \tilde{g} \in C_{12}$ such that $h = g - \tilde{g}$ (i.e., "test" functions, satisfying $h(0) = h(r) = h'(0) = h'(r) = 0$.) A necessary condition for f being an extremal function in C_{12} over which a functional J assumes its minimum is that for any such test function h

  $$\frac{d}{dt}(J(f + th)) = 0 \quad \text{at} \quad t = 0 \ ,$$

  where t is a real parameter.
  Since

  $$\frac{d}{dt}(J(f + th) =$$

$$\frac{d}{dt}(\int_0^r (f'' + th'')^2 dx) =$$
$$\frac{d}{dt}(\int_0^r ((f'')^2 + 2tf''h'' + t^2(h'')^2) dx) =$$
$$2\int_0^r f''h'' dx + 2t\int_0^r (h'')^2 dx$$

we get from the above necessary condition that $\int_0^r f''h'' dx = 0$.

Integrating the last equation twice by parts, and noting that both h and h' vanish on the boundaries, we get that :

$$0 = \int_0^r f''h'' dx = -\int_0^r h'f^{(3)} dx = \int_0^r hf^{(4)} dx \ . \tag{22}$$

Hence, by the arbitrariness of h, we immediately get that :

$$f^{(4)}(x) \equiv 0.$$

Therefore, by considering f(0)=f(r)=0 we get that the extremal function is nothing else than the cubic Hermite spline

$$f(x) = x(x-r)(ax+b) \ ,$$

and by considering also $f'(0) = t_1$ and $f'(r) = -t_2$ we solve for a and b and get that

$$f(x) = x(x-r)(\frac{1}{r^2}(t_1 - t_2)x - \frac{t_1}{r}) \ .$$

Consequently,

$$\int_0^r (f''(x))^2 dx = \frac{4}{r}(t_1^2 + t_2^2 - t_1 t_2) \ .$$

Now, we will show that the above derived f is indeed the global minimum of the functional J, over C_{12}. Take any $g \in C_{12}$, and define $h \stackrel{def}{=} g - f$. Evidently, this h qualifies as a test function ($h(0) = h(r) = h'(0) = h'(r) = 0$). It was already shown above that

$$\frac{d}{dt}(J(f+th) = \int_0^r f''h'' dx + t\int_0^r 2(h'')^2 dx \ .$$

Note that since $f^{(4)}(x) \equiv 0$ we have by Eq. 22 that

$$\int_0^r f''h'' = 0 \ ,$$

and therefore

$$\frac{d}{dt}(J(f+th) = t\int_0^r 2(h'')^2 dx \ .$$

Assuming that $\int_0^r (h'')^2 dx > 0$ (that is, $\int_0^r (g''-f'')^2 dx > 0$) we get that $J(f+th)$, as a scalar function of the real parameter t, assume its unique minimum at $t=0$, and in particular we have that

$$J(f) < J(g) \ .$$

If on the other hand $\int_0^r (g''-f'')^2 dx = 0$ then according to the boundary conditions we have that $g(x) \equiv f(x)$.

□

B Analytic approximation of the "Simpson method" solution

The "$(n=0, \bar{n}=2)$-system" (see Section 4) using the Simpson integration rule ($p=3$) yields for the unknowns λ and $\widetilde{\Psi}_0$ the following two equations

$$\begin{aligned} -2a_0 + \lambda \cos \widetilde{\Psi}_0(\frac{1}{2}) &= 0 \\ \sin \widetilde{\Psi}_0(0) + 4\sin \widetilde{\Psi}_0(\frac{1}{2}) + \sin \widetilde{\Psi}_0(1) &= 0 \ , \end{aligned}$$

which by the form of $\widetilde{\Psi}_0$ essentially means that

$$\begin{aligned} -2a_0 + \lambda \cos\left[\frac{\Psi_1+\Psi_2}{2} + \frac{a_0}{4}\right] &= 0 \\ \sin \Psi_1 + 4\sin\left[\frac{\Psi_1+\Psi_2}{2} + \frac{a_0}{4}\right] + \sin \Psi_2 &= 0 \ . \end{aligned}$$

Assuming small $|\Psi_1+\Psi_2|$ (i.e. $\Phi_1 \simeq \Phi_2$), approximating $\sin(\cdot)$ using its Taylor expansion around Ψ_1 to a first order we get the relation

$$\sin\Psi_1 + \sin\Psi_2 \simeq \sin(\Psi_1+\Psi_2)\cos\Psi_1 \ ,$$

which when introduced into the following representation of the last equation becomes

$$\left[\frac{\Psi_1+\Psi_2}{2} + \frac{a_0}{4}\right] = \arcsin\left[-\frac{\sin\Psi_1+\sin\Psi_2}{4}\right] \ ,$$

resulting in

$$a_0 \simeq -(2+\cos\Psi_1)(\Psi_1+\Psi_2) \ .$$

This implies that

$$\begin{aligned} \widetilde{\Psi}_0 &\approx ((2+\cos\Psi_1)(\Psi_1+\Psi_2))\, s^2 \\ &- ((3+\cos\Psi_1)\Psi_1 + (1+\cos\Psi_1)\Psi_2)\, s + \Psi_1 \ . \end{aligned}$$

Assuming now that $|\Psi_1| + |\Psi_2|$ is small then

$$a_0 \simeq -3(\Psi_1 + \Psi_2) \ .$$

Hence, $\widetilde{\Psi}_0$ is well approximated by

$$\bar{\Psi}(s) \stackrel{def}{=} 3(\Psi_1 + \Psi_2)s^2 - (4\Psi_1 + 2\Psi_2)s + \Psi_1 \ ,$$

for which it is straightforward to show that

$$\Gamma_{inv}(\bar{\Psi}(s)) = 4(\Psi_1^2 + \Psi_2^2 + \Psi_1\Psi_2) \ .$$

C Analysis of the affinity measure used in [30, 26]

In [17, 30] two induction fields were defined, the source field and the sink field. The source field represents the probability of a particle to travel from a source element to all other elements in the image, and the sink field represents the probability of a particle to travel from any element in the image to a sink element. These fields are induced by the affinity measure based on the quantity $P(2|1)$ which represents the probability of the particle to start at an element e_1 at time t_1 and arrive at an element e_2 at time t_2. In fact, $t_{21} = t_2 - t_1$, and the affinity measure between e_1 and e_2 is of the form

$$\bar{A}(e_1, e_2) = \int_{t_1}^{\infty} e^{-t_{21}/\tau_{21}} P(2|1) dt_2 \ ,$$

where the strictly positive parameter τ_{21} is accounted for the decay of particles. Each element of the field of likely completions is obtained by the product of the sum of all source fields at that element with the sum of all sink fields at that element.

Thornber and Williams [26] derived an explicit expression for the summation kernel, $P(2|1)$

$$-log(P(2|1)) = 2log\left(\frac{t_{21}^2 \pi T}{\sqrt{3}}\right) + \frac{6}{T t_{21}}(Q_x + Q_y) \tag{23}$$

where $Q_x = Q_x(x_2 - x_1, \dot{x}_1, \dot{x}_2, t_{21})$, $Q_y = Q_y(y_2 - y_1, \dot{y}_1, \dot{y}_2, t_{21})$, and the strictly positive parameter T are all defined in [26]. We assume that $\|(\dot{x}_1, \dot{y}_1)\| = \|(\dot{x}_2, \dot{y}_2)\| = 1$, so that $\dot{x}_1 = \cos\Psi_1, \dot{y}_1 = \sin\Psi_1, \dot{x}_2 = \cos\Psi_2$, and $\dot{y}_2 = \sin\Psi_2$ (see Fig. 4).

Note that when $|\Phi_1| + |\Phi_2|$ is sufficiently small, having the particle traveling in constant unit speed implies that $P(2|1)$ assumes its significant value at $t_{21} \simeq r$, so that the affinity measure typically behaves as the value of its integrand at $t_{21} = r$. For $t_{21} = r$, we can approximate

$$Q_x + Q_y \simeq \frac{5}{3} - \cos\Phi_1 - \cos\Phi_2 + \frac{1}{3}\cos(\Phi_1 + \Phi_2) \ ,$$

from which by approximating $\cos(\cdot)$ using its Taylor expansion to a second order we get

$$Q_x + Q_y \simeq \frac{1}{3} E_{ang}(\Phi_1, \Phi_2) \ .$$

Substituting this in Eq. (23) and rearranging, we get for $t_{21} = r$ that :

$$P(2|1) \simeq (\frac{\sqrt{3}}{r^2 \pi T})^2 e^{-\frac{6}{Tr} E_{ang}(\Phi_1, \Phi_2)} \ . \tag{24}$$

Under the above assumptions we can evaluate the affinity measure $\bar{A}(e_1, e_2)$ by the value of its integrand at $t_{21} = r$, that is

$$\bar{A}(e_1, e_2) \simeq e^{-r/\tau_{21}} (\frac{\sqrt{3}}{r^2 \pi T})^2 e^{-\frac{2}{Tr} E_{ang}(\Phi_1, \Phi_2)} \ , \tag{25}$$

which practically implies that

$$\bar{A}(e_1, e_2) \simeq A(e_1, e_2) = e^{-r/r_0} e^{-E_{ang}(\Phi_1, \Phi_2)/(r\sigma_0)} \ , \tag{26}$$

for some strictly positive a-priori set parameters r_0 and σ_0.

Note that due to our analysis in Section. 3, when we consider the field induced by an element $e_1 = (P_1, \Psi_1)$, according to the above $A(e_1, e_2)$, the strongest effect at a position P_2 is in the orientation $\Phi_2 = \frac{1}{2}\Phi_1$. We may change this preference by using the generalized form of E_{ang} (Eq. 12) and set the coefficients a and b to fit the desired preference. Also, note that when Φ_1 and Φ_2 are held fixed and r is increased, we see that $e^{-E_{ang}/(r\sigma_0)}$ increases within $A(e_1, e_2)$. This property of the induction also arises from the property of the classical elastica measure Γ_{el} (as is explained in Sec. 2), and has the following intuitive explanation: for any fixed Φ_1 and Φ_2, a particle changing its orientation from Φ_1 to Φ_2 turns less per unit length when r increases.

Finally, we would like to note that although the stochastic completion fields do not explicitly show a preference for scale, such a preference nevertheless arises. By taking the derivative of $A(e_1, e_2)$ with respect to r we can see that when we hold Φ_1 and Φ_2 fixed the strongest affinity between two elements is obtained when

$$r = \sqrt{E_{ang}} \sqrt{\frac{r_0}{\sigma_0}} \simeq \Omega \sqrt{\frac{r_0}{\sigma_0}}$$

(see Sec. 5.2 for Ω).

D Asymptotic smoothness of the kernel G

To show the "asymptotic smoothness" of the kernel $G = G(e_i; e_j) = G(x_i, y_i, \Psi_i, x_j, y_j, \Psi_j)$ note first that due to its symmetry ($G(e_i, e_j) = G(e_j, e_i)$) we can consider only derivatives with respect to x_j, y_j, Ψ_j, and due to its translation and rotation invariance we can fix (x_i, y_i, Ψ_i) conveniently to be $(0, 0, 0)$.

That is, for the current discussion we consider

$$\begin{aligned} G(x,y,\Psi) &= G(0,0,0,x,y,\Psi) \\ &= F_d\left(r/\rho\right) F_t\left(\rho/r\sqrt{\Phi_1^2+\Phi_2^2-\Phi_1\Phi_2}\right) \ , \end{aligned}$$

where $r=\sqrt{x^2+y^2}$, $\Phi_1=\theta=\arctan y/x$, $\Phi_2=\Psi_2-\theta$,

$$F_d(\xi)=1/\xi,\ F_t(\eta)=e^{-\alpha_2\eta^{\beta_2}},\ \alpha_2,\beta_2\in\Re_+ \ .$$

It can be shown that any q-order partial derivative of $G(x,y,\Psi)$ decays fast with r. For employing the multigrid fast summation algorithm it is however enough to consider only *pure* (not mixed) partial derivatives.

Let us focus on $\partial_x^q G$ as representing also $\partial_y^q G$. Having $G=F_dF_t$, we have for any natural number q that

$$\partial_x^q G=\sum_{k=0}^{q} c_k\left(\partial_x^k F_d\right)\left(\partial_x^{q-k}F_t\right) \ ,$$

and so, we can focus on $\partial_x^q F_t$ and $\partial_x^q F_d$ separately. Let us first concentrate on $\partial_x^q F_t$. Considering θ as defined above (see also Fig. 4) we have that $F_t(x,y,\Psi)=F_t(r,\theta,\Psi)$, and therefore

$$\partial_x F_t=\left(\partial_\theta F_t\right)\partial_x\theta+\left(\partial_r F_t\right)\partial_x r \ .$$

Furthermore, it is inductively evident that for any order q we have the same type of expression for $\partial_x^q F_t$ as for $q=1$ above, for instance, for $q=2$

$$\begin{aligned} \partial_x^2 F_t &= \left(\partial_{x\theta}F_t\right)\partial_x\theta+\left(\partial_\theta F_t\right)\partial_x^2\theta+ \\ &\quad \left(\partial_{xr}F_t\right)\partial_x r+\left(\partial_r F_t\right)\partial_x^2 r \\ &= \left(\partial_\theta^2 F_t\right)\left(\partial_x\theta\right)^2+\left(\partial_\theta F_t\right)\partial_x^2\theta+ \\ &\quad \left(\partial_r^2 F_t\right)\left(\partial_x r\right)^2+\left(\partial_r F_t\right)\partial_x^2 r+ \\ &\quad 2\left(\partial_{r\theta}F_t\right)\left(\partial_x r\right)\left(\partial_x\theta\right). \end{aligned} \tag{27}$$

Considering the fast decay of $\partial_x^q F_t$ with r, q being any natural number, note first that

$$\begin{aligned} \partial_x r &= x/r \ , \\ \partial_x^2 r &= 1/r-x^2/r^3 \ , \ \ldots \ , \end{aligned}$$

and hence, inductively, $\exists M_q$, independent of r, such that $\left|\partial_x^k r\right|\le M_q/r^{k-1}$, $\forall\ 0\le k\le q$. Also note that

$$\begin{aligned} \partial_x\theta &= -y/r^2 \ , \\ \partial_x^2\theta &= 2yx^2/r^5 \ , \ \ldots \ , \end{aligned}$$

and thus, it is inductively evident that there exists C_q and $\widetilde{C}_q$, both independent of r, such that $|\partial_x \theta|^k \leq C_q/r^k$, $\forall\ 0 \leq k \leq q$, and $\left|\partial_x^k \theta\right| \leq \widetilde{C}_q/r^k$, $\forall\ 0 \leq k \leq q$.

Regarding $\partial_\theta^q F_t$, for natural numbers q, setting $\beta_2 = 2$ (see the definition of $F_t(\eta)$ above), note that

$$\partial_\theta F_t = (\partial_{\Phi_1} F_t)\,\partial_\theta \Phi_1 + (\partial_{\Phi_2} F_t)\,\partial_\theta \Phi_2 = \partial_{\Phi_1} F_t - \partial_{\Phi_2} F_t\ .$$

Focusing on $\partial_{\Phi_2}^q F_t$, and denoting $g(\Phi_2) = \alpha_2 \rho^2(\Phi_1^2 + \Phi_2^2 - \Phi_1\Phi_2)$, we have that $F_t = e^{-g(\Phi_2)/r^2}$, and therefore

$$\begin{aligned}
\partial_{\Phi_2} F_t &= \left(-g'(\Phi_2)/r^2\right) F_t\ , \\
\partial_{\Phi_2}^2 F_t &= \left(-g''(\Phi_2)/r^2 + (g'(\Phi_2)/r^2)^2\right) F_t\ ,\ \ldots\ ,
\end{aligned}$$

which by taking into account the fact that $\partial_{\Phi_2}^k g(\Phi_2) = 0$, $\forall k \geq 3$, yields inductively that $\exists N_q$ such that $\left|\partial_{\Phi_2}^k F_t\right| \leq N_q/r^{k+1}$, $\forall\ 1 \leq k \leq q$. Focusing now on the more general case of $\partial_{\Phi_1}^{k_1}\partial_{\Phi_2}^{k_2} F_t$, where k_1 and k_2 are natural numbers, and $k = k_1 + k_2$, it is obvious by the same reasoning applied to $\partial_{\Phi_2}^q F_t$ above, that $\exists \widetilde{N}_q$ such that $\left|\partial_{\Phi_1}^{k_1}\partial_{\Phi_2}^{k_2} F_t\right| \leq \widetilde{N}_q/r^k$, $\forall\ 0 \leq k \leq q$. Regarding $\partial_r^q F_t$, note that

$$\begin{aligned}
\partial_r F_t &= -2/r^3 g F_t\ , \\
\partial_r^2 F_t &= \left((-2/r^3 g)^2 + 6/r^4 g\right) F_t\ ,\ \ldots\ ,
\end{aligned}$$

and therefore, inductively, $\exists T_q$ such that $\left|\partial_r^k F_t\right| \leq T_q/r^{k+2}$, $\forall\ 1 \leq k \leq q$. Consequently, $\exists \widetilde{T}_q$ such that

$$\left|\partial_r^{k_1}\partial_\theta^{k_2} F_t\right| \leq \widetilde{T}/r^{k+2}\ ,$$

where k_1 and k_2 are natural numbers, and $k = k_1 + k_2$. Thus, from Eq. 27, and the above derived bounds, we get inductively that for any natural number q, $\exists \widetilde{M_q}$, such that

$$|\partial_x^q F_t| \leq \widetilde{M_q}/r^q\ .$$

Now, as for $\partial_x^q F_d$, having $F_d = \rho/r$, we obviously have that for any natural number q, $\exists \widetilde{L_q}$, independent of r, such that $|\partial_x^q F_d| \leq \widetilde{L_q}/r^{q+1}$.

As for $\partial_\Psi^q G$, note that since $\Psi = \Phi_2 + \theta$ we have that

$$\partial_\Psi^q G = F_d \partial_\Psi^q F_t = F_d \partial_{\Phi_2}^q F_t\ .$$

Thus, concluding from all the above, we get that for any natural number q : $\exists L_q$, independent of r, such that $|\partial_x^q G(x,y,\Psi)| \leq L_q/r^q$, and $\left|\partial_y^q G(x,y,\Psi)\right| \leq L_q/r^q$, and $|\partial_\Psi^q G(x,y,\Psi)| \leq L_q/r^q$.

The fast decay of the derivatives of G with increasing r ensures that when considering the sum of inductions on an element e_i by all other elements we can aggregate the *many far away*

inductions into much fewer inducing representative elements, with as small an error as desired in the total summation by employing a sufficiently high q-order aggregation. In addition, when summing the neighboring (w.r.t. r) inductions on e_i, one should use the smoothness of G with respect to Ψ in order to avoid again summing the *many differently oriented* inductions, this again by employing a q-order aggregation of the many differently oriented inductions into fewer representative, controlling the error by the order of aggregation. In order to use the smoothness of G with respect to Ψ for q-order aggregation one should note that the q-order derivatives of G with respect to Ψ decays fast.

In addition to the q-order aggregation described above one should also use for fast summation the fast decay of $G(x_i, y_i, \Psi_i, x_j, y_j, \Psi_j)$ with respect to *any* of its six arguments in order to sum all inductions acting upon certain neighborhoods only into some appropriate representatives of these neighborhoods, interpolating from these representative sums the induction acting upon any desired element, while controlling the error by the order of interpolation.

Employing fast summation of induction fields (see Sec. 5.3), all of the above considerations may and should be exploited within different scales of spatial and angular farness between elements. For instance, given a scaled radius of curvature $\rho(l, w)$, the fast decay of the derivatives of the kernel G should be considered with respect to the scaled spatial distance r/ρ.

References

[1] T. D. Alter, and R. Basri, "Extracting salient curves from images: an analysis of the saliency network," *International Journal of Computer Vision*, forthcoming.

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

[3] A. Brandt, "Multilevel computations of integral transforms and particle interactions with oscillatory kernels," *Comp. Phys. Comm.*, **65**:24–38, 1991.

[4] A. Brandt and J. Dym, "Fast computation of multiple line integrals," *SIAM J. Sci. Comp.*, to appear.

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

[6] J. Dolan and R. Weiss, "Perceptual Grouping of Curved Lines," *Proc. IUW89*:1135–1145, Palo Alto, CA., 1989.

[7] L. E. El'sgol'ts, "Calculus of variations," *International series of monographs on pure and applied mathematics*, Pergamon press, Vol. 19, 1961

[8] D. J. Field, A. Hayes, and R. F. Hess, "Contour Integration by the Human Visual System: Evidence for a Local "Association Field"," *Vision Res.*, **33**(2):173–193, 1993.

[9] S. Grossberg and E. Mingolla, "The Role of Illusory Contours in Visual Segmentation," *The Perception of Illusory Contours*, Petry, S. and Meyer, G. (eds.), Springer-Verlag, New York:116–125, 1987.

[10] G. Guy and G. Medioni, "Inferring Global Perceptual Contours from Local Features," *Proc. DARPA Image Understanding Workshop*:881–892, 1993.

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

[12] L. Herault and R. Horaud, "Figure-ground discrimination, a combinatorial optimization approach," *IEEE Pattern Analysis and Machine Intelligence*, **15**, 1993

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

[14] G. Kanizsa, "Organization in Vision," Praeger, New York, 1979.

[15] M. K. Kapadia, M. Ito, C. D. Gilbert, and G. Westheimer, "Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys," *Neuron*, **15**:843–856, 1995.

[16] P. J. Kellman, and T. F. Shipley, "A theory of visual interpolation in object perception," *Cognitive Psychology*, **23**:141–221, 1991

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

[18] M. Nitzberg and D. Mumford, "The 2.1-D Sketch," *Proc. of the 3rd Intl. Conf. on Computer Vision*:138–144, Osaka, Japan, 1990.

[19] C. E. Pearson (ed.), "Handbook of applied mathematics - Second Edition," Van Nostrand Reinhold Company, New York, 1983.

[20] P. Parent and S. W. Zucker, "Trace inference, curvature consistency and curve detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**:823–889, 1989.

[21] U. Polat and D. Sagi, "Lateral Interactions Between Spatial Channels: Suppression and Facilitation Revealed by Lateral Masking Experiments," *Vision Res.*, **33**(7):993–999, 1993.

[22] N. Rubin, R. Shapley, and K. Nakayama, "Rapid propagation speed of signals triggering illusory contours," *Invest. Ophth. and Vis. Sci. Suppl.*, **36**:1037, 1995.

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

[24] S. Sarkar and K. Boyer, "Quantitative Measures of Change based on Feature Organization: Eigenvalues and Eigenvectors," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*:478–483, San Francisco, CA., 1996.

[25] A. Sha'ashua and S. Ullman, "Structural Saliency: The Detection of Globally Salient Structures Using a Locally Connected Network," *2nd ICCV*:321–327, 1988,

[26] K. K. Thornber and L. R. Williams, "Analytic Solution of Stochastic Completion Fields," *Biological Cybernetics*, **75**:141–151, 1996.

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

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

[29] L. R. Williams, and A. R. Hanson, "Perceptual Completion of Occluded Surfaces," *Proc. of IEEE Computer Vision and Pattern Recognition*, Seattle, WA., 1994.

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

[31] L. R. Williams, and D. W. Jacobs, "Local Parallel Computation of Stochastic Completion Fields," *CVPR*:161–168, 1996.

[32] S. Yen and L. Finkel, "Pop-Out of Salient Contours in a Network Based on Striate Cortical Connectivity," *Investigative Opthalmology and Visual Science (ARVO)*, **37**(3), 1996.

[33] S. W. Zucker, and R. A. Hummel, "Toward a Low-Level Description of Dot Clusters: Labeling Edge, Interior, and Noise Points," *CGIP*, **9**(3):213–234, 1979.

[34] S. W. Zucker, C. David, A. Dobbins, and L. Iverson, "The Oraganization of Curve Detection: Coarse Tangent Fields and Fine spline Coverings," *Proc. ICCV88*:568–577, Tampa, FL.,Dec. ,1988.