A Tangent Bundle Theory for Visual Curve Completion

Guy Ben-Yosef, *Student Member*, *IEEE*, and Ohad Ben-Shahar, *Member*, *IEEE*

Abstract—Visual curve completion is a fundamental perceptual mechanism that completes the missing parts (e.g., due to occlusion) between observed contour fragments. Previous research into the *shape* of completed curves has generally followed an "axiomatic" approach, where desired perceptual/geometrical properties are first defined as axioms, followed by mathematical investigation into curves that satisfy them. However, determining psychophysically such desired properties is difficult and researchers still debate what they should be in the first place. Instead, here we exploit the observation that curve completion is an early visual process to formalize the problem in the *unit tangent bundle* $\mathbf{R}^2 \times \mathcal{S}^1$, which abstracts the primary visual cortex (V1) and facilitates exploration of basic principles from which perceptual properties are later *derived* rather than imposed. Exploring here the elementary principle of *least action* in V1, we show how the problem becomes one of finding minimum-length admissible curves in $\mathbf{R}^2 \times \mathcal{S}^1$. We formalize the problem in variational terms, we analyze it theoretically, and we formulate practical algorithms for the reconstruction of these completed curves. We then explore their induced visual properties vis-à-vis popular perceptual axioms and show how our theory predicts many perceptual properties reported in the corresponding perceptual literature. Finally, we demonstrate a variety of curve completions and report comparisons to psychophysical data and other completion models.

Index Terms—Visual completion, curve completion, tangent bundle, inpainting.

1 Introduction

Visual curve completion is a perceptual phenomenon in which the visual system fills in the missing parts between boundary fragments to facilitate the perception of complete objects. When the object is fragmented due to occlusion (Figs. 1A and 1B), the completion is usually called *amodal* [25]. When the object is illusory and its completed boundary curves are subjective (Fig. 1C), the completion is known as *modal*.[1] In the latter case, observers typically report clear intensity edges along the subjective contours and the perception of an illusory shape that is slightly brighter than the surrounding regions [8], [25], [37]. In both amodal and modal completion, a significant part of the process is considered low-level and local, affected only a little by context (Fig. 1D) or visual experience (Fig. 1E). Like other problems of perceptual organization [37], it too has been studied from different perspectives of the vision sciences, and in particular by the computational, perceptual, and neurophysiological communities.

Due to its low-level and local characteristics, studies of curve completion usually assume that the completed curve is *induced* by the two oriented line segments at the point of occlusion (hereinafter the *inducers*). Referring to these inducers, the curve completion problem is typically divided into two parts. First, the *grouping problem* deals with the organization of candidate inducers into pairs that indeed induce completed curves (see Fig. 2A). The second, the *shape problem*, deals with the characterization and reconstruction of the shape of the completed curve between two given inducers. In this paper, we focus on the latter problem, with the goal of formalizing and solving the shape of perceptually completed curves between two given inducers.

Suppose we are given an image region where boundaries are missing (e.g., blue circle in Fig. 1A) and the two inducers between which the curve is completed (e.g., the white segments in Fig. 1A). Clearly, there are infinitely many ways to complete a curve between these two inducers (some possible completions are shown in yellow in Fig. 1A), from which our visual system appears to choose a singular and unique shape (illustrated in magenta in Fig. 1A). This (somewhat illusive) shape is the objective of computational curve completion in general and of this paper in particular:

Problem 1. *Given the position and orientation of two inducers* $p_0 = [x_0, y_0; \theta_0]$ *and* $p_1 = [x_1, y_1; \theta_1]$ *in the image plane, find the shape of the "correct" perceptual curve that passes between these inducers.*

While we consider as "correct" the completion that agrees the most with perceptual evidence (see below), the notion of "correctness" in the context of our completion task is clearly ill defined. Since formulating visual perception is difficult, one may consider other alternatives also. Most desirable, perhaps, is completing the *actual shape* of the occluded (or subjective) object at the time the image was captured. But

1. The term *modal* is often used in this context to indicate completed contours which are "phenomenally present with the usual characteristics of visual modality." On the other hand, the term *amodal* completion is used when the completed structure is "present in the perception but does not have the phenomenal characteristics of sensory modally." While both phenomena have been reported earlier, these descriptions are attributed to Kanizsa [25, pp. 194-195].

• *The authors are with the Department of Computer Science, Ben-Gurion University of the Negev, PO Box 653, Beer-Sheva 84105, Israel. E-mail: {guybeny, ben-shahar}@cs.bgu.ac.il.*

Manuscript received 27 Dec. 2010; revised 8 Sept. 2011; accepted 6 Nov. 2011; published online 30 Dec. 2011.
Recommended for acceptance by M. Lindenbaum.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2010-12-0985.
Digital Object Identifier no. 10.1109/TPAMI.2011.262.

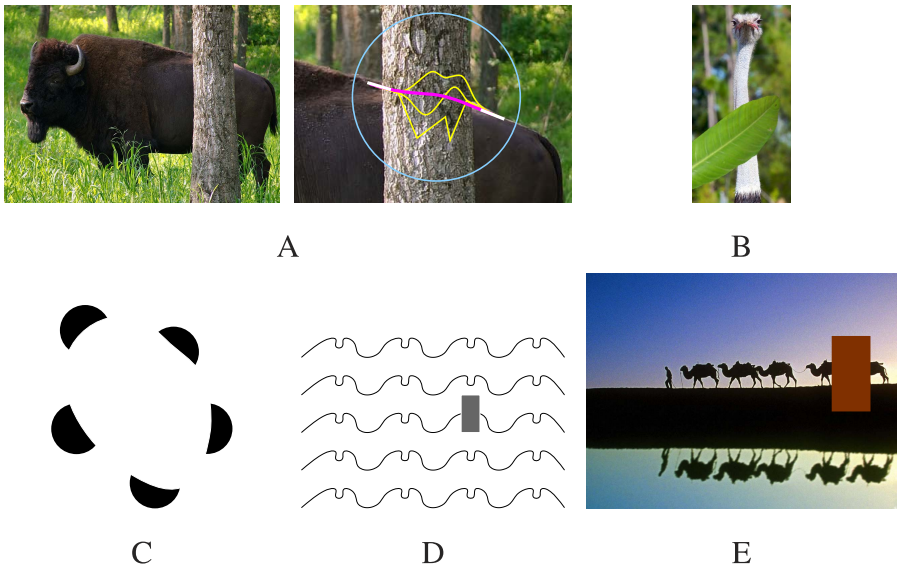


Fig. 1. Phenomenology of visual completion. (A) Curve completion is typically defined as the problem of constructing the single *perceptual* shape (magenta) between two given inducing segments (white), out of the infinitely many possible curves which connect these inducers (e.g., in yellow). (B) Another visual scene with amodal completion. (C) Modal completion that gives rise to an illusory blobby shape. (D) Curve completion is strong enough to override visual context and explicit knowledge of what the occluded part is. Here, most observers report the completion of a convex shape behind the occluder, although all contextual cues indicate the presence of a dent. (E) The same phenomenon of "seeing versus thinking" [25] can be demonstrated with familiar objects in photographs. Covering part of a caravan, the occluder here gives rise to the perception of an excessively long camel, one that every observer "knows" is impossible. Indeed, curve completion can override life-long visual experience.

clearly, this requires prior knowledge of the occluded shape (if nothing else, then at least for the evaluation of the completed shape), which renders the completion task meaningless. Clearly, while guessing the actual shape may be possible as part of the completion task (can you guess the shape of the missing boundaries in Fig. 1B?), the apparent arbitrariness of shapes makes this guess easily tricked (see Fig. 2B to see if your guess was correct).

Instead of the actual shape, one may seek the "typical" or likely shape of the occluded part of the object, which is a weaker but possibly more accessible constraint. Still, this too requires knowledge of the type and the part of the occluded object (to allow access to its typical shape in the occluded part), and hence its recognition or classification from the image data. Naturally, this conflicts with the early nature of the completion problem, and its accepted role as *facilitating* higher level tasks (like recognition) rather than being dependent on them (e.g., [27, page 159], [51, page 5], [56, page 837], or [39, page 47], to name but a few).

Given the observations and inherent difficulties above, much of the previous computational curve completion literature has incorporated *perceptual* considerations, either explicitly or implicitly (e.g., looking for the curve that is *chosen by the visual system* [51, page 1], modeling the shape, salience, and sharpness of *perceived* contours [56, page 838], or seeking the most "*pleasing*" curve [27, page 161] are just a few of the stated goals). Since formalizing the visual process is difficult, most previous studies have addressed the problem by defining constraints on the "desired" solution in the image plane. Unfortunately, however, almost all of

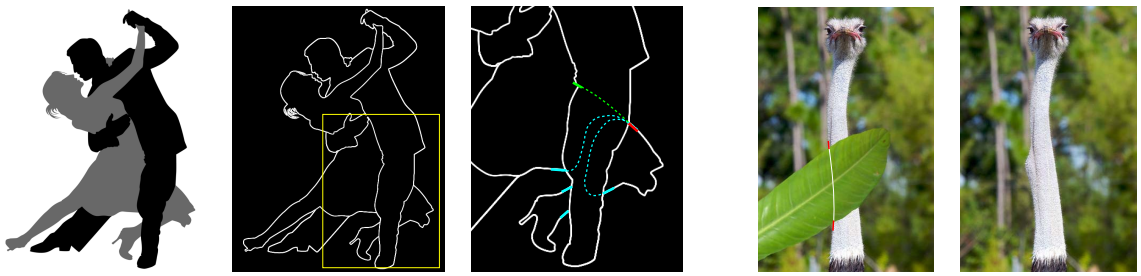


Fig. 2. The grouping and shape aspects of the curve completion problem. (A) The *grouping* phase deals with the grouping of inducers into pairs between which perceptually completed curves are formed. While in the selected region of this figure (also shown is its edge map) all cyan inducers are possible candidates for pairing with the red inducer (which leads to the possible completed curves such as those plotted in cyan), it seems that perceptual completion occurs only with the green inducer (which might produce the completion in green). (B) The *shape* problem deals with the inference of the completed shape between a given pair of inducers. As we argue, the "correct" computational completion should match the perceived one rather than the actual physical shape behind the occluder. Here, the ostrich's neck is occluded by a leaf and human observers tend to report completions similar to the one plotted in white. However, the apparent arbitrariness of physical shapes makes this guess easily tricked, as would be the case if at the moment there was a small piece of food in the ostrich's throat. Since we cannot predict the exact state of the occluded object, its actual shape and boundaries at the time the image is captured cannot serve as a measure of correctness for shape completion.

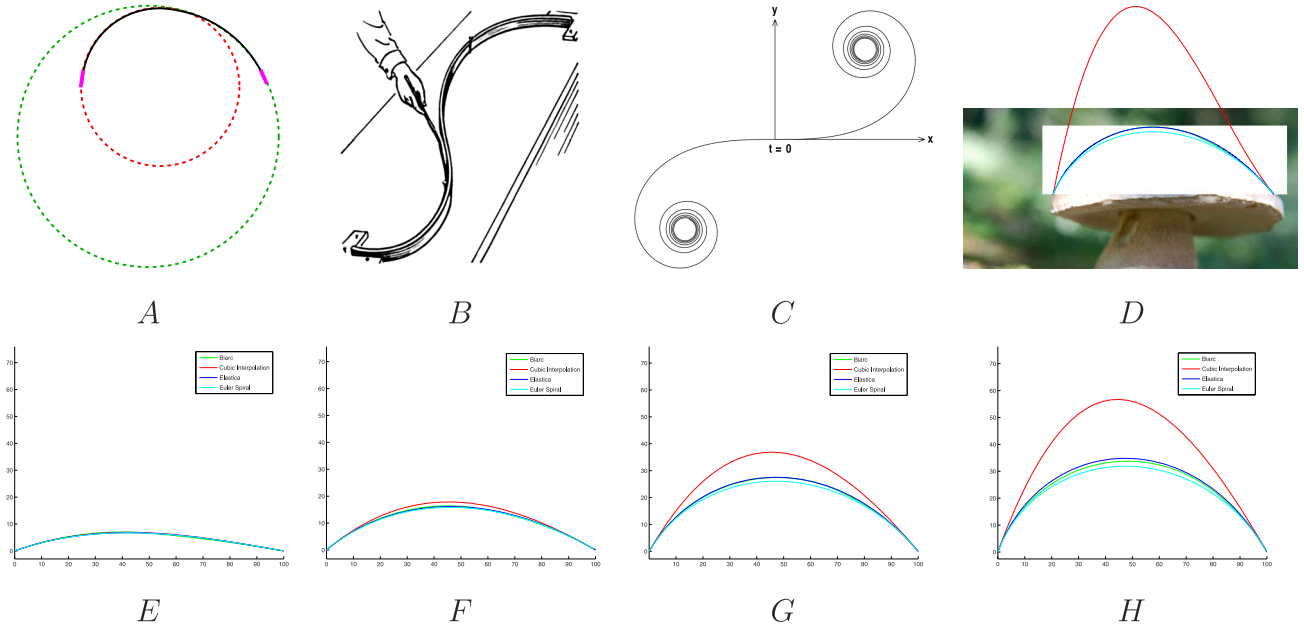


Fig. 3. Previous computational studies on curve completion. (A) The biarc curve (in black) connecting two inducers (in magenta) consists of two circular arcs (one in green and one in red) that are tangent both to an inducer and to each other. (B) Serving as inspiration for elastica models, a flat mechanical spline used by draftsmen in older days was used to determine the "smoothest" possible shape. The spline is fixed in two (or more) positions on the plate, in which it can be constrained to specific orientations as well (as in this illustration). Under these conditions, the spline takes a shape in which its total bending energy is minimal. Since the bending energy at each point is proportional to the squared curvature of the spline at that point, the shape of the spline is that of the elastica. (C) An example of an Euler spiral curve, whose curvature changes linearly with arclength, i.e., $\kappa(s) = a \cdot s + b$, for some constants a and b. **(D)-(H)** Results of our own MATLAB implementation to previous popular curve completion algorithms, given the position and orientation of two inducers: The Biarc model [42], [51] is plotted in green, the Cubic interpolation model [6] is plotted in red, the Elastica model [21] is plotted in blue, and the Euler spiral model [27] is plotted in cyan. Please use the electronic version to zoom in for a better view.

these constraints have been inspired more by intuition and mathematical elegance, and less by perceptual findings or neurophysiological principles. In the next section we review some of the computational models, the axiomatic approach which often motivates them, and the degree of compatibility to perceptual and neurophysiological evidence. We then turn to present a new and rigorous mathematical curve completion theory which is motivated by the structure of the visual system and uses a single basic principle of "least action" in the visual cortical space to provide, among other things, accurate explanations and predictions for existing perceptual evidence.

2 Previous Work

2.1 Computational Studies

Among the first to address the shape problem (Problem 1) in a computational framework was Ullman [51], whose seminal work suggested that completed curves are constrained by certain geometrical/perceptual properties:

- *Isotropy*—The completed curve should be invariant to rigid transformations.
- *Smoothness*—The completed curve is analytic (or in some cases, differentiable once).
- *Total minimum curvature*—The integral of curvature along the curve should be as small as possible.
- *Extensibility*—Any two arbitrary tangent inducers along a completed curve C should generate the same shape as the shape of the portion of C connecting them.

In seeking the curve that uniquely satisfies these axiomatic properties, Ullman suggested a *biarc curve*, according to which the completed shape between two inducers consists of two circular arcs, each tangent both to an inducer and to the other arc (Fig. 3A). Since the number of such biarc pairs is infinite, the selected pair is the one that generates the minimal total curvature. While the biarc model was designed to satisfy the four original axioms, it violates smoothness (in the strict sense, since it is differentiable only once), it does not minimize total curvature in general (it does so only within the family of biarc curves), and it was also shown to violate extensibility in some cases [6]. While Ullman did not present a closed-form solution to his biarc model, a mathematical solution based on a one-dimensional nonlinear optimization was introduced later by Rutkowski [42] (see Figs. 3D, 3E, 3F, 3G, and 3H for our own implementation). Importantly, however, Ullman did suggest a solver based on a parallel network of simple computational nodes, reminiscent of the computational infrastructure in the primary visual cortex.

This perspective taken by Ullman triggered a wealth of computational curve completion research which follow what we call the *axiomatic approach*—a quest for the unique completed curve that satisfies a specified set of *predefined* desired characteristics. Indeed, subsequent studies have suggested a variety of other desired axioms, such as

- *Scale invariance*—the completed shape should be independent of the viewing distance.
- *Roundedness*—the shape of a completed curve induced by two *cocircular* inducers should be a circle.

- *Total minimum change of curvature*—integral of the *derivative* of curvature along the curve should be as small as possible.

However, since some of the suggested axioms conflict with each other, the axiomatic approach has invigorated a continuous debate on the scope of each axiom and the "correct" set of perceptual axioms to be used in the first place. Part of our contribution in this paper is to show that perceptual properties of completed curves can be derived (rather than imposed) by applying basic principles in the computational space in which the completion is computed.

Perhaps the most studied axiom since Ullman's work is the axiom of total minimum curvature. Inspired by the shape of mechanical splines (see Fig. 3B) and centuries-long work done by mathematicians (see [32] for a mathematical historical account), the planar "curve of least energy" is known to minimize the total bending energy:

$$\int a + b\,\kappa(s)^2 ds, \tag{1}$$

where a and b are some constants. This family of curves, known as *"elastica,"* defies known analytical solutions, but its appealing geometrical properties have inspired polynomial approximations first in drafting [44] and computer aided geometric design [22], and later in typography [28], and visual curve completion [6]. Although they enjoy simplicity, closed-form formulation, and computational efficiency, these (cubic) polynomials deviate significantly from elastica in many cases (see Fig. 3), a fact that led to more rigorous investigation of elastica by Horn [21] and later Mumford [35]. Slightly different than the original elastica formulation, however, in the context of curve completion between two inducers, they sought the curvature-minimizing curve with *unrestricted length* that minimizes (1) (for $a = 0,\ b = 1$) and satisfies the boundary conditions of the two inducers as defined in Problem 1. The Euler-Lagrange equation applied to this functional leads to a differential equation that the solution must satisfy, which in arclength parameterization takes the form

$$\left(\frac{d\theta}{ds}\right)^2 = \frac{1}{c}\sin(\theta + \phi), \tag{2}$$

where θ is the tangential angle of the curve at each position [21]. To use this equation for solving the elastica problem one needs to resolve the two parameters c and ϕ. Since no closed-form analytical solution is known, Horn suggested a solution using *elliptic integrals* [21] while Mumford did so based on *theta functions* [35].

Although in many cases the elastica solution deviates only slightly from the biarc model ([21] and Fig. 3), theoretically it conflicts with both the scale-invariance and the roundedness properties. Hence, subsequent attempts to model curve completion have modified the axiom of total minimum curvature to address additional properties as well. In particular, Weiss [53] suggested a scale invariant elastica model that minimizes the functional $L\int_0^1 \kappa^2 ds$, where L is the (unspecified) length of the curve. While Weiss could not find a closed-form analytical solution to this problem, closed-form *approximations* for both the elastica and the scale-invariant elastica were derived by Sharon et al. [46]. Since the latter associates energy not only to total curvature but also to length, the results tend to be "flatter" than the original elastica shape. At the same time, scale invariant elastica complies with roundedness but still violates the axiom of extensibility.

In an attempt to incorporate as many axioms as possible into a single computational model that prioritizes the roundedness property, Kimia et al. [27] have suggested replacing the minimum total curvature axiom with the minimization of total *change* in curvature. This property is captured by the minimization of the functional $\int (\frac{d\kappa}{ds})^2 ds$ subject to the boundary conditions $[x_0, y_0; \theta_0]$ and $[x_1, y_1; \theta_1]$. This model immediately entails a class of curves in which the curvature changes linearly, often known in the mathematical literature as *Euler spirals* [33]. Yet, since there could be many Euler spirals connecting two inducers, Kimia et al. [27] disambiguated the solution by picking the shortest possible Euler spiral as the completed curve (see Fig. 3 for examples). In addition to the roundedness and smoothness that this model satisfies by construction, the Euler spiral model also complies with the axioms of extensibility and scale invariance. Unfortunately, however, the roundedness and scale invariance axioms were refuted psychophysically in recent perceptual literature (see below), a fact that undermines the validity of the Euler spiral as an appropriate completion model.

Given the dominance of axiomatic models to curve completion, it should be mentioned that a somewhat different, nonaxiomatic approach was taken by Williams and Jacobs [56] in their stochastic model to curve completion, which employs assumptions on the *generation process* rather than on the desired final shape. Indeed, although it was not verified against perceptual findings, in their work they have argued that the completed curve is the most likely random walk in a 3D discrete lattice of positions and orientations. Still different, though similarly nonaxiomatic (in the perceptual sense), are other approaches that model the generation process as a variational segmentation procedure (e.g., [43], [24]). Although our theory differs from all these previous nonaxiomatic approaches in all of the theory, the assumptions, and the implications, it does share the same philosophy that shape constraints can be *derived* from more basic principles rather than imposed as axioms. Unlike all of these nonaxiomatic models, however, we will also show that our theory entails completed shapes that agree and predict recent perceptual and psychophysical findings.

With the variety of curve completion axioms and curve completion models, the question of what the "correct" model is and what the "right" axioms to employ are still remains unanswered.[2] Surely addressing this question requires extensive psychophysical research. But since the different predictions by the different models are often (though not always) small (see Figs. 3D, 3E, 3F, 3G, and 3H) and since it is difficult to *measure* the exact shape of modally or amodally completed perceptual curves, psychologists

2. It should be mentioned that in addition to the explicit research on curve completion, other relevant work comes from the problem of curve integration and detection, and in particular, from the ecological investigation of curvilinear continuity in natural images (e.g., [14], [13], [9]). However, the frequent relation of these data to physical (rather than perceptual or illusory) contours, the common application of axiomatic visual properties that derive from this fact (e.g., scale invariance in [39]), and the relevance of such data mainly to short range interactions all make the ecological and statistical approach less constructive in our context.

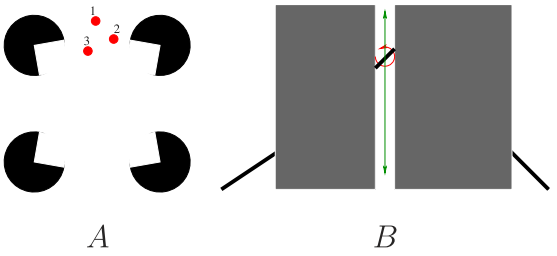


Fig. 4. Perceptual experimental studies on curve completion. (A) The dot localization paradigm (after [18]). A point probe (in red, numbered by the order of appearance) is presented for a short time on top of a modally or amodally completed figure. Observers are asked whether this point is localized either inside or outside the completed boundary and the distribution of responses is used to determine the likely completed shape. (B) The oriented probe localization method (after [11]). Observers view a completion scene and are asked to position and orient a segment probe in a narrow slit in the occluder.

have been focusing on other ways to characterize them, typically qualitatively rather than quantitatively. Some of these findings are reviewed in the next section.

2.2 Perceptual and Psychophysical Insights

The phenomenon of visual completion had been first reported more than a century ago by Schumann [45], and later gained much interest due to the extensive contributions and striking demonstrations by Kanizsa [25]. Traditionally, both modal and amodal completion are considered a result of the same visual mechanisms, or what has become known as the *identity hypothesis*. Advocated strongly by Kellman and Shipley [26] (who showed how modal configurations can be often reduced to amodal configurations), the identity hypothesis continues to be debated to this day (e.g., [47]). In this paper, we assume that both completion phenomena are a product of the same process, but we do note that there is nothing intrinsic in our theory that prevents their future separation.

Most of the *perceptual* research on curve completion, from Kanizsa and until recent years, has focused on the *grouping problem* (see Section 1). One popular theory in this context is the *relatability theory* [26], which suggests that two inducers could be part of the same completed contour if their linear extensions intersect in an acute outer angle. This condition was later shown to reflect the existence of a smooth curve with no inflection points connecting the two inducers [49].

In recent years, however, perceptual and psychophysical research on the *shape problem* has become increasingly more dominant, and in particular, some work has been focusing on measuring and characterizing *geometrical* features of visually completed curves using different experimental paradigms. For example, the *dot localization paradigm* (e.g., [18], [50], [15]) is a method where observers are asked to localize a point (or line) probe either inside or outside an amodally or modally completed boundary (Fig. 4A), and the distribution of responses is used to determine the likely completed shape. Often in these experiments the inducers are placed symmetrically in a cocircular configuration, which could be used to very explicitly examine the validity of the roundedness axiom. Somewhat surprisingly, and despite the popularity and the intuitive appeal of this axiom, recent results show that visually completed curves are somewhat flatter than circular arcs [18], [11], thus putting in doubt the validity of the biarc [51], the Euler spiral [27], and the scale-invariant elastica [46] models. It should be mentioned that in these experiments the distribution of results was similar for modal and amodal completions, hence supporting the identity hypothesis. It was further shown that the completed contour is perceived (i.e., constructed) as quickly as 120 msec, a result indicative of a preattentive early visual process.

In an alternative experimental paradigm, the *oriented probe localization* method [48], [11], [12], observers view a curve completion scene and are asked to localize the position and orientation of a probe to match the perceived completed contour (see Fig. 4B). Recent studies that employed either oriented probe localization, dot localization, or other paradigms have demonstrated that the perceptually completed curve becomes *flatter* as the distance between the inducers increases [15], [48], [11], [47, page 457], a dependency on scale that violates the scale invariance property that is a basic axiom in several of the completion models discussed above.

While even the most updated perceptual and psychophysical findings are still partial, they do indicate that none of the existing axiomatic models is valid vis-à-vis the operation of the human visual system. Since accurate perceptual data may be difficult to measure and quantify accurately (e.g., while scale dependency was demonstrated, the precision of present-day experimental paradigms is unlikely to resolve this dependency in quantitative terms), these studies also question the utility of the axiomatic approach as a whole. In this work, we therefore suggest focusing on theories in which perceptual properties of completed curves are derived from more basic (and nonperceptual) principles, rather than being imposed as axioms. In particular, while here we focus on the computational aspects of such a theory (and keep the full-fledged perceptual validation for future work), we do show how it qualitatively predicts these very same perceptual findings that existing models fail to replicate correctly.

2.3 Neurophysiological Insights

Since curve completion is a perceptual task performed by the visual system, it is worth reviewing basic neurophysiological aspects of the latter. The short time in which completed curves are constructed (e.g., [18], [40]) is indicative of an early vision process that takes place as early as the primary visual cortex (V1). Some properties of this cortical area indeed motivate much of our proposed theory.

The primary visual cortex (or V1) is one of the most intensively studied areas in the brain. The seminal work by Hubel and Wiesel [23] has indicated that V1 constitutes of *orientation selective cells* at all orientations (and at various scales) for all retinal positions, i.e., for each "pixel" in the visual field (see Fig. 5). This was captured by the so-called *ice cube* model, suggesting that V1 is continuously divided into full-range orientation *hypercolumns*, each associated with a different image (or retinal) position [23]. Hence, an image contour is represented in V1 as an activation pattern of all those cells that correspond to the oriented tangents along the curve's arclength (Fig. 5B). Subsequent work has shown that orientation selective neurons in two hypercolumns are able to interact via *long range horizontal connections* [41], [16], [5] to facilitate contextual computations that could

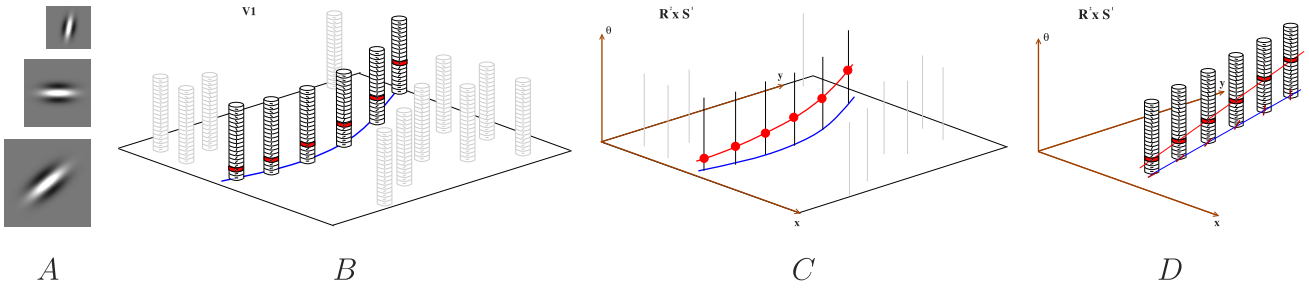


Fig. 5. The unit tangent bundle as an abstraction for the organization and mechanisms of the primary visual cortex. (A) Orientation selective cells in V1 have receptive fields of different scales and orientation tuning. (Shown here is the even-symmetric type only, coded by intensity.) (B) The primary visual cortex is organized in orientation hypercolumns [23], which implies that every retinal position is covered by neurons of all orientations. Thus, for each "pixel" p in the visual field we can think of a vertical array of orientation-selective cells extending over p and respond selectively (shown in red) according to the stimulus that falls on that "pixel." (C) The organization of V1 implies it can be abstracted as the unit tangent bundle $\mathbf{R}^2 \times \mathcal{S}^1$, where vertical fibers are orientation hypercolumns, and the activation pattern to image curve α (in blue) becomes "lifted" curve β (in red) such that $\alpha(t)$ and $\beta(t)$ are linked by the admissibility constraint in (3). The only aspect of $\mathbf{R}^2 \times \mathcal{S}^1$ not captured in this sketch is the periodicity of its orientation dimension. Thus, the reader should think of the plotted space as quotient space [55] $\mathbf{R}^3/\sim$, where the equivalence structure is such that $(x_1, y_1, z_1) \sim (x_2, y_2, z_2)$ if and only if $x_1 = x_2$, $y_1 = y_2$, and $|z_1 - z_2| = \pi$. (D) Not every curve in $\mathbf{R}^2 \times \mathcal{S}^1$ is admissible and, in particular, linear curves (e.g., in red) are usually inadmissible since their linear progression of the orientation (i.e., height) contradicts the "straightness" of the projected curve in the image plane (in blue).

provide local geometrical properties (e.g., curvature) of visual curves (e.g., [3]).

The participation of early visual neurons in the representation of curves is not limited to viewable curves only, and was shown to extend to completed or illusory curves as well. Among the first to show this were von der Heydt et al. [52], who have examined the activity of cells in the visual cortex of Macaque monkeys during presentation of modally completed shapes. They have found that approximately a third of the orientation selective cells in V2 (which are similar in structure to those in V1) fire when illusory contours move across their receptive field, and that many of these cells respond nearly the same to real and subjective contours. More recent studies by Grosof et al. [17] and Lee and Nguyen [31] have further extended these findings to area V1. All this evidence further supports the conclusion that curve completion is predominantly an early visual process that takes place as early as the primary visual cortex.

3 Curve Completion in the Tangent Bundle $\mathbf{R}^2 \times \mathcal{S}^1$

The basic neurophysiological findings mentioned above suggest that in the limit we can abstract orientation hypercolumn as infinitesimally thick "fibers," and place each of them at the position in the image plane that is associated with the hypercolumn. Doing so, one obtains an abstraction of V1 by the space $\mathbf{R}^2 \times \mathcal{S}^1$, as illustrated in Fig. 5. This space is an instance of a fundamental construct in modern differential geometry, the *unit tangent bundle* [36] associated with $\mathbf{R}^2$. A tangent bundle of a manifold S is the union of all tangent spaces at all points of S [36]. Similarly, a *unit* tangent bundle is the union of *unit* tangent spaces at all points of S, where each of these spaces constitutes unit tangent vectors only. In our case, it therefore holds [36] that:

Definition 1. *Let $I = \mathbf{R}^2$ the image plane. $T(I) \triangleq \mathbf{R}^2 \times \mathcal{S}^1$ is the unit tangent bundle of I.*

Recall our observation from Section 2.3 that an image contour is represented in V1 as an activation pattern of all those cells that correspond to the oriented tangents along the curve's arclength. Indeed, V1 represents image curves in a "lifted" fashion where both position and tangent orientation are represented *explicitly* along the path [20]. Given the $\mathbf{R}^2 \times \mathcal{S}^1$ continuous abstraction and remembering that the tangent orientation of a regular curve is a continuous function, we immediately observe that a regular image curve $\alpha(t)$ is represented as another regular curve $\beta(t)$ in $T(I)$ (Figs. 5B and 5C).

The last observation leads us to our main idea: If curve completion (like many other visual processes) is mostly an early visual process in V1 (cf., Section 2.3), and if V1 can be abstracted as the space $T(I)$, then perhaps the completion process should be investigated in *this* space, rather than in the image plane I. In this paper, we offer such a mathematical investigation whereby curve completion is carried out in $T(I)$, followed by projection to I. Part of our motivation for this idea is that unlike the debatable perceptual axioms in the image plane, the $T(I)$ space, as an abstraction of the cortical machinery, may offer more basic (and not necessarily perceptual) completion principles from which perceptual properties emerge as a consequence.

It should be mentioned that in addition to Williams and Jacobs [56] mentioned above, the space of positions and orientations in vision was already used in several cases for tasks such as curve integration, texture processing, curve completion, and image inpainting (e.g., [1], [2], [38], [7], [4], [19]). As will be discussed later, our contribution is quite different than the previous attempts to address visual completion in the roto-translational or the unit tangent bundle spaces—it is a variational approach rather than one based on diffusion [7], it provides simpler and more direct mathematical insights and facilitates the analysis of perceptual properties (unlike in [38], [7]), and it is unique in incorporating a relative scaling between the spatial and angular dimensions, a critical act for making these dimensions commensurable (unlike in [38], [7], [4]). All these issues are further elaborated in the rest of the paper.

3.1 Admissibility in $T(I)$

At first sight, curve completion in $T(I)$ may not be that different than curve completion in the image plane I, except, possibly, for the higher dimension involved. This intuition,

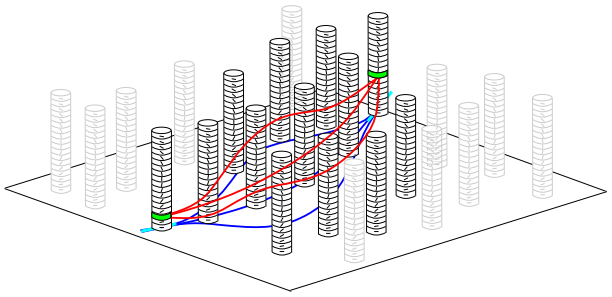


Fig. 6. Curve completion in the tangent bundle. When switching to $T(I)$, the curve completion problem becomes a problem of curve construction between boundary points (represented by the green cells) rather than between oriented inducers (plotted here in cyan). Similar to the problem in the image plane, in the tangent bundle there is also an infinite number of curves connecting two given endpoints (some of them are plotted here in red, while their projections to I are plotted in blue). Thus, as before, we will need additional constraints or principles to narrow down the set of solutions.

unfortunately, is incorrect. We first observe that when switching to $T(I)$, the curve completion problem (see Section 1) becomes a problem of curve construction between boundary *points* (e.g., as shown in green in Fig. 6), rather than between *oriented inducers* (as illustrated in cyan in Fig. 6). More importantly, as we now discuss, the completed curves in $T(I)$ cannot be arbitrary. In fact, the class of curves that we can consider in the first place is quite constrained.

Let $\alpha(t) = [x(t), y(t)]$ be a regular curve in I. Its associated curve in $T(I)$ is created by lifting α to $\mathbf{R}^2 \times \mathcal{S}^1$, yielding a curve $\beta(t) = [x(t), y(t), \theta(t)]$, which satisfies (using Newton's notation for differentiation)

$$\tan\theta(t) = \frac{\dot{y}(t)}{\dot{x}(t)}, \quad where \quad \dot{x}(t) \equiv \frac{dx}{dt}, \quad \dot{y}(t) \equiv \frac{dy}{dt}. \tag{3}$$

We emphasize that $\alpha(t)$ and $\beta(t)$ are intimately linked by (3) (which is sometimes referred to as the *Frobenious integrability condition*), and that α is the projection of β back to I. An example of such corresponding curves is shown in Fig. 5C.

One can immediately notice that while every image curve can be lifted to $T(I)$, not all curves in $T(I)$ are lifted versions of some image curve. We therefore define:

Definition 2. *A curve $\beta(t) = [x(t), y(t), \theta(t)] \in T(I)$ is called* admissible *if and only if $\exists \alpha(t) = [x(t), y(t)]$ such that (3) is satisfied.*

There are more inadmissible curves in $T(I)$ than admissible ones and examples for both admissible and inadmissible curves are shown in Figs. 5C and 5D. We also note that if we consider (3) as a differential equation, a curve is admissible if and only if it is an integral curve of this equation. Therefore, any completion mechanism in $T(I)$ is restricted to admissible curves only, and we shall refer to (3) accordingly as the *admissibility constraint*.

3.2 "Minimum Action" Completion in $T(I)$

What curve completion principles could we adopt in $T(I)$? In general, since both are vector spaces, any principle that one could use in I is a candidate principle for completion in $T(I)$ also, subject to the admissibility constraint. However, when we recall that $T(I)$ is an abstraction of V1, first candidates for completion principle should perhaps attempt to capture likely behavior of *neuron populations* rather than axiomatic *perceptual* criteria. Perhaps the simplest of such principles is a "minimum energy consumption" or "minimum action" principles, according to which the cortical tissue would attempt to link two boundary points (i.e., active cells) with the minimum number of additional active (i.e., energy consuming) cells that give rise to the completed curve. In the abstract this becomes the case of the *shortest admissible path* in $T(I)$ connecting two endpoints $[x_0, y_0, \theta_0]$ and $[x_1, y_1, \theta_1]$ (where admissibility must apply all along the curve, including its endpoints). Note that while such a shortest curve in I is necessarily a straight line, most linear curves in $T(I)$ are "inadmissible" in the sense of Definition 2. Since the shortest *admissible* curve in $T(I)$ has a nontrivial projection in the image plane, we hypothesize that the "minimum action" principle in $T(I)$ corresponds to the visually completed curve. The geometrical and perceptual properties of this curve can then be *induced* from this first principle rather than *imposed* as axioms. In the bulk of our paper we develop this theory in a rigorous manner, and we examine its entailed perceptual properties.

4 Minimum Length Curve in the Tangent Bundle

Given the motivation, arguments, and insights above, we are now able to define our curve completion problem formally. Let p_0 and p_1 be two given endpoints in $T(I)$ which represent two oriented inducers in the image plane I. We seek the shortest admissible path in $T(I)$ between these two given endpoints, i.e., the curve that minimizes

$$\mathcal{L} = \int_{p_0}^{p_1} \sqrt{\dot{\beta}(t)^2}dt = \int_{p_0}^{p_1} \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2 + \dot{\theta}(t)^2}dt, \tag{4}$$

subject to the admissibility constraint.

While this initial objective function employs a euclidean metric for $\mathbf{R}^2 \times \mathcal{S}^1$, a natural question that arises relates to the units and relative scale of the different dimensions in this space. Indeed, while x and y are measured in meters (or other length units), θ is measured in radians. Furthermore, the hypercolumnar organization of V1 suggests that the "cost" (or cortical distance) for moving one orientation unit is not necessarily similar to moving one spatial unit. Hence, to balance dimensions in the arclength integral and to facilitate relative scale between the spatial and angular coordinates, a proportionality constant $\hbar$ in units of $\frac{meters}{radians}$ should be incorporated in (4) (in a manner reminiscent of many physical proportionality constants such as the reduced Planck constant, which proportions the energy of a photon and the angular frequency of its associated electromagnetic wave). We thus generalize the distance measure between points in $T(I)$ and formulate our curve completion problem as follows:

Problem 2. *Given two endpoints $p_0 = [x_0, y_0, \theta_0]$ and $p_1 = [x_1, y_1, \theta_1]$ in $T(I)$, find the curve $\beta(t) = [x(t), y(t), \theta(t)]$ that minimizes the functional*

$$\mathcal{L}(\beta) = \int_{t_0}^{t_1} \sqrt{\dot{x}^2 + \dot{y}^2 + \hbar^2\dot{\theta}^2}dt, \tag{5}$$

while satisfying the boundary conditions $\beta(t_0) = p_0$ and $\beta(t_1) = p_1$ and the admissibility constraint from (3).

The rest of our paper is devoted to solving this problem formally and to analyzing the properties of the solution. Before we do so, we note that there were few previous independent attempts to address the curve completion problem in roto-translational spaces. For example, it was suggested by Petitot [38] that perceptually completed curves may be modeled as "geodesics contours" in V1, which he suggested to consider as the shortest path in the *jet space* $\mathbf{R}^2 \times P$, where P is the real line of $\tan(\theta)$, the tangent of orientation angles. Like Horn [21], Petitot suggested (but did not present) a solution based on elliptic integrals (whose practicality is questionable), and offered no evaluation vis-à-vis perceptual findings. A different approach by Citti and Sarti [7] considered the curve completion problem via the reconstruction of minimal *surfaces* in $\mathbf{R}^2 \times \mathcal{S}^1$ by means of a diffusion-based iterative procedure that attempts to group pairs of inducers and find the completed shape between them, all at the same time. Unfortunately, this ad hoc treatment of the grouping phase (c.f., Fig. 2A) is prone to yielding erroneous results and it completely ignores the rich literature on the grouping process and its relationship to other (possibly higher level) visual mechanisms. In addition, and beyond being less biologically plausible, this approach cannot be applied when the completion scene does not involve regions (as in Fig. 1D), it does not address the need to make the spatial and angular dimensions commensurable, and no attempt was made to relate it to perceptual findings. Compared to these few previous studies, our curve completion theory yields a completely different mathematical analysis using different mathematical tools, which not only yields clear, replicable, and easily implementable solution, but also facilitates a full examination of perceptual properties (Section 4.4) and extensive experimental validation.

Finally, Ben-Yosef and Ben-Shahar [4] have recently addressed a preliminary and simpler version of Problem 2 where the objective functional lacks the square root and a proportionality constant, i.e.,

$$\mathcal{L}(\beta) = \int_{t_0}^{t_1} [\dot{x}^2 + \dot{y}^2 + \dot{\theta}^2] dt. \tag{6}$$

Unfortunately, in addition to deviating from the minimum action principle as a guiding rule,[3] the functional in (6) is also nonintrinsic, i.e., its value depends on the particular parameterization of the curve. Clearly, both of these properties are undesired.

Unlike the above previous studies, here we address the general curve completion problem by 1) making the spatial and angular dimensions commensurable and scaled properly, and 2) applying the minimum action principle in the unit tangent bundle genuinely and directly. We offer a rigorous variational analysis in the continuous domain in a manner that facilitates not only a simple and complete solutions to Problem 2, but an extensive analysis of perceptual properties as well (unlike in, e.g., [38], [7]). Our analysis facilitates a numerical solution which is replicable and easily implementable, from which experimental results are easy to produce for future validation against perceptual and psychophysical data.

3. Note that without the square root, the functional in (6) no longer represents length in $\mathbf{R}^2 \times \mathcal{S}^1$.

4.1 Theoretical Differential Analysis

Let $\alpha(s) = [x(s), y(s)]$ be an image curve given in an arclength parameterization, whose corresponding lifted curve in $T(I)$ is

$$\beta(s) = [x(s), y(s), \theta(s)]. \tag{7}$$

Representing all admissible curves in $T(I)$ in this form, the functional $\mathcal{L}$ from (5) becomes

$$\mathcal{L}(\beta) = \int_0^l \sqrt{\dot{x}(s)^2 + \dot{y}(s)^2 + \hbar^2 \dot{\theta}(s)^2} ds, \tag{8}$$

where l is the total length of $\alpha(s)$ and the admissibility constraint (3) can be written as

$$\begin{aligned} \cos\theta(s) &= \dot{x}(s), \\ \sin\theta(s) &= \dot{y}(s). \end{aligned} \tag{9}$$

Given these observations, the following is our main theoretical result:

Theorem 1. *Of all admissible curves in $T(I)$, those that minimize the functional in (8) belong to a two parameter family (c, ϕ), which is defined by the following differential equation:*

$$\left(\hbar \frac{d\theta}{ds}\right)^2 = \frac{c^2}{\sin^2(\theta + \phi)} - 1. \tag{10}$$

Proof. Observe first that l in (8) is unknown. To facilitate the application of the calculus of variation to our problem we therefore begin by slightly changing the representation of this equation. Let $\kappa(s) = \dot{\theta}(s)$ be the curvature of α and let $R(s) \triangleq \frac{1}{|\kappa(s)|}$ be its corresponding radius of curvature at each point along its arclength. At this stage, assume also that θ is a monotonic function of s (this assumption, which amounts to α having no inflection points, will be removed later) and hence that it can be used to parameterize the curve. Let the constant $\sigma \in \{-1, 1\}$ be the sign of curvature. Since $\frac{d\theta}{ds} = \sigma \frac{1}{R}$, it follows that $ds = \sigma R d\theta$, which allows us to rephrase the admissibility constraint in (9) as follows:

$$\begin{aligned} dx &= \cos\theta ds = \sigma R \cos\theta d\theta, \\ dy &= \sin\theta ds = \sigma R \sin\theta d\theta. \end{aligned} \tag{11}$$

Substituting (11) into (8) we get

$$\begin{aligned} \mathcal{L}(\beta) &= \int_0^l \sqrt{\left(\frac{dx}{ds}\right)^2 + \left(\frac{dy}{ds}\right)^2 + \hbar^2 \left(\frac{d\theta}{ds}\right)^2} ds \\ &= \int_0^l \sqrt{\frac{\sigma^2 R^2 \cos^2(\theta) d\theta^2}{ds^2} + \frac{\sigma^2 R^2 \sin^2(\theta) d\theta^2}{ds^2} + \hbar^2 \sigma^2 \frac{d\theta^2}{ds^2}} ds, \end{aligned}$$

from which we immediately obtain the following new objective function for our minimization problem, this time in terms of the tangential orientation θ of the curve:

$$\mathcal{L}(\beta) = \sigma \int_{\theta 0}^{\theta 1} \sqrt{R(\theta)^2 + \hbar^2}\, d\theta. \tag{12}$$

By using this form to describe the curve, we are at risk of ignoring the boundary conditions on positions $[x_0, y_0]$ and $[x_1, y_1]$ that must be introduced back into the

problem.[4] This can be done by adding certain constraints that force the projection of the induced curve to pass through these two end points. For that, note that one can derive the following identities for the spans Δx and Δy of the end points:

$$\Delta x \triangleq x_1 - x_0 = \int_0^l \dot{x} = \int_{\theta_0}^{\theta_1} \sigma R \cos\theta d\theta,$$
$$\Delta y \triangleq y_1 - y_0 = \int_0^l \dot{y} = \int_{\theta_0}^{\theta_1} \sigma R \sin\theta d\theta,$$

from which we can rewrite the following integral constraints on the desired R and θ functions:

$$\int_{\theta_0}^{\theta_1} \left[\sigma R\cos\theta - \frac{\Delta x}{\Delta\theta}\right] d\theta = 0,$$
$$\int_{\theta_0}^{\theta_1} \left[\sigma R\sin\theta - \frac{\Delta y}{\Delta\theta}\right] d\theta = 0,$$

where $\Delta\theta = \theta_1 - \theta_0$. These additional constraints can now be incorporated into our new functional in (12) using two arbitrary Lagrange multipliers λ_x and λ_y. The result is the following minimization problem in terms of θ:

$$\mathcal{L}(\beta) = \sigma \int_{\theta_0}^{\theta_1} \left[\sqrt{R^2+\hbar^2} + \lambda_x\left(R\cos\theta - \frac{\Delta x}{\sigma\Delta\theta}\right) + \lambda_y\left(R\sin\theta - \frac{\Delta y}{\sigma\Delta\theta}\right)\right] d\theta. \tag{13}$$

Note that although θ is measured in radians, $\tan\theta = \frac{\dot{y}}{\dot{x}}$, $\cos\theta$, and $\sin\theta$ are dimensionless and therefore do not unbalance the units in the functional.

With (13) representing the variational problem to be solved, the corresponding Euler-Lagrange equation becomes

$$\frac{d}{dR}\left(\sqrt{R^2+\hbar^2} + \lambda_x\left(R\cos\theta - \frac{\Delta x}{\sigma\Delta\theta}\right) + \lambda_y\left(R\sin\theta - \frac{\Delta y}{\sigma\Delta\theta}\right)\right) = 0,$$

which can be further simplified to

$$\frac{R}{\sqrt{R^2+\hbar^2}} + \lambda_x\cos\theta + \lambda_y\sin\theta = 0. \tag{14}$$

Renaming our free parameters more conveniently by $\lambda_x = \frac{1}{c}\sin\phi$ and $\lambda_y = \frac{1}{c}\cos\phi$, we can further simplify (14) as follows:

$$\frac{R}{\sqrt{R^2+\hbar^2}} = -\frac{1}{c}\sin(\phi+\theta). \tag{15}$$

Finally, taking the square of both sides of (15) and remembering that $R = 1/\dot{\theta}$, we obtain the desired (10).

To generalize the proof to inflectional curves as well, assume that the shortest admissible path between two endpoints p_0 and p_n in $\mathbf{R}^2 \times \mathcal{S}^1$ is (at least) twice continuously differentiable and has a finite number of $n-1$ inflection points $p_1, p_2, \ldots, p_{n-1}$. (This is a reasonable assumption since we do not expect the optimal curve to wiggle too much, let alone an infinite number of times.) The curve is thus a union of n noninflectional curve segments, each of which is described by R from (14), with a possible change of sign between them. Hence, the behavior of R^2 (and therefore of $\dot{\theta}^2$) is valid to all curve segments and thus to the entire length of the inflectional curve as a whole. □

4. Note that the initial conditions on the *orientation* of the two inducers become very explicit in this form and are embedded directly in the limits of the integral in (12).

4.2 Analytical Solution via Elliptic Integrals

The similarity of (10) to the elastica equation (cf., (2)) and the Elastica-Pendulum equation [32] suggests that a closed-form analytic solution is unlikely. However, a solution based on elliptic integrals has been offered to the Elastica problem [21], [32], and it is therefore reasonable to suspect that elliptic integrals may fit in our case as well. Indeed, in our second theoretical result in this paper we show how noninflectional curves from Theorem 1 and (10) can be expressed in terms of elliptic integrals.

Theorem 2. *Let $\alpha(\theta) = [x(\theta), y(\theta)]$ be the image projection of a curve β in $T(I)$ that belongs to the family of (10). Suppose α is noninflectional and thus it can be parameterized by its tangential angle θ. Then,*

$$\begin{aligned} x(\theta) = x_0 &- \hbar\cos\phi\left(\sqrt{c^2-\sin^2(\theta+\phi)} - \sqrt{c^2-\sin^2(\theta_0+\phi)}\right) \\ &- c\hbar\sin\phi\left(E\left(\frac{1}{c^2},\theta+\phi\right) - E\left(\frac{1}{c^2},\theta_0+\phi\right)\right. \\ &\left. - F\left(\frac{1}{c^2},\theta+\phi\right) + F\left(\frac{1}{c^2},\theta_0+\phi\right)\right), \\ y(\theta) = y_0 &- \hbar\sin\phi\left(\sqrt{c^2-\sin^2(\theta+\phi)} - \sqrt{c^2-\sin^2(\theta_0+\phi)}\right) \\ &- c\hbar\cos\phi\left(E\left(\frac{1}{c^2},\theta+\phi\right) - E\left(\frac{1}{c^2},\theta_0+\phi\right)\right. \\ &\left. - F\left(\frac{1}{c^2},\theta+\phi\right) + F\left(\frac{1}{c^2},\theta_0+\phi\right)\right), \end{aligned} \tag{16}$$

where F, E are the incomplete elliptic integrals of the first and second kinds.

Proof. We start with the observation that

$$\frac{d\theta}{dx} = \frac{d\theta}{ds} \Big/ \frac{dx}{ds},$$

in which $\frac{d\theta}{ds}$ can be expressed directly in terms of (10). When the sign of curvature is fixed (wlog, positive) and s is the arclength (such that $\frac{dx}{ds} = \cos\theta$), the above term amounts to

$$\frac{d\theta}{dx} = \frac{\sqrt{c^2-\sin^2(\theta+\phi)}}{\hbar\sin(\theta+\phi)} \cdot \frac{1}{\cos\theta}. \tag{17}$$

By rearranging (17) and integrating we get

$$x(\theta_s) = \int_0^{x_s} dx = I_x \triangleq \hbar \int_{\theta_0}^{\theta_s} \frac{\cos\theta\sin(\theta+\phi)}{\sqrt{c^2-\sin^2(\theta+\phi)}} d\theta, \tag{18}$$

where for the convenience of reading we denote the tangential angle of the curve as θ_s and we save θ to denote the parameter of integration. (We also assume that $x_0 = 0$; otherwise x_0 should simply be added to I_x.) To solve I_x we first substitute $\mu = \theta + \phi$, such that $d\theta = d\mu$. We thus get

$$I_x = \hbar \int_{\theta_0+\phi}^{\theta_s+\phi} \frac{\cos(\mu-\phi)\sin\mu}{\sqrt{c^2-\sin^2\mu}}\,d\mu \tag{19}$$

or

$$I_x = \hbar\cos\phi \int_{\theta_0+\phi}^{\theta_s+\phi} \frac{\cos\mu\sin\mu}{\sqrt{c^2-\sin^2\mu}}\,d\mu + \hbar\sin\phi \int_{\theta_0+\phi}^{\theta_s+\phi} \frac{\sin^2\mu}{\sqrt{c^2-\sin^2\mu}}\,d\mu.$$

Consider the two addends of the last expression and call their corresponding integrals I_1 and I_2 for the left and right terms, respectively. The solution for I_1 is simple:

$$I_1 = \int_{\theta_0+\phi}^{\theta_s+\phi} \frac{cos\mu\sin\mu}{\sqrt{c^2-\sin^2\mu}}\,d\mu = \Big|_{\theta_0+\phi}^{\theta_s+\phi} - \sqrt{c^2-\sin^2\mu}, \tag{20}$$

while I_2 can be solved via elliptic integrals

$$I_2 = \int_{\theta_0+\phi}^{\theta_s+\phi} \frac{\sin^2\mu}{\sqrt{c^2-\sin^2\mu}}\,d\mu = -c\int_{\theta_0+\phi}^{\theta_s+\phi} \frac{-\frac{1}{c^2}\sin^2\mu}{\sqrt{1-\frac{1}{c^2}\sin^2\mu}}\,d\mu$$
$$= -c\int_{\theta_0+\phi}^{\theta_s+\phi} \left(\frac{1-\frac{1}{c^2}\sin^2\mu}{\sqrt{1-\frac{1}{c^2}\sin^2\mu}} - \frac{1}{\sqrt{1-\frac{1}{c^2}\sin^2\mu}}\right)d\mu$$

or

$$I_2 = -c\int_{\theta_0+\phi}^{\theta_s+\phi} d\mu\sqrt{1-\frac{1}{c^2}\sin^2\mu} \;\; + c\int_{\theta_0+\phi}^{\theta_s+\phi} \frac{d\mu}{\sqrt{1-\frac{1}{c^2}\sin^2\mu}}. \tag{21}$$

Fortunately, the first addend in (21) is the incomplete elliptic integral of the second kind (usually denoted as E), and the second addend in (21) is the incomplete elliptic integral of the first kind (usually denoted as F). Thus, we get

$$x(\theta_s) = -\hbar\cos\phi\left(\sqrt{c^2-\sin^2(\theta_s+\phi)} - \sqrt{c^2-\sin^2(\theta_0+\phi)}\right) - c\hbar\sin\phi\left(E\left(\frac{1}{c^2},\theta_s+\phi\right) - E\left(\frac{1}{c^2},\theta_0+\phi\right) - F\left(\frac{1}{c^2},\theta_s+\phi\right) + F\left(\frac{1}{c^2},\theta_0+\phi\right)\right). \tag{22}$$

(Note that (10) implies that $c^2 > 1$, which makes the parameter $\frac{1}{c^2}$ of the elliptic integrals smaller than 1.)

The entire derivation thus far can be repeated in a similar way to find $y(\theta_s)$, for which we apply (10) to the identity

$$\frac{d\theta}{dy} = \frac{d\theta}{ds}\Big/\frac{dy}{ds}.$$

Again, when the sign of curvature is fixed, and s denotes arclength (such that $\frac{dy}{ds} = \sin\theta$), the above term becomes

$$\frac{d\theta}{dy} = \frac{\sqrt{c^2-\sin^2(\theta+\phi)}}{\hbar\sin(\theta+\phi)}\cdot\frac{1}{\sin\theta}. \tag{23}$$

Again, by rearranging (23) and integrating we get

$$y(\theta_s) = \int_0^{y_s} dy = I_y \triangleq \hbar\int_{\theta_0}^{\theta_s} \frac{\sin\theta\sin(\theta+\phi)}{\sqrt{c^2-\sin^2(\theta+\phi)}}\,d\theta, \tag{24}$$

and by using the same substitution $\mu = \theta + \phi$ we eventually obtain

$$I_y = \hbar\cos\phi\int_{\theta_0+\phi}^{\theta_s+\phi} \frac{\sin^2\mu}{\sqrt{c^2-\sin^2\mu}}\,d\mu - \hbar\sin\phi\int_{\theta_0+\phi}^{\theta_s+\phi} \frac{\cos\mu\sin\mu}{\sqrt{c^2-\sin^2\mu}}\,d\mu. \tag{25}$$

This term involves the same integrals I_2 and I_1 that we have solved previously. Thus,

$$y(\theta_s) = -c\hbar\cos\phi\left(E\left(\frac{1}{c^2},\theta_s+\phi\right) - E\left(\frac{1}{c^2},\theta_0+\phi\right) - F\left(\frac{1}{c^2},\theta_s+\phi\right) + F\left(\frac{1}{c^2},\theta_0+\phi\right)\right) - \hbar\sin\phi\left(\sqrt{c^2-\sin^2(\theta_s+\phi)} - \sqrt{c^2-\sin^2(\theta_0+\phi)}\right). \tag{26}$$

□

4.3 Numerical Solution

The analysis shown in Sections 4.1 and 4.2 is still short of resolving the parameters c, ϕ of the specific curve from the family of (10) which passes through the given tangent bundle boundary points p_0 and p_1. Apparently, these parameters can be solved by applying the end point constraint to (16), i.e., by solving

$$\begin{aligned} x(\theta_1) &= x_1, \\ y(\theta_1) &= y_1, \end{aligned} \tag{27}$$

which become the following pair of equations in the two unknowns c and ϕ (note that all other parts are known):

$$x_1 = x_0 - \hbar\cos\phi\left(\sqrt{c^2-\sin^2(\theta_1+\phi)} - \sqrt{c^2-\sin^2(\theta_0+\phi)}\right) - c\hbar\sin\phi\left(E\left(\frac{1}{c^2},\theta_1+\phi\right) - E\left(\frac{1}{c^2},\theta_0+\phi\right) - F\left(\frac{1}{c^2},\theta_1+\phi\right) + F\left(\frac{1}{c^2},\theta_0+\phi\right)\right),$$
$$y_1 = y_0 - \hbar\sin\phi\left(\sqrt{c^2-\sin^2(\theta_1+\phi)} - \sqrt{c^2-\sin^2(\theta_0+\phi)}\right) - c\hbar\cos\phi\left(E\left(\frac{1}{c^2},\theta_1+\phi\right) - E\left(\frac{1}{c^2},\theta_0+\phi\right) - F\left(\frac{1}{c^2},\theta_1+\phi\right) + F\left(\frac{1}{c^2},\theta_0+\phi\right)\right). \tag{28}$$

Unfortunately, given the severely nonlinear nature of this system it seems unlikely that an analytical solution is viable. Since we must resort to numerical methods, however, it seems reasonable to extend the scope of our sought after solutions to general (i.e., possibly inflectional) curves, rather than noninflectional curves only,[5] a limitation that is intrinsic to (28).

To find a numerical solution in the general case, we therefore abandon (28) and turn to formulating an ODE-based solution with boundary conditions. To begin, we notice that to facilitate the computation of inflectional curves, it is preferable to solve the second order ODE that is obtained by differentiating (10):

$$\hbar^2 \frac{d^2\theta}{ds^2} = -\frac{c^2 \cos(\theta+\phi)}{\sin^3(\theta+\phi)}. \tag{29}$$

This way we avoid determination of the sign of the square root when it is applied to (10). Still, at first sight, this approach is problematic since it appears that the number of constraints in our problems is smaller than its degrees of freedom (or free parameters). Indeed, (29) represents a family of planar curves in *Whewell* form (i.e., an equation that relates the tangential angle of the curve with its arclength), which induces an image parametric curve via the following integrations:

$$\begin{aligned} x(s) &= x_0 + \int_0^s \cos\theta(\tilde{s}) d\tilde{s}, \\ y(s) &= y_0 + \int_0^s \sin\theta(\tilde{s}) d\tilde{s}. \end{aligned} \tag{30}$$

Thus, a single and unique curve from our family of solutions is determined by 7 degrees of freedom: θ_0, $\dot{\theta}_0$ (or put differently, the curvature κ_0 at p_0), ϕ, and c are needed to resolve a unique $\theta(s)$ function via (29), and x_0, y_0, and l are then needed to determine the curve's coordinate functions $x(s)$ and $y(s)$ from the first to the second inducer via (30). At the same time, the curve completion problem provides only six constraints expressed by the two given inducers:

$$\begin{array}{lll} x(0)=x_0 & y(0)=y_0 & \theta(0)=\theta_0 \\ x(l)=x_1 & y(l)=y_1 & \theta(l)=\theta_1. \end{array}$$

Fortunately, this initial observation does not imply that our problem is underdetermined (and therefore lacks unique solutions) since it turns out that c can be expressed in terms of κ_0 and θ_0. To do so we evaluate (10) at θ_0:

$$\left(\hbar \frac{d\theta}{ds}\Big|_{\theta_0}\right)^2 = \frac{c^2}{\sin^2(\theta_0+\phi)} - 1,$$

which results in the following identity:

$$c^2 = (\hbar^2 {\kappa_0}^2 + 1) \cdot \sin^2(\theta_0+\phi). \tag{31}$$

5. Indeed, much of the perceptual curve completion literature suggests that perceptually completed curves are unlikely to incorporate inflections points (e.g., [26], [49], [11], c.f. Section 2.2). However, following recent psychophysical work (e.g., [11], [49]), we argue that the discrimination between these perceptual states is not dichotomous but rather graded, and that inflectional completed curves are indeed occasionally perceived (as may be experienced for certain boundary conditions; see Fig. 9 for several examples), albeit with possibly smaller perceptual strength.

Substituting (31) in (29) we now obtain

$$\ddot{\theta} = \frac{-({\kappa_0}^2 + \frac{1}{\hbar^2}) \cdot \sin^2(\theta_0+\phi)\cos(\theta+\phi)}{\sin^3(\theta+\phi)}, \tag{32}$$

in which c no longer participates.

Following these algebraic manipulations, we therefore assert that our curve completion problem can be answered by solving (32), and then use the resolved parameters to construct the completed curve with (30). One standard numerical technique for solving such ODE is based on nonlinear optimization that seeks the values of the equation parameters that satisfy the given boundary conditions. In our case, this entails the following general algorithm:

1. Make an initial guess regarding the values of the parameters κ_0, ϕ, and l.
2. Construct a curve of length l starting from $p_0 = [x_0, y_0, \theta_0]$ in a way that obeys (32).
3. Evaluate the correctness of the parameters by assessing the error between the obtained end point of the constructed curve (i.e., the point $[x(l), y(l), \theta(l)]$) and the desired end point ($p_1 = [x_1, y_1, \theta_1]$).
4. Use the error $E(\kappa_0, \phi, l)$ between these two tangent bundle points to update the parameters before iterating back to step 2.

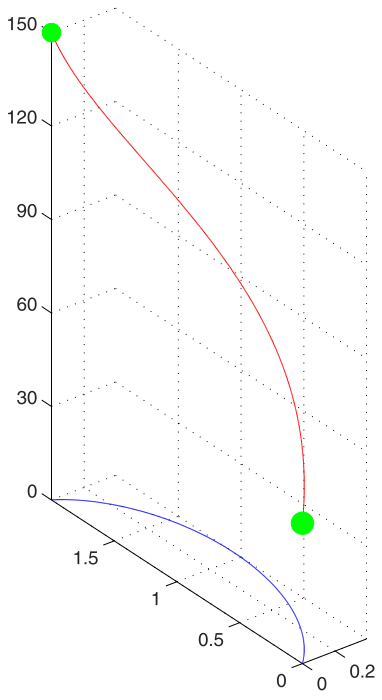
More specifically, for each iteration i with a given starting point p_0 and parameter values κ_0 and ϕ we first solve the differential equation (32) via Euler's method (though more sophisticated methods could be used too, of course). Initializing arclength $s_0 = 0$ at the beginning of each iteration we compute

$$\begin{aligned} s_{n+1} &\triangleq s_n + h, \\ \kappa_{n+1} &\triangleq \kappa(s_{n+1}) = \kappa(s_n + h) \approx \kappa(s_n) + h \cdot \ddot{\theta}(s_n) \\ &= \kappa_n + h \cdot \frac{-({\kappa_0}^2 + \frac{1}{\hbar^2}) \cdot \sin^2(\theta_0+\phi)\cos(\theta_n+\phi)}{\sin^3(\theta_n+\phi)}, \\ \theta_{n+1} &\triangleq \theta(s_{n+1}) = \theta(s_n + h) \approx \theta(s_n) + h \cdot \kappa(s_n) \\ &= \theta_n + h \cdot \kappa_n, \\ y_{n+1} &\triangleq y(s_{n+1}) \approx y(s_n) + h \cdot \dot{y}(s_n) \\ &= y_n + h \cdot \sin\theta_n, \\ x_{n+1} &\triangleq x(s_{n+1}) \approx x(s_n) + h \cdot \dot{x}(s_n) \\ &= x_n + h \cdot \cos\theta_n, \end{aligned}$$

where h is a preselected step size and the error is of order $O(h)$. The curve $\beta(s_i) = [x(s_i), y(s_i), \theta(s_i)]$ computed by this step is then evaluated at $s_n = l$ (i.e., at step $n = l/h$) to obtain the point $[x_{end}, y_{end}, \theta_{end}] = [x(l), y(l), \theta(l)]$ and the error $E(\kappa_0, \phi, l)$ associated with the current value of the parameters is computed by

$$E(\kappa_0, \phi, l) = \|[x_1, y_1, \theta_1] - [x_{end}, y_{end}, \theta_{end}]\|.$$

The new values for κ_0, ϕ, and l are then computed by gradient descent on $E(\kappa_0, \phi, l)$. A demonstration of a minimum curve (and its image projection) that is generated by this procedure is shown in Fig. 7A. Also, to validate our numerical process we compared curves from different models according to their length in $T(I)$, length in I, and



A

Model	Length in I	Curvature in I	Length in $T(I)$
$p_0 = [0, 0, 20^\circ]$ and $p_1 = [1, 0, -10^\circ]$			
Biarc	1.0128	0.3896	1.1820
Cubic Interpolation	1.0232	0.3655	1.1848
Elastica	1.0126	**0.3608**	1.1720
Euler Spiral	1.0122	0.3656	1.1809
Minimum Length in Tangent Bundle	**1.0112**	0.3683	**1.1687**
$p_0 = [0, 0, 40^\circ]$ and $p_1 = [1, 0, -30^\circ]$			
Biarc	1.0685	1.5054	1.6394
Cubic Interpolation	1.0918	1.5354	1.6697
Elastica	1.0688	**1.4816**	1.6368
Euler Spiral	1.0657	1.4949	1.6417
Minimum Length in Tangent Bundle	**1.0569**	1.5309	**1.6287**
$p_0 = [0, 0, 60^\circ]$ and $p_1 = [1, 0, -50^\circ]$			
Biarc	1.1852	3.2117	2.2646
Cubic Interpolation	1.3082	3.8697	2.4215
Elastica	1.1889	**3.1929**	2.2668
Euler Spiral	1.1725	3.2118	2.2698
Minimum Length in Tangent Bundle	**1.1353**	3.5262	**2.2435**
$p_0 = [0, 0, 70^\circ]$ and $p_1 = [1, 0, -60^\circ]$			
Biarc	1.2741	4.1437	2.6092
Cubic Interpolation	1.6053	6.2661	3.0180
Elastica	1.2899	**4.1157**	2.6197
Euler Spiral	1.2523	4.1711	2.6117
Minimum Length in Tangent Bundle	**1.1798**	4.9496	**2.5758**

B

Fig. 7. (A) The curve of minimum length in the tangent bundle as computed by our numerical procedure in Section 4.3. The shortest admissible path in $T(I)$ between points (marked here in green) $p_0 = [0, 0, 45^\circ]$ and $p_1 = [0, 2, 150^\circ]$ for $\hbar = 1$ is shown in red, and its projection to I is shown in blue. (B) Comparison between computationally completed curves according to their length in $I(\int \sqrt{\dot{x}^2 + \dot{y}^2} dt)$, total curvature squared in $I(\int \dot{\kappa}^2 dt)$, and length in $T(I)(\int \sqrt{\dot{x}^2 + \dot{y}^2 + \hbar^2 \dot{\theta}^2} dt)$ for $\hbar = 1$, for various boundary conditions. See corresponding plots in Figs. 3E, 3F, 3G, and 3H.

total (squared) curvature in I. Naturally, this evaluation should show that our model possesses the smallest length in $T(I)$, while the elastica model produces the smallest total curvature in I. These expectations are indeed confirmed in the results of this quantitative evaluation, which we show in Fig. 7B.

To conclude this section we note that in certain cases the numerical optimization process can be simplified significantly, as expressed in the following proposition:

Theorem 3. *If the sought-after solution is limited to curves $\beta(s)$ whose image projection $\alpha(s)$ is noninflectional, then the number of free parameters that should be determined via optimization is reduced from three (κ_0, ϕ, and l) to two (κ_0 and ϕ) only.*

Proof. If $\alpha(s)$ is restricted to be noninflectional, its sign of curvature must be constant (wlog, positive) along its entire length. Hence, we can safely take the square root of (10) to write the curvature as follows:

$$\kappa = \frac{d\theta}{ds} = \frac{\sqrt{c^2 - \sin^2(\theta + \phi)}}{\hbar \sin(\theta + \phi)}. \tag{33}$$

Integrating (33) we get

$$\int_{\theta_0}^{\theta_1} \left[\hbar \frac{\sin(\theta + \phi)}{\sqrt{c^2 - \sin^2(\theta + \phi)}} \right] d\theta = \int_0^l ds,$$

or

$$I_0 = \hbar \int_{\theta_0}^{\theta_1} \left[\frac{\sin(\theta + \phi)}{\sqrt{c^2 - 1 + \cos^2(\theta + \phi)}} \right] d\theta = l.$$

To evaluate the integral I_0 we substitute

$$z = \cos(\theta + \phi),$$

such that $dz = -sin(\theta + \phi)d\theta$, and obtain

$$I_0 = \hbar \int_{\theta_0}^{\theta_1} \frac{-dz}{\sqrt{c^2 - 1 + z^2}}.$$

Remembering that $c^2 > 1$, we substitute $a^2 = c^2 - 1 > 0$ and get

$$I_0 = -\hbar \int_{\theta_0}^{\theta_1} \frac{dz}{\sqrt{a^2 + z^2}} = -\hbar \Big|_{\theta_0}^{\theta_1} \ln \left| z + \sqrt{z^2 + a^2} \right|. \tag{34}$$

Finally, we obtain the following:

$$\begin{aligned} l = \hbar (&\ln |\cos(\theta_0 + \phi) + \sqrt{\cos^2(\theta_0 + \phi) + c^2 - 1}| \\ &- \ln |\cos(\theta_1 + \phi) + \sqrt{\cos^2(\theta_1 + \phi) + c^2 - 1}|), \end{aligned} \tag{35}$$

which expresses l in terms of c and ϕ, and by application of (31), in terms of κ_0 and ϕ (the other two free parameters in the optimization process). □

4.4 Dependency on Scale and Other Visual Properties

Our discussion so far illustrates how the problem of curve completion can be formulated and solved in the space that abstracts the early visual cortical regions where this perceptual process is likely to occur. Since the theory, and the single principle of "minimum action" that guides this solution, are "nonperceptual," it is important to understand what perceptual properties they entail, and how these predictions correspond to existing perceptual findings and the geometrical axioms reviewed earlier. This section studies this question in a rigorous manner.

As suggested in the beginning of this section, the value of the $\hbar$ constant could have a significant influence over the shape of curves of minimum length in the tangent bundle, as it implicitly controls the relative contribution of total length in I versus total curvature in I during the minimization process,[6] or put differently, the relative scale ratio (i.e., the proportion between units of measurement) of the length and orientation axes in the unit tangent bundle. In this context, the behavior at the limits of $\hbar$ provides important qualitative insights regarding its effect. On one hand, if $\hbar$ is very small, the minimization process becomes similar to minimization of length in I (subject to boundary conditions) and we therefore expect the resultant curve to straighten (or "flatten"). On the other hand, when $\hbar$ is very large, the minimization process is dominated by the minimization of orientation derivative (again, subject to boundary conditions), a condition that resembles (qualitatively) the classical elastica and converges to a particular shape. To show these properties formally, we rewrite (10) as

$$\left(\frac{d\theta}{ds}\right)^2 = \frac{c^2}{\hbar^2 \sin^2(\theta+\phi)} - \frac{1}{\hbar^2} = \frac{\tilde{c}}{\sin^2(\theta+\phi)} - \frac{1}{\hbar^2}, \tag{36}$$

where $\tilde{c} = \frac{c}{\hbar}$ is a renaming of the constant that fits the boundary conditions. Using this representation we now observe that when $\hbar \to \infty$, (36) converges to

$$\left(\frac{d\theta}{ds}\right)^2 = \frac{\tilde{c}}{\sin^2(\theta+\phi)}. \tag{37}$$

(Note that in such case c can also be very large and thus balance $\hbar$ in $\tilde{c} = \frac{c}{\hbar}$.) The shape of the curve described by this expression can be shown experimentally to be rather rounded, although it does *not* describe a circular arc (see Fig. 8A for an example using particular initial conditions). To examine the curve behavior when $\hbar \to 0$, we return to (10), which now becomes

$$0 = \frac{c^2}{\sin^2(\theta+\phi)} - 1$$

or

$$\sin(\theta+\phi) = \pm c. \tag{38}$$

Hence, since in the limit $\sin(\theta(s)+\phi)$ must be constant for all s, $\theta(s)$ becomes constant also, which implies a linear curve. A demonstration of several curves for the same inducer pair but different values of $\hbar$ is shown in Fig. 8A.

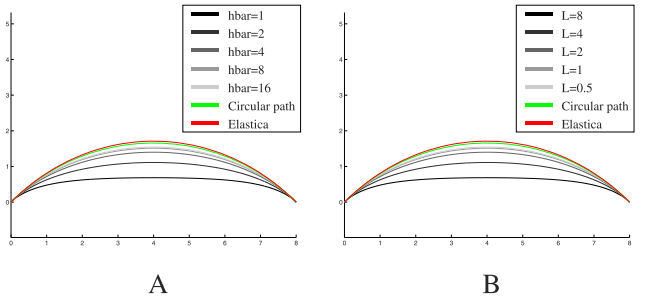


Fig. 8. Minimum length tangent bundle curves at different scales (or "viewing distances") and different $\hbar$ values. (A) Effect of $\hbar$. Shown here are examples for two fixed inducers ($[0,0,45°]$ and $[0,8,-45°]$) and different values of $\hbar$. Note the convergence of the limiting shape for large $\hbar$. (B) Effect of viewing scale. Shown here are examples for a fixed $\hbar = 1$ and different scale values. To obtain this graph we have positioned the two inducers at $[0,0,45°]$ and $[0,L,-45°]$, with $L = 0.5, 1, 2, 4, 8$, computed the optimal curve, and then scaled each of them by a factor of $8/L$ so their shape can be compared on the same plot. Note the "flattening" with increased scale and the identical results to decreasing $\hbar$. For comparison, we also plotted the corresponding elastica and circular curves (in red and green, respectively), the latter also demonstrating the deviation of our model from roundedness.

Intuitively, changing $\hbar$ amounts to changing the viewing scale of a particular completion task, i.e., applying a global scale transform on the initial conditions. Indeed, viewing distance affects the projected distance between the inducers on the retina (i.e., the image plane) and hence the cortical distance between the hypercolumns of the inducing neurons in V1 (i.e., the tangent bundle). Hence, one could expect that the effect of scaling on the resultant minimum length curve would be similar to changing $\hbar$, or in other words, that minimum length in the tangent bundle does not provide scale independence. To show this formally, we first observe that any (i.e., not necessarily minimal) image curve $\alpha_L(\tilde{t})$ traveling from $p_0 = [0,0,\theta_0]$ to $p_1 = [L,0,\theta_1]$ can be written as a scaled version of some other curve $\alpha_1(t)$ traveling from $[0,0,\theta_0]$ to $[1,0,\theta_1]$:

$$\alpha_L(t) = L \cdot \alpha_1(t) = [L \cdot x(t), L \cdot y(t)], t \in [t_0, t_1],$$

where

$$\alpha_1(t) = [x(t), y(t)] \;\; t \in [t_0, t_1] \text{ and } x(t_0) = 0, \; x(t_1) = 1.$$

Hence, the lifted $T(I)$ curve $\beta_L(t)$ that corresponds to $\alpha_L(t)$ becomes

$$\begin{aligned}\beta_L(t) &= \left[L \cdot x, L \cdot y, tan^{-1}\left(\frac{L \cdot \dot{y}}{L \cdot \dot{x}}\right)\right] \\ &= [L \cdot x(t), L \cdot y(t), \theta(t))] \quad t \in [t_0, t_1],\end{aligned}$$

where $[x(t), y(t), \theta(t)]$ is the tangent bundle curve that corresponds to $\alpha_1(t)$. Note now that the total arclength of β_L is

$$\begin{aligned}\mathcal{L}(\beta_L) &= \int_{t_0}^{t_1} \sqrt{L^2(\dot{x}^2 + \dot{y}^2) + \hbar^2 \dot{\theta}^2}\, dx \\ &= L \cdot \int_{t_0}^{t_1} \sqrt{\dot{x}^2 + \dot{y}^2 + \left(\frac{\hbar}{L}\right)^2 \dot{\theta}^2}\, dx.\end{aligned} \tag{39}$$

The identity in (39) indicates that one could minimize either functional to obtain the desired result. However,

6. Note that since $\int_{t_0}^{t_1} \sqrt{\dot{x}^2 + \dot{y}^2} dt$ amounts to total length in the image plane while $\int_{t_0}^{t_1} \sqrt{\dot{\theta}^2} dt$ represents total curvature in the image plane, $\hbar$ in (5) can be interpreted as balancing between these two terms.

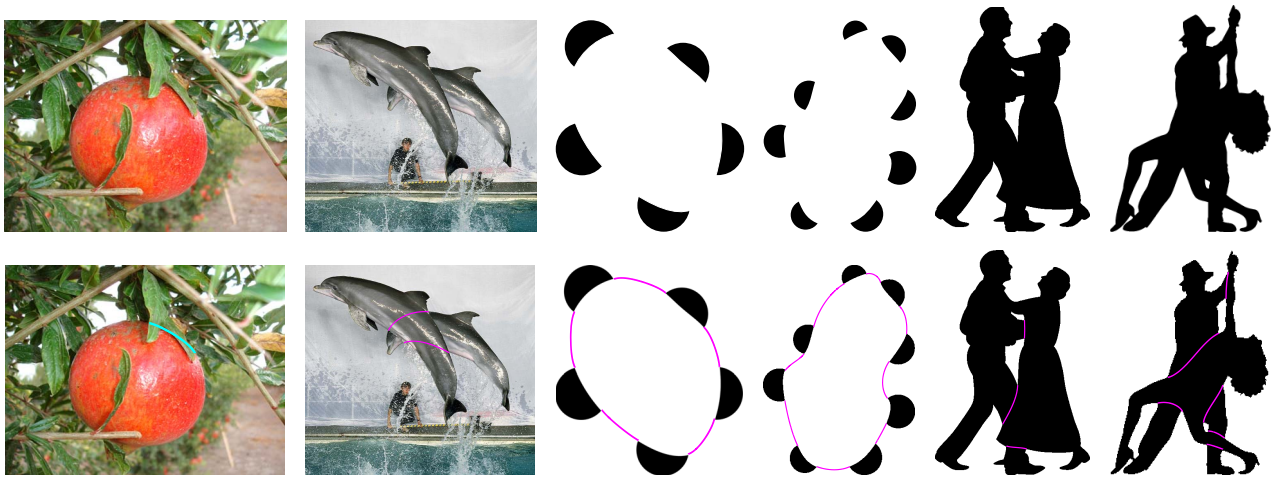


Fig. 9. Examples of curve completion via minimum length in the tangent bundle applied in different amodal and modal settings. Magenta completions reflect $\hbar = 2$ and pixel size (i.e., viewing scale) of 0.01 in all cases.

notice that the left functional reflects the scaling up of the visual input by L, while the right functional represents the scaling down of $\hbar$ by L. Hence, we conclude that changing global scale has an effect inverse to changing $\hbar$. We therefore expect a correspondence between $L \to 0$ and $\hbar \to \infty$, and vice versa. The two panels of Fig. 8 demonstrate this correspondence experimentally, using two symmetric inducers as initial conditions.

In addition to scaling issues, several properties of our model can be pointed out regarding the six axioms of curve completion mentioned in Section 2.1. First, since our solution is not linked to any specific frame, it is trivially isotropic. Note that since the rotated minimum length curve also satisfies (10) (for $\tilde{c} = c$ and $\tilde{\phi} = \phi + \rho$, where ρ is the angle of rotation), the solution is invariant under rotations. Second, since the solution minimizes total arclength in $T(I)$, it must be extensible in that space and hence in the image plane also. Third, since the completed curves can be described by a differential equation (10), they clearly satisfy the axiom of smoothness.

Obviously, the analysis of scale that was discussed above indicates that our theory generates scale *variant* solutions or, put differently, it does not satisfy the axiom of scale invariance. Another axiom where our model departs from prior solutions is the axiom of roundedness, since it is easy to confirm that the case of constant curvature ($\frac{d\theta}{ds} = const$) does *not* satisfy (10). At first sight these two properties could undermine the utility of our model, but given the refutation of both scale invariance and roundedness at the perceptual and psychophysical level (as discussed in Section 2.2), we consider these properties an important advantage of our theory rather than a limitation. That these properties were derived as emergent properties rather than imposed as axioms is yet another benefit of our approach as a whole.

4.5 Experimental Results

While this paper is primarily theoretical and while the full utility of our proposed new theory requires psychophysical verification (which is part of our short-term future research), here we have experimented with our results by applying them to selected instances of curve completion problems. Some results are shown in Figs. 9, 10, and 11 and demonstrate that the completed curves correspond well with the perceptual outcome (and, as is evident, with the physical occluded curve as well). Clearly, determination of the "correct" $\hbar$ is a matter of perceptual calibration or anatomical and physiological considerations,[7] both of which are outside the scope of this theoretical and computational paper. For the demonstrated examples we have calibrated $\hbar$ manually after setting the scale (i.e., pixel size in length units) arbitrarily. Inducer orientation was measured manually from the images and all initial data were fed to the numerical algorithm from Section 4.3. In all results the parameters κ_0, ϕ, and l were optimized up to an error $E(\kappa_0, \phi, l) \leq 10^{-5}$. The resultant curves of minimum arclength in $T(I)$ were then projected to the image plane I and plotted on the missing parts of the image. In several cases we also show a comparison to other common models, such as the biarc, the cubic interpolation, the elastica, and the Euler spiral models.

It should be emphasized that results such as those shown in Figs. 9, 10, and 11, as good as they appear, do not necessarily indicate that a certain model is better than others. Indeed, until a comprehensive evaluation is carried out against perceptual findings, no such conclusion can be made. However, the presented results do suggest *quantitative* methods to assess completions by computer vision systems. In particular, the use of synthetic occluders (as in Fig. 10) on a large collection of objects may facilitate an evaluation of completion results against the statistics of natural image contours. More importantly, quantitative validation against perceptual findings (in the spirit of

7. We stress that $\hbar$ is not a *free* parameter of the theory that needs to be fitted to each image, stimulus, or viewing condition. This constant needs to be determined only once via psychophysical experiments (perhaps using paradigms like those shown in Fig. 4). One may also reason about the value of $\hbar$ based on anatomical and physiological data. Although a nontrivial task, this may be done by considering the density of neurons in the primary visual cortex and their columnar and hypercolumnar layout, as well as other anatomical parameters such as the cortical magnification factor, magnification scaling, eccentricity, and the cortical extent of horizontal connections neurons.

Fig. 10. Curve completion via minimum length in the tangent bundle (shown in magenta) applied in different natural scenes with synthetic occluders. Pixel size in all examples was set to 0.01 and $\hbar$ was set to 2. To compare our results to various curve completion algorithms we also present the completions due to the biarc (green), cubic interpolation (red), elastica (blue), and Euler spiral (cyan) models for the same pixel size (note that in some cases different models overlap and occlude each other). Please zoom in using the digital version to observe the differences better.

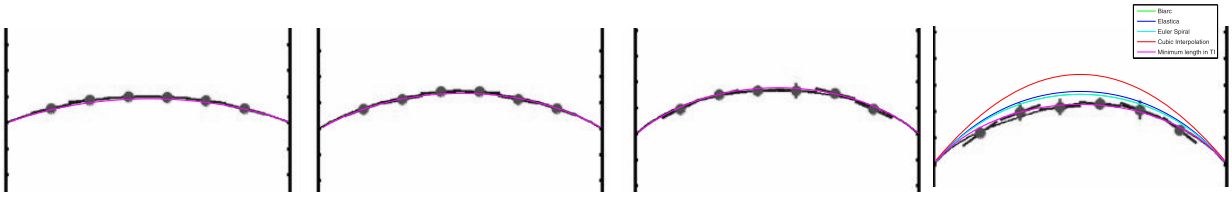


Fig. 11. Comparison to psychophysics. Completion as the curve of minimum length in the tangent bundle is shown (magenta) for four pairs of symmetric inducers with boundary orientation θ and $-\theta$, where $\theta = 20^\circ, 30^\circ, 40^\circ, 50^\circ$ (ordered from left to right panel, top to bottom). Mode outputs are plotted on top of the average of repeated completions by one human observer (reproduced from Fulvio et al. [11], their Fig. 4, observer 1). The proportionality constant $\hbar$ was set to 1.5 (for pixel size of 0.01) after it was calibrated to match the perceptual evidence in the first three cases. It was then tested on the fourth (rightmost) case and compared to other models (biarc, cubic interpolation, elastica, and Euler spiral) for the same viewing scale (i.e., pixel size). While this is by no means an exhaustive perceptual verification, it does emphasize that different models could make significantly different predictions, especially when the orientation difference between the inducers increases.

Fig. 11) could provide comparative evaluation of different models vis-à-vis the goal of finding the "correct" perceptual completions (please refer again to Section 1 for the discussion about the notion of "correctness" in this context).

5 Discussion and Future Directions

This paper proposes a new theory of curve completion in the unit tangent bundle, the latter being the space which abstracts the early areas in the visual cortex where much of the curve completion process presumably occurs.

Employing a universal principle of "minimum action" which strives to minimize neural energy consumption, we have shown that curve completion amounts to finding the shortest admissible path in $\mathbf{R}^2 \times \mathcal{S}^1$. We have proven differential properties of this path, solved it analytically in terms of elliptic integrals, showed how it can be found numerically, and derived its perceptual properties.

As we have shown, the basic principle of minimal neural energy consumption entails parametric-free completions[8] that minimize a combination of both total curvature and total length in the image plane (see Section 4). In this sense our model naturally expresses two basic Gestalt principles [54]. The first, the principle of *good continuation*, is often formalized as minimization of curvature [51]. The second, the principle of *proximity*, is naturally formulated as minimization of total length. While such combinations have been explored in both the perceptual literature (e.g., [10]) and the computational community (e.g., [53], [46], [56]), our theory does so as a result of minimizing a single basic (and nonvisual) principle, from which perceptual insights are *derived* rather than *imposed*.

Since the tangent bundle is a natural space for examining *combinations* of image plane properties (as we also discussed in Section 4.4), it is natural to test additional completion principles in $T(I)$ and explore their implications. For example, one could attempt to study the "tangent bundle curve of least energy" as a direct extension to the curve of least energy (i.e., the elastica) in the image plane (e.g., [21]). Such a criterion may reflect a biological mechanism which attempts to minimize properties of the *connections* between neurons, rather than the mere number of activated cells that represent the completed curve. It could also relate to the curvature-related fashion by which cells link to their neighbors [3]. In either case, the fact that the medium in which our curves are embedded is three-dimensional implies that their energy might consist not only a bending energy but also a *twisting* energy [30], [34]. Abstracting an infinitesimally thick rod in 3D space, a curve β would therefore be associated with an energy described more formally by

$$\mathcal{E}(\beta) = \mathcal{E}_B(\beta) + \mathcal{E}_T(\beta) = \int [a\kappa_\beta(s)^2 + b\tau_\beta(s)^2]ds, \qquad (40)$$

where a, b are the elasticity coefficients of the rod, κ_β is the curvature of β such that κ_β^2 is proportional to the bending energy $\mathcal{E}_B$, and τ_β is the *torsion* of β such that τ_β^2 is proportional to the twisting energy $\mathcal{E}_T$ [30], [34]. Considered in the image plane (i.e., in terms of the projected curve α), (40) would amount to exploring a certain combination of both total curvature of α and total *change of curvature* of α (as would happen if one were to combine elastica [21], [35] and the Euler spiral [27] in one model). Interestingly enough, our early exploration of such directions indicates that a unique solution to this problem requires boundary conditions that include not only inducer position and orientation, but also their *curvatures*, a type of condition that has been speculated also in the perceptual literature [50], [49], [48]. A full and rigorous investigation of such a theory is currently underway as a part of our short term future research.

8. We reemphasize that $\hbar$ is *not* a free parameter one needs to fit for each image or viewing conditions. Rather, it is a constant that remains fixed (per observer or even within a species) and therefore needs to be calibrated only once.

As illustrated in Figs. 9, 10, and 11, the completions that are generated by our tangent bundle model often match the desired perceptual outcome and the behavior of natural image contours. Obviously, it is not expected that all curves in natural images would belong to this family of curves (cf., Fig. 2B). Our results do suggest, however, that an examination of natural contour statistics in the context of curve completion is an important topic. Following up on recent initial work on this issue (e.g., [13]), this is another goal of our short term future research.

One aspect of the curve completion problem that has gained interest in recent studies is the "strength of the perceptual completion" between two given inducers [26], [11], [12]. Highly related to the grouping problem (cf. Section 2.2), this topic too may be informed by our tangent bundle theory. Since the total length of admissible curves in the unit tangent bundle $\mathbf{R}^2 \times \mathcal{S}^1$ represents the number of active cells in V1 and since each neural computation carries some measure of uncertainty, it may be possible to associate each completed curve with a measure of *total uncertainty* that is proportional to the arclength $\mathcal{L}$ in (5). We hypothesize that the perceptual strength of a completed curve (in terms of accuracy and consistency with human performance [11], [12]) decreases with growing uncertainty, i.e., decreases with the number of active cells. In other words, perceptual strength is inversely related to $\mathcal{L}$. Unlike the binary measure of relatability [26], this proposed type of perceptual strength is a *graded and continuous* geometrical measure for grouping (cf. Section 2.2)) and it may be used to resolve grouping decisions when multiple combinations are possible (as is the case of most nontrivial completion scenes).

Finally, it is interesting to reflect on the link between our work and early ideas of the Gestalt movement, and in particular their notion of "tendency toward minimum energy." Inspired by Euler, Lagrange, Hamilton, and others, who have shown that the central equations of mechanics can be derived by a tendency toward minimum energy (or a *stationary state*), Gestatalists like Kohler (e.g., [29]) hypothesized that the principles of *perception* are also the outcome of a "development in the direction of minimum energy" [29, page 52], or what was encapsulated in the German word *Pragnänz*, which sometimes is translated as "good form." In this paper we too have formulated a classical perceptual organization problem in terms of a tendency toward minimum energy. Unlike previous energy minimization formulations of the same problem, however, we have done so by an abstraction of what may be considered as the "true" energy of a visual completed curve, namely, the number of active orientation-selective cells in the primary visual cortex. Whether or not this really is the appropriate energy to use is left here as an open question, with the observation that the tangent bundle framework offers many interesting and more elaborate alternatives to explore the curve completion problem.

ACKNOWLEDGMENTS

This work was funded in part by the European Commission in the 7th Framework Programme (CROPS GA no. 246252) and the Israel Science Foundation (ISF) grant No. 1245/08. The authors also thank the generous support of the Frankel fund, the Paul Ivanier center for Robotics Research, and the Zlotowski Center for Neuroscience at Ben-Gurion University.

REFERENCES

[1] O. Ben-Shahar, "The Perceptual Organization of Visual Flows," PhD thesis, Yale Univ., 2003.

[2] O. Ben-Shahar and S.W. Zucker, "The Perceptual Organization of Texture Flows: A Contextual Inference Approach," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 4, pp. 401-417, Apr. 2003.

[3] O. Ben-Shahar and S.W. Zucker, "Geometrical Computations Explain Projection Patterns of Long Range Horizontal Connections in Visual Cortex," *Neural Computation,* vol. 16, no. 3, pp. 445-476, 2004.

[4] G. Ben-Yosef and O. Ben-Shahar, "Minimum Length in the Tangent Bundle as a Model for Curve Completion," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2010.

[5] W.H. Bosking, Y. Zhang, B. Schofield, and D. Fitzpatrick, "Orientation Selectivity and the Arrangement of Horizontal Connections in the Tree Shrew Striate Cortex," *The J. Neuroscience,* vol. 17, no. 6, pp. 2112-2127, 1997.

[6] M. Brady, W.E.L. Grimson, and D.J. Langridge, "Shape Encoding and Subjective Contours," *Proc. Am. Assoc. for Artificial Intelligence Conf.,* pp. 15-17, 1980.

[7] G. Citti and A. Sarti, "A Cortical Based Model of Perceptual Completion in the Roto-Translation Space," *J. Math. Imaging and Vision,* vol. 24, no. 3, pp. 307-326, 2006.

[8] S. Coren, "Subjective Contours and Apparent Depth," *Psychological Rev.,* vol. 79, no. 4, pp. 359-367, 1972.

[9] J.H. Elder and R.M. Goldberg, "Ecological Statistics of Gestalt Laws for the Perceptual Organization of Contours," *J. Vision,* vol. 2, pp. 324-353, 2002.

[10] C. Fantoni and W. Gerbino, "Contour Interpolation by Vector-Field Combination," *J. Vision,* vol. 3, pp. 281-303, 2003.

[11] J.M. Fulvio, M. Singh, and L.T. Maloney, "Precision and Consistency of Contour Interpolation," *Vision Research,* vol. 48, pp. 831-849, 2008.

[12] J.M. Fulvio, M. Singh, and L.T. Maloney, "An Experimental Criterion for Consistency in Interpolation of Partly Occluded Contours," *J. Vision,* vol. 9, pp. 1-19, 2009.

[13] W.S. Geisler and J.S. Perry, "Contour Statistics in Natural Images: Grouping across Occlusions," *Visual Neuroscience,* vol. 26, pp. 109-121, 2009.

[14] W.S. Geisler, J.S. Perry, B.J. Super, and D.P. Gallogly, "Edge Co-Occurrence in Natural Images Predicts Contour Grouping Performance," *Vision Research,* vol. 41, no. 6, pp. 711-724, 2001.

[15] W. Gerbino and C. Fantoni, "Visual Interpolation in Not Scale Invariant," *Vision Research,* vol. 46, pp. 3142-3159, 2006.

[16] C.D. Gilbert, "Horizontal Integration and Cortical Dynamics," *Neuron,* vol. 9, pp. 1-13, 1992.

[17] D.H. Grosof, R.M. Shapley, and M.J. Hawken, "Macaque v1 Neurons Can Signal 'Illusory' Contours," *Nature,* vol. 365, pp. 550-552, 1993.

[18] S.E. Guttman and P.J. Kellman, "Contour Interpolation Revealed by a Dot Localization Paradigm," *Vision Research,* vol. 44, pp. 1799-1815, 2004.

[19] R.K. Hladky and S.D. Pauls, "Minimal Surfaces in the Roto-translational Group with Applications to a Neuro-Biological Image Completion Model," *J. Math. Imaging and Vision,* vol. 36, pp. 1-27, 2010.

[20] W.C. Hoffman, "The Visual Cortex Is a Contact Bundle," *Applied Math. and Computation,* vol. 32, pp. 137-167, 1989.

[21] B.K.P. Horn, "The Curve of Least Energy," *ACM Trans. Math. Software,* vol. 9, no. 4, pp. 441-460, 1983.

[22] J. Hoschek and D. Lasser, *Fundamentals of Computer Aided Geometric Design.* A.K. Peters, Ltd., 1989.

[23] D. Hubel and T. Wiesel, "Functional Architecture of Macaque Monkey Visual Cortex," *Royal Soc. of London, Series B,* vol. 198, pp. 1-59, 1977.

[24] Y.M. Jung and J. Shen, "First-Order Modeling and Stability Analysis of Illusory Contours," *J. Visual Comm. and Image Representation,* vol. 19, pp. 42-55, 2008.

[25] G. Kanizsa, *Organization in Vision: Essays on Gestalt Perception.* Praeger Publishers, 1979.

[26] P.J. Kellman and T.F. Shipley, "A Theory of Visual Interpolation in Object Perception," *Cognitive Psychology,* vol. 23, pp. 141-221, 1991.

[27] B.K. Kimia, I. Frankel, and A.M. Popescu, "Euler Spiral for Shape Completion," *Int'l J. Computer Vision,* vol. 54, nos. 1-3, pp. 159-182, 2003.

[28] D.E. Knuth, "Mathematical Typography," *Bull. of the Am. Math. Soc.,* vol. 1, pp. 337-372, 1979.

[29] W. Köhler, "Physical Gestalten," *A Source Book of Gestalt Psychology,* W.D. Ellis, ed., pp. 17-53, Routledge & Kegan Paul Ltd., 1920.

[30] L.D. Landau and E.M. Lifshitz, *Theory of Elasticity.* Pergamon, 1986.

[31] T.S. Lee and M. Nguyen, "Dynamics of Subjective Contours Formation in the Early Visual Cortex," *Proc. Nat'l Academy of Sciences USA,* vol. 98, no. 4, pp. 1907-1911, 2001.

[32] R. Levien, "The Elastica: A Mathematical History," Technical Report UCB/EECS-2008-103, Electrical Eeng. and Computer Science Dept., Univ. of California, Berkeley, Aug. 2008.

[33] R. Levien, "The Euler Spiral: A Mathematical History," Technical Report UCB/EECS-2008-111, Electrical Eng. and Computer Science Dept., Univ. of California, Berkeley, Sept. 2008.

[34] A.E.H. Love, *A Treatise on the Mathematical Theory of Elasticity.* Dover Publications, 1944.

[35] D. Mumford, "Elastica in Computer Vision," *Algebric Geometry and Its Applications,* B. Chandrajit, ed., Springer-Verlag, 1994.

[36] B. O'Neill, *Semi-Riemannian Geometry with Applications to Relativity.* Academic Press, 1983.

[37] S.E. Palmer, *Vision Science: Photons to Phenomenology.* The MIT Press, 1999.

[38] J. Petitot, "Neurogeometry of v1 and Kanizsa Contours," *Axiomathes,* vol. 13, nos. 3/4, pp. 347-363, 2003.

[39] X. Ren, C. Fowlkes, and J. Malik, "Learning Probabilistic Models for Contour Completion in Natural Images," *Int'l J. Computer Vision,* vol. 77, nos. 1-3, pp. 47-63, 2008.

[40] D.L. Ringach and R. Shapley, "Spatial and Temporal Properties of Illusory Contours and Amodal Boundary Completion," *Vision Research,* vol. 36, no. 19, pp. 3037-3050, 1996.

[41] K.S. Rockland and J.S. Lund, "Widespread Periodic Intrinsic Connections in the Tree Shrew Visual Corte," *Science,* vol. 215, no. 19, pp. 1532-1534, 1982.

[42] W.S. Rutkowski, "Shape Completion," *Computer Vision, Graphics, and Image Processing,* vol. 9, pp. 89-101, 1979.

[43] A. Sarti, R. Malladi, and J.A. Sethian, "Subjective Surfaces: A Method for Completing Missing Boundaries," *Int'l J. Computer Vision,* vol. 46, no. 32, pp. 201-221, 2002.

[44] I.J. Schoenberg, "Contributions to the Problem of Approximation of Equidistant Data by Analytic Functions," *Quarterly of Applied Math.,* vol. 4, pp. 44-99, 1946.

[45] F. Schumann, "Beitrage zur Analyse der Gesichtswhrnehmungen," *Zeitschrift für Psychologie,* vol. 33, pp. 161-185, 1904.

[46] E. Sharon, A. Brandt, and R. Basri, "Completion Energies and Scale," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 10, pp. 1117-1131, Oct. 2000.

[47] M. Singh, "Modal and Amodal Completion Generate Different Shapes," *Psychological Science,* vol. 15, no. 7, pp. 454-459, 2004.

[48] M. Singh and J.M. Fulvio, "Visual Extrapolation of Contour Geometry," *Proc. Nat'l Academy of Sciences USA,* vol. 102, pp. 939-944, 2005.

[49] M. Singh and D.D. Hoffman, "Completing Visual Contours: The Relationship between Relatability and Minimizing Inflections," *Perception and Psychophisics,* vol. 61, pp. 943-951, 1999.

[50] H. Takeichi, "The Effect of Curvature on Visual Interpolation," *Perception,* vol. 24, pp. 1011-1020, 1995.

[51] S. Ullman, "Filling in the Gaps: The Shape of Subjective Contours and a Model for Their Creation," *Biological Cybernetics,* vol. 25, pp. 1-6, 1976.

[52] R. von der Heydt, E. Peterhans, and G. Baumgartner, "Illusory Contours and Cortical Neuron Responses," *Science,* vol. 224, pp. 1260-1262, 1984.

[53] I. Weiss, "3D Shape Representation by Contours," *Computer Vision, Graphics, and Image Processing,* vol. 41, pp. 80-100, 1988.

[54] M. Wertheimer, "Untersuchungen zur Lehre von der Gestalt," *Psychologische Forschung,* vol. 4, pp. 301-350, 1923.

[55] S. Willard, *General Topology.* Addison-Wesley Publishing Company, 1970.

[56] L.R. Williams and D.W. Jacobs, "Stochastic Completion Fields: A Neural Model of Illusory Contour Shape and Salience," *Neural Computation,* vol. 9, no. 4, pp. 837-858, 1997.

Guy Ben-Yosef received BSc degrees in physics and in computer science and the MSc degree in computer science from Ben Gurion University of the Negev, Israel, in 2005 and 2007, respectively. He is currently working toward the PhD degree in computer science at Ben Gurion University under the supervision of Ohad Ben-Shahar. He is interested in theoretical vision sciences and in integrating psychophysical and neurophsyiological experimental findings for rigorous computational theories. He was the winner of the best research award in image processing, computer vision, and video processing at the Israeli Machine Vision Conference 2011, and the Intel award for excellent Israeli PhD students in 2011. He is a student member of the IEEE and the IEEE Computer Society.

Ohad Ben-Shahar received the BSc and MSc degrees in computer science from the Technion, Israel Institute of Technology, and the PhD degree in computer science from Yale University, New Haven, Connecticut, in 2003. He is an associate professor in the Department of Computer Science at Ben Gurion University of the Negev (BGU) and the director of the BGU Interdisciplinary Computational Vision Lab. His main area of research is in computational vision and image analysis, where he is focusing primarily on issues related to the differential geometrical foundations of perceptual organization and early vision. His work is explicitly multidisciplinary and his computational research is often endowed by investigations into human perception, visual psychophysics, and computational neuroscience of biological vision. He is the recipient of the 2007 Psychobiology Young Investigator Award and his research is funded by the Israel Science Foundation (ISF), the US Air Force Office of Scientific Research (AFOSR), the Deutsche Forschungsgemeinschaft (DFG) in Germany, the US National Science Foundation (NSF), the US National Institute for Psychobiology, and the European Union (FP7). He is a member of the IEEE and the IEEE Computer Society.