First-Order Modeling and Stability Analysis of Illusory Contours

Yoon Mo Jung and Jianhong (Jackie) Shen *

Abstract

In visual cognition, illusions help elucidate certain intriguing latent perceptual functions of the human vision system, and their proper mathematical modeling and computational simulation are therefore deeply beneficial to both biological and computer vision. Inspired by existent prior works, the current paper proposes a first-order energy-based model for analyzing and simulating illusory contours. The lower complexity of the proposed model facilitates rigorous mathematical analysis on the detailed geometric structures of illusory contours. After being asymptotically approximated by classical active contours, the proposed model is then robustly computed using the celebrated level-set method of Osher and Sethian (*J. Comput. Phys.*, **79**:12-49, 1988) with a natural supervising scheme. Potential cognitive implications of the mathematical results are addressed, and generic computational examples are demonstrated and discussed.

Keywords: illusion, illusory contours, energy model, local stability, contour decomposition, real and imaginary, kinks, spans, turns, level-set method, supervision.

1 Introduction

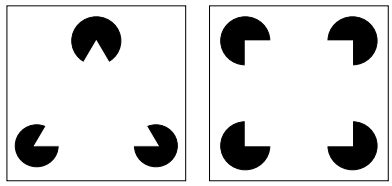
In System Theory [22], input-output analysis has been a major tool for partial or complete identification of black-box systems. In cognitive vision science, the study of various visual illusions follows exactly the same spirit. Cognitive scientists have designed numerous intriguing inputs of image signals, so that the distorted or transformed outputs (as reported by an average human observer) can help reveal some crucial latent properties of the human vision system (e.g., the remarkable works [1, 15, 17, 12]). *Illusory contours* are such a well known class of visual illusions, and the current paper develops a mathematical model to characterize, analyze, and simulate generic illusory contours. Our work has been closely inspired by many existent modeling works, especially by Sarti, Malladi, and Sethian [27], and Zhu and Chan [37, 38].

We set out by emphasizing first that by no means the current mathematical model purports to be absolutely faithful in the context of neuron science or cortical vision. Rather, the present work still remains *phenomenological*, as [27, 37, 38] do. In fact, it is precisely our ultimate goal to be able to develop a neurally rational model that incorporates all the important discoveries by neuron scientists (e.g., [18, 34, 33, 36, 25]). The current work and all the aforementioned inspiring predecessors may thus pave the valuable stepping stones towards this goal.

Fig. 1 shows two examples of illusory contours known as *Kanizsa triangle and square* [12, 27, 37]. The illusory or *imaginary* white triangle and square pop out almost instantly to a normal observer. The human vision system is capable of interpolating non-existent edges and making

*This work has been partially supported by the NSF (USA) under grant number DMS-0604510. Mail address of the authors: School of Mathematics, University of Minnesota, 206 Church Street SE, Minneapolis, MN 55455, USA. Corresponding author: Prof. Jackie Shen, jhshen@math.umn.edu; (612) 625-3570 (Tel); (612) 626-2017 (Fax); www.math.umn.edu/~jhshen.

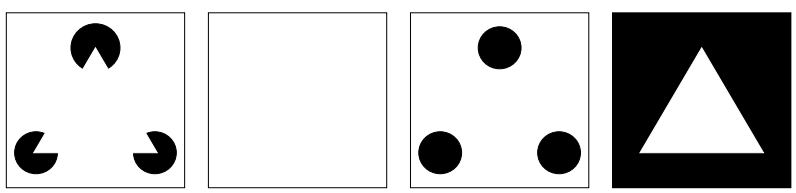
meaningful visual organization of both the *real* and *imaginary* edge segments. From the viewpoints of modeling and computation, such illusory perception has also been called *segmentation with missing boundaries* [27], and is closely related to some earlier works (e.g., [20, 21]).



(a) (b)

Figure 1: Two classical examples of illusory contours: Kanizsa's triangle and square.

Qualitatively, the perception of illusory contours can be deciphered by the Gestalt framework [11, 12]. In a typical example of illusory contours, the illusory target objects share the same gray level as their background and become invisible to ordinary detectors. The illusory targets however do leave their footprints, mainly through the cues of occlusion, e.g., the visible corners in Fig. 1. An average human vision system can organize these cues and develop a so-called 2.1-D sketch of the scene [21], i.e., to separate and complete objects in ordered (according to the relative depth to a viewer) multiple layers. The multi-layer sketch of Kanizsa's triangle is displayed in Fig. 2, for example. We also refer the reader to the recent work of Shen [32] on an abstract study of the occlusion phenomenon and 2.1-D models.



(a) (b) (c) (d)

Figure 2: Multi-layer decomposition of Kanizsa's triangle: (a) image of Kanizsa's triangle, (b) far background, (c) three disks in the middle range, and (d) the illusory triangle closest to a viewer.

Quantitatively, however, it is a nontrivial task to properly model and compute illusory contours. There are two notable recent works that have primarily influenced the present work in developing plausible mathematical models and robust computational schemes.

In [27], Sarti, Malladi, and Sethian first proposed a variational-PDE model for computing illusory contours based on a *reference point* within an image. Given a reference point which models the focus of visual attention, a surface is then constructed and flattened except along the existing *real* edges. The level sets of the surface function are able to connect the imaginary contours. Define the edge indicator function by [27]

$$g(x) = g(x_1, x_2) = \frac{1}{1 + (|\nabla G_\sigma(x) * u(x)|/\beta)^2},$$

$$G_\sigma(\xi) = \frac{\exp(-(\xi/\sigma)^2)}{\sigma\sqrt{\pi}}.$$

Then the model is to minimize the g-weighted area of the surface $S : x = (x_1, x_2) \to (x_1, x_2, \Phi)$:

$$\min_\Phi \int_\Omega g(x)\sqrt{1 + \Phi_{x_1}^2 + \Phi_{x_2}^2}\, dx.$$

Our model shares a similar computational formulation, but is more self-contained in terms of modeling and analysis since it is directly built on curves rather than on reference points and surfaces. (We must point out that, in visual cognition, attention focuses do play important roles in a number of cognitive tasks [12].)

Later in [37], Zhu and Chan proposed a more complex level-set based variational model. Let $u : \Omega \to \mathbb{R}$ be an image defined on the domain Ω, with $u = 1$ on the objects and $u = -1$ otherwise. The key component of Zhu and Chan's model is the *signed distance* function from solving the eikonal equation :

$$\frac{\partial d}{\partial t} = \text{sign}(u)(1 - |\nabla d|),$$

with the initial data $d(x, t = 0) = u(x) = u(x_1, x_2)$. Their variational model is then to minimize the following energy for the level-set function ϕ which codes the desired illusory contours:

$$\begin{aligned} E(\phi) = &\int_\Omega \left((1 + \mu C(d)\kappa^+(d))|d|\delta(\phi)|\nabla\phi| + \lambda H(d)H(\phi)\right) dx \\ &+ \int_\Omega \left(a + b\left|\nabla \cdot \left[\frac{\nabla\phi}{|\nabla\phi|}\right]\right|^2\right) |\nabla\phi|\delta(\phi)dx, \end{aligned} \tag{1}$$

where δ denotes Dirac's delta, H the Heaviside function, and C a C^1 differentiable cut-off function. The $|d|\delta(\phi)|\nabla\phi|$ term measures the distance from the zero level-set to the existent boundaries and $C(d)\kappa^+(d)|d|\delta(\phi)|\nabla\phi|$ measures the distance to concave corners. The second term of the first integral measures the overlap between the existent objects and the interior of ϕ. The last integral represents Euler's elastica energy which was first proposed by Mumford [19], and later applied to geometric image inpainting by Chan, Kang, and Shen [3], and Esedoglu and Shen [9]. Zhu and Chan [37] showed outstanding performance of this model for several general examples. Our proposed model is much less complex than this high-order geometric model, and is also independent of the signed distance function. The essence of our proposed model is to be able to capture the most salient features of illusory perception in a simple and analyzable manner, at the unsurprising cost of losing certain high-order details. (Zhu and Chan [38] later also employed prior shape information to capture illusory contours.)

Mainly inspired by these pioneering works on quantitative modeling, we propose a new energy-based model for studying illusory contours. The model is directly formulated on admissible contours and does not depend upon reference points nor their associated surfaces. The model has a much lower order of complexity which however yields good leading-order approximation to most generic illusory contours (with exact capturing in the canonical cases of Kanizsa's triangle or square). The relationship between this new model and the more complex models of Zhu and Chan is formally analogous to that between the first-order and second-order Taylor expansions in Calculus. Most remarkably, the lower complexity of the new model allows to almost completely characterize the structures of illusory contours, which provides a solid stepstone towards more general quantitative analysis in the future. To our best knowledge, such rigorous characterization has been unprecedented in the literature.

The second major feature of the new model is that in terms of computation, it can be efficiently implemented by classical active contours and level-set based algorithms of Osher and Sethian [24, 23, 29, 28]. In particular, we introduce a key but natural step of supervision for capturing robustly illusory contours, which can also be considered as an innate component of the new model itself.

Finally, our theory shows that any *nonempty* local minima to the proposed energy model must be anchored or pinned somewhere along the *existing* edges of a given configuration. As a result, the

initial input (contour or the level-set function) to the evolution equation must be spacious enough to accommodate all existing edges within. This is crucial for avoiding any "leakage" of existing edge information any bit of which could contribute to the emergence of an illusory contour. Further elaboration can be found in the main text.

The paper has been organized as follows. In Section 2, we develop and analyze the new model for illusory contours that emerge from any generic configuration of real *objects*. We are able to rigorously characterize the geometric structures of illusory contours when identified as the local minima of the proposed model. In Section 3, the model is further extended to process configurations with real *contours*. In Section 4, an approximate or relaxed model is then developed based on the classical active-contour model of Kass, Witkin and Terzopoulos [13], which allows the proposed model to be actually computed via asymptotic approximation. In Section 5, the level-set method of Osher and Sethian [23, 24, 29] is employed to compute the stable local minima of the model, under the help of a key step of supervision. Finally, the details of numerical implementation as well as the computational performance of the proposed model are provided in Section 6.

2 Illusory contours as stable local minima and characterization

In this section, we propose a contour-based energy model, and model illusory contours as the *local minima* of such an energy. The particular form of the energy allows detailed and rigorous analysis of the geometric and topological structures of admissible local minima. Such in-depth mathematical characterization is, to our best knowledge, unique in the literature.

Note that a local minimum corresponds to an energy valley and is stable *only* to local perturbations. This differs our model from many other efforts in modeling illusory contours as global minima. Aiming at local instead of global minima has also been motivated by cognitive sensitivities of illusory perception (of light, geometry, or objects, etc) to various factors including visual attention, focus points, orientations, and scaling, etc. Such visual factors may disrupt or erase an observer's ongoing illusory experience, while global minima often remain stable to these external perturbations.

In later sections we then develop an efficient level-set-based computational algorithm that supports stable stagnation at such local minima.

2.1 Kinks, outer and inner spans θ, and a generic configuration Q

Let Ω denote the entire open image domain, which could be the complete plane $\mathbb{R}^2$ or a finite rectangle area. Let $Q = \{Q_i \mid i = 1 : k\}$ be a *configuration* of *disjoint* compact domains whose boundaries are *piecewise-C^2* simple closed curves. Denote the boundaries by

$$\partial Q = \partial Q_1 \cup \partial Q_2 \cup \cdots \cup \partial Q_k.$$

Definition 2.1 (A kink and its outer span θ and inner span θ^*)**.** A point $y \in \partial Q$ is said to be a kink if the two tangent lines from the two sides are *not* identical. If y is a kink, its outer span θ_y is the angle between the two tangent lines when moving *continuously* from one arm to the other *outside* Q (Fig. 3). We further define $\theta_y^* = 2\pi - \theta_y$ to be the inner span.

A cusp is a singular kink point y for which either the outer span $\theta_y = 0$ or the inner span $\theta_y^* = 0$ (Fig. 3). It will be interesting in vision psychology to understand the cognitive effects of cusps in various scales. In the current work we shall assume that the given configuration Q only contains

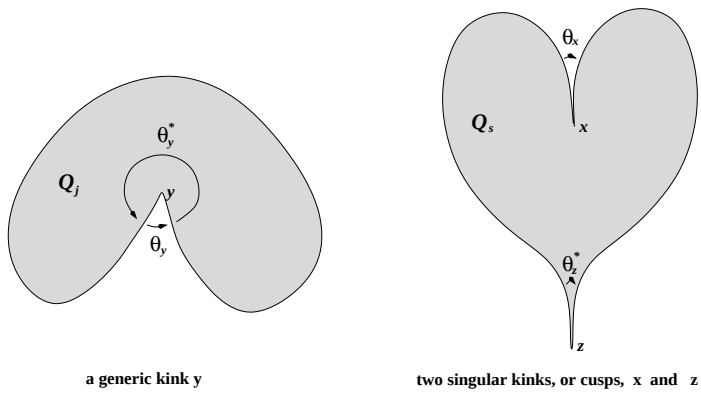


Figure 3: **Left panel:** a generic kink y with outer span $\theta_y > 0$ and inner span $\theta_y^* = 2\pi - \theta_y > 0$. **Right panel:** two examples of singular kinks or cusps; an inward cusp at z with $\theta_z^* = 0$ and an outward cusp x with $\theta_x = 0$. The current work discusses *generic* configurations containing only generic kinks. Perceptual effects of cusps have scarcely been experimented in visual cognition.

generic kinks so that $\theta_y, \theta_y^* > 0$. Such a configuration Q is said to be *generic* consequently. Almost all classical examples correspond to generic configurations.

Definition 2.2 (Kink set K and minimum spans $\theta_{\min}$ and $\theta^*_{\min}$)**.** Let $Q \subset \Omega$ be a *generic* configuration. Define its kink set $K[Q]$ to be the collection of all kinks of Q, and its minimum outer span by

$$\theta_{\min} = \theta_{\min}[Q] = \min\{\theta_y \mid y \in K[Q]\},$$

and minimum inner span by

$$\theta^*_{\min} = \theta^*_{\min}[Q] = \min\{\theta_y^* \mid y \in K[Q]\} = \max\{2\pi - \theta_y \mid y \in K[Q]\}.$$

Since ∂Q is compact and piecewise C^2, $K[Q]$ must be finite. Consequently, both $\theta_{\min}$ and $\theta^*_{\min}$ must be positive as well.

2.2 Real and imaginary decomposition of a contour

Let Γ be any piecewise C^2 simple closed curve in Ω *that does not intersect the interior of* Q. Define its real and imaginary parts by

$$\Gamma_{\text{re}} = \Gamma \cap \partial Q, \qquad \Gamma_{\text{im}} = \Gamma \setminus \partial Q. \tag{2}$$

For an illusory contour Γ, the imaginary part corresponds precisely to the illusory interpolation (Fig. 4). (The names "real" and "imaginary" have been formally motivated by complex analysis in which a complex number z can be decomposed into real and imaginary parts.)

Notice that since both Γ and ∂Q are compact and thus *closed* sets, Γ_{re} must be closed in Ω and Γ_{im} *relatively* open in Γ. We shall only consider a curve Γ for which Γ_{im} contains *finitely* many connected components to avoid unnecessary technicalities:

$$\Gamma_{\text{im}} = \gamma^1_{\text{im}} \cup \gamma^2_{\text{im}} \cup \cdots \cup \gamma^m_{\text{im}}.$$

Since $\Gamma_{\text{im}} = \Gamma \cap Q^c$ (with $Q^c = \mathbb{R}^2 \setminus Q$), the intersection of a simple closed (and thus locally connected) curve and an open domain, Γ_{im} must be locally connected [14]. Consequently, each

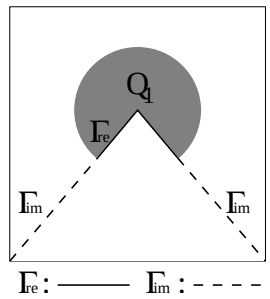


Figure 4: Decomposition of Γ into real and imaginary parts (partial view of Kanizsa's triangle).

connected component γ_{im}^j must be a connected curve segment with at most two endpoints. *In the current work, we assume that the two endpoints are distinct.* (Otherwise, the entire curve Γ becomes a loop which touches ∂Q only at a single point. Such a loop can easily shrink to decrease the energy to be proposed immediately next, and is hence irrelevant to the local minima.) Let $\partial\gamma_{\text{im}}^j$ denote the two boundary endpoints, and $\partial\Gamma_{\text{im}} = \partial\gamma_{\text{im}}^1 \cup \cdots \cup \partial\gamma_{\text{im}}^m$.

2.3 Energy model and the structural theorem for local minima

For a given generic configuration Q in a 2-D open domain Ω, define the class of admissible curves

$$\mathcal{C} = \{\Gamma \mid \Gamma \text{ is simple closed, piecewise } C^2 \text{ with finitely many pieces, and } \Gamma \cap Q^\circ = \text{empty}\}. \quad (3)$$

For any admissible curve $\Gamma \in \mathcal{C}$, we define the energy

$$E_o[\Gamma] = E_o[\Gamma \mid \alpha, \beta] = \alpha \int_{\Gamma_{\text{re}}} ds + \beta \int_{\Gamma_{\text{im}}} ds, \quad (4)$$

where s denotes the arc-length parameter and $0 < \alpha \ll \beta$. The subscript "o" in E_o signifies that this energy is for *object*-based configuration Q. In the next section, we will introduce an energy E_c for *contour*-based configurations. Notice that E_o is the weighted sum of the lengths of Γ_{re} and Γ_{im}. We are interested in the behavior and characteristics of the local minimizers to E_o, in particular, in how well they can quantify more familiar features of illusory contours in visual cognition, which have often been subconsciously processed by human observers.

Lemma 2.3. *Let $\Gamma \in \mathcal{C}$ be a local minimizer of E_o. Then each connected component of Γ_{im} must be a straight line segment.*

Proof. Suppose otherwise that one connected component γ of Γ_{im} is not a straight line segment. Then there must exist at least one point $b \in \gamma$, within *any* of whose small neighborhoods, there can be found two points $a, c \in \gamma$ on the two sides of b such that $\{a, b, c\}$ are not collinear. We then replace $\gamma|_{[a,c]}$ by the straight line segment joining a and c. Since Γ and hence γ are piecewise C^2 and $\gamma \subset Q^c$ (an open set), as long as the neighborhood is small enough, the straight line substitute does not intersect the rest of Γ nor Q. Denote the new curve after such a micro-surgery by Γ'. Then $\Gamma' \in \mathcal{C}$ and $E_o[\Gamma'] < E_o[\Gamma]$ since the two only differ in between a and c. This contradicts to the assumption that Γ is a local minimizer since such micro-surgeries can occur in *any* local neighborhood of b. □

Corollary 2.4. *Suppose $\Gamma \in \mathcal{C}$ is a local minimizer of E_o with an empty real part $\Gamma_{\text{re}} = \varnothing$. Then $\Gamma = \varnothing$. As a result, any* non-empty *local minimizer must be pinned somewhere along ∂Q.*

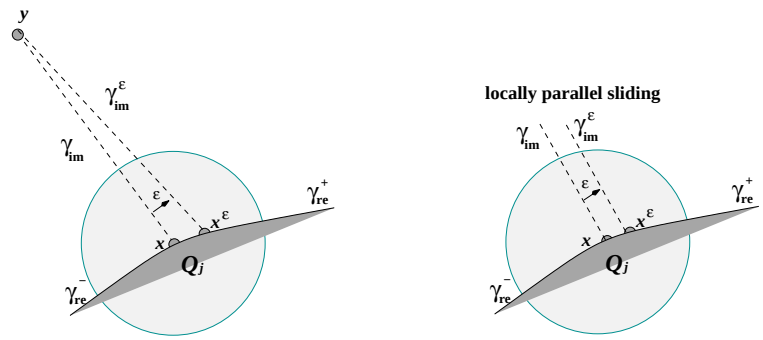


Figure 5: Local perturbation at a junction point $x \in \partial\Gamma_{\text{im}} \subset \partial Q$ can be approximated (in the first order of the displacement distance ε) by a simple parallel translation of the attached imaginary ray γ_{im}, due to Lemma 2.3 and the same mechanism that makes direct solar rays appear approximately parallel when reaching onto the Earth. Parallel shifting simplifies most computation.

Proof. By the preceding lemma, the entire $\Gamma = \Gamma_{\text{im}}$ must be a straight line segment. On the other hand, $\Gamma \in \mathcal{C}$ has to be a closed curve. The two could both hold only when Γ is empty. □

This is a desirable property of the model, since in reality illusory contours can indeed only arise from the cognitive interpolation of the *existing* edges. An illusory shape cannot solely emerge out of featureless blank areas.

Furthermore, the corollary also unveils one indispensable ingredient of any successful computing scheme for the model. That is, during the dynamic process of such a scheme, every point of ∂Q must be visited at least once, since without any *a priori* bias, every part of ∂Q can potentially contribute to Γ_{re}. In particular, the choice of an initial contour has to be properly made in a curve evolution scheme. One shall see from the later computational development how all these can be naturally implemented.

We now investigate the detailed geometrical and topological structures of the minimizers. In the following analysis, the local perturbation of a junction point $x \in \partial\Gamma_{\text{im}}$ along ∂Q (see $x \to x^{\varepsilon} \in \partial Q$ in Fig. 5) has always been accomplished by a simple parallel translation of the attached imaginary ray γ_{im}. This is justified by the preceding lemma, as well as the fact that when the other endpoint y is relatively far away, the perturbed imaginary ray $\overline{yx^{\varepsilon}}$ due to $x \to x^{\varepsilon}$ is parallel to $\overline{yx}$ in the first order (of ε) within small neighborhoods of x (referring to Fig. 5 and its caption).

Lemma 2.5 (Two distinct imaginary rays cannot share a hinge)**.** *Suppose $\Gamma \in \mathcal{C}$ is a local minimizer to $E_o = E_o[\cdot|\alpha, \beta]$, and γ^1 and γ^2 are two distinct connected components (which are straight line segments by the preceding lemma). Then there exists some critical ratio r_c, such that as long as $r = \alpha/\beta < r_c$, $\partial\gamma^1 \cap \partial\gamma^2$ must be empty.*

Proof. (See the companion illustration in Fig. 6) Otherwise, assume that γ^1 and γ^2 share a hinge at a point z (i.e., the point O in the figure). Then z must be a *real* point, and one can assume that $z \in \partial Q_j$ for some unique j.

Locally in a small diskette neighborhood of z, the two straight rays γ^1 and γ^2 segment the disk into two parts: an "inner" part that intersects with Q_j and an "outer" part. Let θ denote the outer span angle.

First suppose $\theta \in (0, \pi)$. Then any corner-cutting local deformation near z (as shown in the left panel of Figure 6) can decrease energy E_o since the the real part of the energy is unchanged but the imaginary one shrinks. Thus this scenario cannot occur.

Then one must have $\theta \geq \pi$. For ∂Q_j, similarly there are two *real* tangent rays at z, denoted by μ^1 and μ^2, so that the acute cone of μ^1 and γ^1 contains neither of the other two rays. Let ϕ_i denote the acute angle between μ^i and γ^i, $i = 1, 2$. Without loss of generality, assume that $\phi_1 \leq \phi_2$. Since

$$\phi_1 + \phi_2 \leq (2\pi - \theta) - \theta^*_{\min} \leq \pi - \theta^*_{\min},$$

one has

$$\phi_1 = \min(\phi_1, \phi_2) \leq \frac{\pi}{2} - \frac{\theta^*_{\min}}{2}.$$

Then as shown in the right panel of Fig. 6, a small parallel sliding (with perpendicular distance ε) of the root of the imaginary ray γ^1 along the real edge of ∂Q_j in the direction of μ^1 will introduce an energy increment of:

$$\alpha \frac{\varepsilon}{\sin \phi_1} - \beta \frac{\varepsilon}{\tan \phi_1} = \frac{\varepsilon \beta}{\sin \phi_1} \left(\frac{\alpha}{\beta} - \cos \phi_1 \right).$$

Define $r_c = \cos(\pi/2 - \theta^*_{\min}/2) = \sin(\theta^*_{\min}/2) > 0$. Then as long as $r = \alpha/\beta < r_c$, the leading order of energy increment is in fact negative, which again contradicts to the given assumption. To conclude, as long as $r < r_c$, two distinct imaginary line segments cannot share a hinge. □

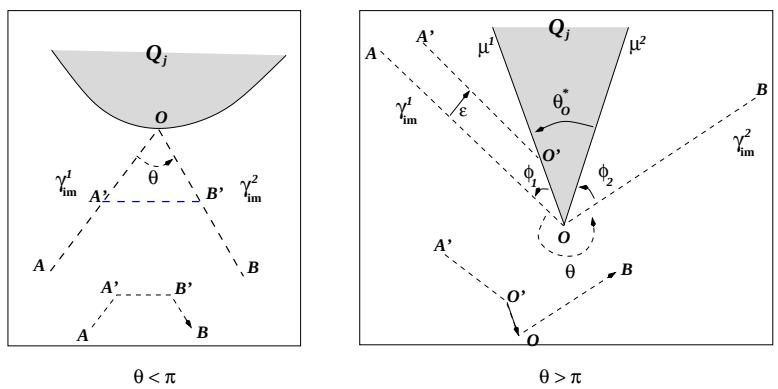


Figure 6: Companion illustration for the proof of Lemma 2.5: Two distinct imaginary rays γ^1_{im} and γ^2_{im} (of a local minimum contour Γ) cannot share a common hinge O (Lemma 2.5). **Left panel:** if $\theta < \pi$, a local perturbation of $\overline{AOB}$ to $\overline{AA'B'B}$ decreases the total energy E_o. **Right panel:** if $\theta \geq \pi$, a local perturbation of $\overline{AOB}$ to $\overline{A'O'OB}$ still decreases the total energy (referring to the computation in the proof of Lemma 2.5, especially the arising of the critical ratio $r_c = \alpha/\beta$ and the constraint $r < r_c$).

Definition 2.6 (Junction set $J[\Gamma]$ and a turn ϕ_z at a junction point z)**.** For a minimizer $\Gamma \in \mathcal{C}$, we define its junction set to be

$$J[\Gamma] = \partial \Gamma_{\text{im}}, \qquad \text{i.e., all endpoints of imaginary segments.}$$

By definition, one must have $J[\Gamma] \subseteq \partial Q$. For any junction point $z \in J[\Gamma]$, in one direction Γ goes to a uniquely associated imaginary ray, and let $\boldsymbol{t}_{\text{im}} = \boldsymbol{t}_{\text{im}}(z)$ denote this *outgoing* unit vector of orientation, and in the other direction comes a curve segment of Γ_{re} (according to the preceding lemma) and let $\boldsymbol{t}_{\text{re}} = \boldsymbol{t}_{\text{re}}(z)$ denote this *incoming* unit vector of orientation (Fig. 7). The turn at z is then defined to be the angle:

$$\phi_z = \arccos(\boldsymbol{t}_{\text{im}} \cdot \boldsymbol{t}_{\text{re}}) \in [0, \pi].$$

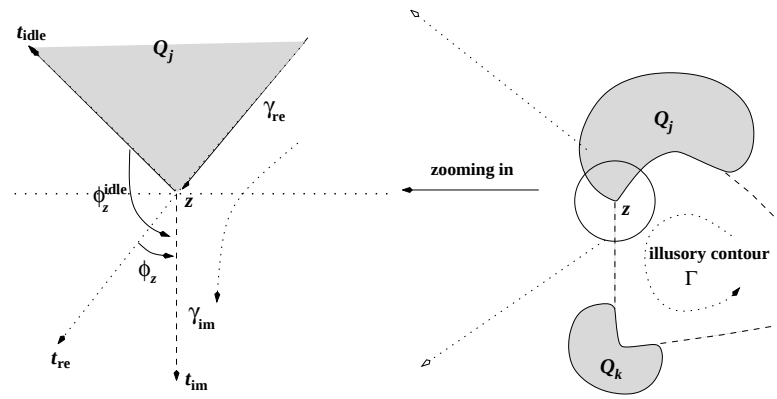


Figure 7: A permissible geometric layout at a junction point $z \in J[\Gamma]$ with the turn $\phi_z < \pi/2$ and idle turn $\phi_z \geq \pi/2$, according to the results established in Lemmas 2.7 and 2.10.

Lemma 2.7 (Acute turns at junction points)**.** *Suppose $\Gamma \in \mathcal{C}$ is a local minimum of E_o and $z \in J[\Gamma]$ is a junction point. Then the turn $\phi_z < \pi/2$.*

Proof. Suppose otherwise that $\phi_z \geq \pi/2$. Then if one slides the root z of the imaginary ray γ_{im} along the *incoming* part of Γ_{re} while keeping the imaginary rays parallel, as well as other parts of Γ intact, as in the left panel of Fig. 8, the leading order energy increment becomes:

$$-\beta \frac{\varepsilon}{\tan(\pi - \phi_z)} - \alpha \frac{\varepsilon}{\sin(\pi - \phi_z)} < 0,$$

which contradicts to the assumption of Γ being a local minimum. (Notice that in the critical case when $\phi_z = \pi/2$, the first term degenerates into a second-order change $O(\varepsilon^2)$. Thus the result still holds due to the second term.) □

This lemma has an intuitive interpretation in terms of visual cognition. Illusory contours more or less are smooth and straight extrapolations of existing real boundaries [5, 6], and consequently the turns at the junction points should generally be as small as possible.

Definition 2.8 (Maximum turn of a local-minimum contour Γ)**.** Suppose $\Gamma \in \mathcal{C}$ is a local minimum of E_o. Define the maximum turn $\phi_{\max}$ of Γ by

$$\phi_{\max} = \max\{\phi_z \mid z \in J[\Gamma]\}.$$

Then the preceding lemma implies that $\phi_{\max} < \pi/2$.

Definition 2.9 (Idle angle ϕ^{idle} at a junction point)**.** Let $\Gamma \in \mathcal{C}$ be a local minimum to E_o, and $z \in J[\Gamma]$ a junction point. Define the imaginary tangent $\boldsymbol{t}_{\text{im}} = \boldsymbol{t}_{\text{im}}(z)$ as in Definition 2.6. Further define the *idle* tangent $\boldsymbol{t}_{\text{idle}} = \boldsymbol{t}_{\text{idle}}(z)$ to be the unit outgoing tangent vector of $\partial Q \setminus \Gamma_{\text{re}}$ (i.e., the *idle* real boundaries that do not participate in the formation of Γ) at z (Fig. 8). Define the idle angle at z by

$$\phi_z^{\text{idle}} = \arccos(\boldsymbol{t}_{\text{idle}} \cdot \boldsymbol{t}_{\text{im}}).$$

Lemma 2.10 (Behavior of idle angles at junctions)**.** *Suppose $\Gamma \in \mathcal{C}$ is a local minimum to $E_o[\cdot \mid \alpha, \beta]$ for* any *$\alpha, \beta > 0$, and $z \in J[\Gamma]$ is one of its junction point. Then $\phi_z^{\text{idle}} \geq \pi/2$.*

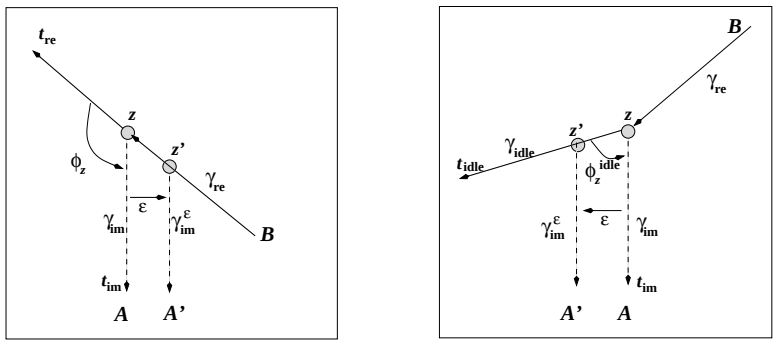


Figure 8: Companion illustration for the proofs of Lemmas 2.7 and 2.10. **Left panel:** the turn ϕ_z at a junction point $z \in J[\Gamma]$ cannot be blunt, otherwise the local perturbation of z to z' and $\gamma_{\rm im}$ to $\gamma^{\varepsilon}_{\rm im}$ decreases both the real and imaginary energies in E_o. **Right panel:** the idle turn $\phi^{\rm idle}_z$ cannot be acute for sufficiently small α/β, otherwise the perturbation shown will cause a net drop in the total energy E_o (referring to the computation in the proof of Lemma 2.10).

Proof. Suppose otherwise that $\phi^{\rm idle}_z < \pi/2$. Then if one slides the root z of the imaginary ray $\gamma_{\rm im}$ along the *outgoing* idle part of $\partial Q \setminus \Gamma_{\rm re}$ while keeping the imaginary rays parallel, as well as other parts of Γ intact, as in the right panel of Fig. 8, the leading order energy increment becomes:

$$\alpha \frac{\varepsilon}{\sin \phi^{\rm idle}_z} - \beta \frac{\varepsilon}{\tan \phi^{\rm idle}_z} = \frac{\varepsilon \beta}{\sin \phi^{\rm idle}_z} \left(\frac{\alpha}{\beta} - \cos(\phi^{\rm idle}_z) \right). \tag{5}$$

Since Γ is a local minimum for *any* $\alpha, \beta > 0$, one must have

$$-\cos(\phi^{\rm idle}_z) \geq 0, \qquad \text{or} \qquad \phi^{\rm idle}_z \geq \pi/2.$$

□

Combining all the characterizations established in the lemmas of this section, we are ready to state the main theorem of this paper.

Theorem 2.11 (Characterization of a local minimum)**.** *Let Q be a generic configuration on an open domain Ω, and $\theta_{\rm min}, \theta^*_{\rm min} > 0$ its minimum outer and inner spans. Suppose a given simple closed curve $\Gamma \in \mathcal{C}$ satisfies the following structural conditions:*

(i) (Imaginary behavior) each connected component $\gamma_{\rm im}$ of $\Gamma_{\rm im}$ is a straight line segment, and no two distinct components share a common hinge;

(ii) (Junction behavior) at any junction point $z \in J[\Gamma]$, the turn $\phi_z < \pi/2$, and the idle angle $\phi^{\rm idle}_z \geq \pi/2$.

Let $\phi_{\rm max}$ denote the maximum turn on $J[\Gamma]$. Then there exists a critical ratio $r_c = r_c(\theta_{\rm min}, \phi_{\rm max}) < 1$, such that for any α and β with $r = \alpha/\beta < r_c$, Γ is a local minimum to the energy $E_o[\cdot \mid \alpha, \beta]$.

Proof. We shall only outline the main steps of the proof and leave the details to the reader, which are very similar to the proofs of the above ensemble of lemmas. Let $J[\Gamma]$ denote the junction set of such a given Γ, and define

$$\Gamma^{\circ}_{\rm re} = \Gamma_{\rm re} \setminus J[\Gamma].$$

Then Γ is segmented into three distinct parts:

$$\Gamma = \Gamma_{\rm im} \cup \Gamma_{\rm re}^{\circ} \cup J[\Gamma],$$

each of which leads to a distinct energy variation pattern due to local admissible perturbations.

First, it is easy to see that under Condition (i) any local perturbations at a point of $\Gamma_{\rm im}$ will increase the energy.

Next, $\Gamma_{\rm re}^{\circ}$ can be further partitioned into kink points $\Gamma_{\rm re}^{\circ} \cap K[Q]$ and the majority rest of smooth points. At a smooth point y, any *generic* local variation drifts a *real* local segment γ into Q^c and converts it into an imaginary one, and hence introduces a net leading-order energy increment of $(\beta - \alpha) \int_\gamma ds$, which is always positive for $r = \alpha/\beta < 1$. At a kink point y on $\Gamma_{\rm re}^{\circ}$, one defines $r_1 = \sin\theta_{\max}/2 > 0$. Then simple trigonometric calculus shows that for any α and β with $\alpha/\beta < r_1$, any local admissible displacement of the kink point always increases the energy.

Finally, one turns to the local perturbations at any junction point $z \in J[\Gamma]$. By Condition (ii) on the acute turns, one must have that the maximum turn $\phi_{\max} < \pi/2$. Define $r_2 = \cos\phi_{\max} > 0$. Then trigonometric calculus shows that any local drifting of z into the *incoming* real contour $\Gamma_{\rm re}$ will increase the energy as long as $\alpha/\beta < r_2$. On the other hand, any local drifting of z into the *idle* real edge $\partial Q \setminus \Gamma_{\rm re}$ always increases energy due to Condition (ii) on $\phi_z^{\rm idle} \geq \pi/2$, for any $\alpha, \beta > 0$. The details are very similar to the proofs of Lemmas 2.7 and 2.10.

In combination, one defines the critical ratio $r_c = \min(r_1, r_2, 1)$. Then for any α and β with $\alpha/\beta < r_c$, the above analysis in all the three scenarios shows that the energy always increases under any local perturbations of Γ. Thus Γ must be a local minimum to $E_o[\cdot \mid \alpha, \beta]$. □

By this theorem, it is straightforward to establish the following corollary, which demonstrates that local minima of E_o are indeed reasonable models for illusory contours, at least in terms of the leading-order approximation.

Corollary 2.12. *Both Kanizsa's triangle and square are local minimizers to $E_o[\Gamma \mid \alpha, \beta]$ for sufficiently small α/β.*

Remark 2.13. In the preceding main theorem, only the condition $\phi_z^{\rm idle} \geq \pi/2$ could be relaxed to $\phi_z^{\rm idle} \geq \pi/2 - \delta$ for some constant $\delta = \delta(\alpha, \beta) \geq 0$ with $\delta \to 0$ as $r = \alpha/\beta \to 0$ (due to Eqn. (5)). The preceding lemmas show that all the other conditions are necessary and cannot be further relaxed.

3 Real contour bundles and object-contour mixtures

In this section, we briefly discuss contour-based illusory contours, as well as those arising from a mixture configuration of both objects and contours. Computationally, however, it can be handled by exactly the same scheme for the object-based model, as to be numerically demonstrated later on.

Let $L = \{l_i\}_{i=1}^n$ be a bundle of disjoint compact contour segments on a given open image domain Ω. To be distinguished from illusory contours, each l_i shall be called a *real* contour (Fig. 9). Let $\partial l_i = \{a_i, b_i\}$ denote the two endpoints of l_i. Define

$$\partial L = \cup_{i=1}^n \partial l_i = \cup_{i=1}^n \{a_i, b_i\}, \qquad L^\circ = L \setminus \partial L.$$

As for object-based illusory contours, we define the class of contour-based admissible contours by

$$\mathcal{C}_c = \{\Gamma \mid \Gamma \text{ is simple closed, piecewise } C^2, \text{ and } \Gamma \cap L^\circ = \text{empty}\}.$$

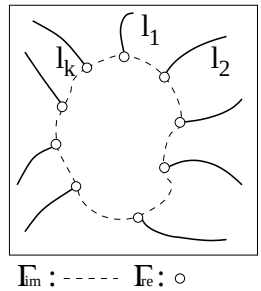


Figure 9: Decomposing a candidate contour Γ into its real part Γ_{re} and imaginary part Γ_{im}.

For each admissible contour $\Gamma \in \mathcal{C}_c$, we define, in the same spirit of the previous object-based case, the real-imaginary decomposition (Fig. 9):

$$\Gamma_{\text{re}} = \Gamma \cap \partial L, \quad \Gamma_{\text{im}} = \Gamma \setminus \partial L = \Gamma \setminus \Gamma_{\text{re}}.$$

As a result, the previous object-based energy E_o is modified to the following contour-based:

$$E_c[\Gamma] = -\alpha \mathcal{H}^0(\Gamma_{\text{re}}) + \beta \int_{\Gamma_{\text{im}}} ds = -\alpha \mathcal{H}^0(\Gamma_{\text{re}}) + \beta \int_{\Gamma} ds, \tag{6}$$

with $\alpha, \beta > 0$. Here $\mathcal{H}^0$ denotes the 0-dimensional Hausdorff measure, or equivalently, the atomic counting measure.

The major difference of E_c from the previous object-based energy E_o resides in the first negative term. Under this new formulation, any real end point from ∂L acts like an energy basin, in order to be able to anchor prospective illusory contours.

The local minimizers of E_c are much easier to characterize.

Theorem 3.1. *$\Gamma \in \mathcal{C}_c$ is a local minimizer with a non-empty imaginary part Γ_{im} if and only if it is a polygon whose vertices are all anchored within ∂L.*

Proof. First suppose Γ is a local minimizer with a non-empty imaginary part Γ_{im}. For any connected component γ of Γ_{im}, the same argument as in Lemma 2.3 applies since a local perturbation of γ only modifies the length integral in the β-term. Thus γ must be a straight line segment. Since γ is arbitrary, $\Gamma \in \mathcal{C}_c$ has to be a polygon whose vertices are anchored within ∂L.

Conversely, suppose $\Gamma \in \mathcal{C}_c$ is a polygon whose vertices belong to ∂L. Any local variation *within* a side γ of Γ increases the β-term line integral while leaving the atomic α-term unchanged, since γ is straight. On the other hand, any nonzero local perturbation of a vertex of Γ will drift it into $\Omega \setminus L$ (since the perturbation must be kept within the admissible class $\mathcal{C}_c$), and consequently cause a leading-order energy increment of α. Hence Γ is a local minimizer, and the proof is complete. □

Finally, some examples in the literature also involve illusory contours from a mixture configuration of both objects and contours. For this reason, we consider a mixture model briefly. Consider an image which is composed of compact and mutually disjoint real objects Q and real contours L. Let Γ be an admissible piecewise C^2 simple closed curve in Ω that does not intersect $Q^\circ \cup L^\circ$ (defined previously). We define the real and imaginary part of Γ as follows:

$$\begin{aligned} \Gamma_{\text{re}} &= \Gamma_{\text{re_o}} \cup \Gamma_{\text{re_c}} = (\Gamma \cap \partial Q) \cup (\Gamma \cap \partial L), \\ \Gamma_{\text{im}} &= \Gamma \setminus \Gamma_{\text{re}} = \Gamma \setminus (\partial Q \cup \partial L). \end{aligned}$$

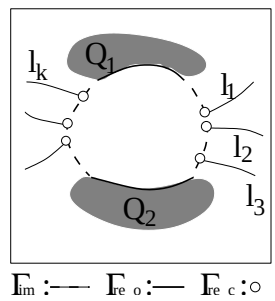


Figure 10: Decomposition of a candidate curve Γ in an object-contour mixture configuration.

Notice that here the real part has been further categorized into object-based real part $\Gamma_{\text{re_o}}$ and contour-based real part $\Gamma_{\text{re_c}}$. We then define a mixture energy functional E_m in the form of:

$$E_m[\Gamma] = \alpha_o \int_{\Gamma_{\text{re_o}}} ds - \alpha_c \mathcal{H}^0(\Gamma_{\text{re_c}}) + \beta \int_{\Gamma_{\text{im}}} ds.$$

where $0 < \alpha_o \ll \beta$ and $\alpha_c > 0$.

The characterization of a local minimizer to this mixture energy can be readily obtained from a straightforward combination of the previous two major theorems. In particular, for example, if Γ is a local minimizer, then two distinct connected components γ^1 and γ^2 of the imaginary part *can* share a common hinge, but the hinge point must belong to ∂L and not to ∂Q.

Next we discuss how to compute these models in real applications. In particular, we will demonstrate numerically that for practical applications, a single relaxed model is sufficient to handle all the previous three theoretical models of E_o, E_c, and E_m.

4 Relaxation of the model: approximation by active contours

The main models have been built upon the real-imaginary decomposition of any candidate contour Γ, which in turn depends upon the identifiability of the *real* edge set ∂Q. Hence arise two major practical issues for implementing these models realistically. First, for general images, one has to perform a pre-processing of edge detection in order to obtain ∂Q, which is easier for binary images but highly nontrivial for general continuous-tone images. Second, even provided with ∂Q, it is well known that direct computation of curves is highly challenging due to the densening or coarsening effects of discrete points as well as potential topological changes like splitting and merging.

The current section resolves the first issue, while the next section will address the second.

The main result of this section is to show that the edge-detection process can be implicitly integrated into an approximate model, which surprisingly turns out to be the more familiar active contour model [8, 7, 13, 16]. That is, the proposed new model for illusory contours can be asymptotically approximated by active contours when restricted in the class of admissible contours $\mathcal{C}$.

First consider an ideal binary image $u(x) = \chi_Q(x)$, i.e., the characteristic function of a generic configuration Q. Let $\eta(x)$ be any C^∞ positive function which is compactly supported on the unit disk in the plane, and for any scale $\sigma > 0$, define the mollifier $\eta_\sigma(x) = \frac{1}{\sigma^2}\eta(\frac{x}{\sigma})$, and the mollification $u_\sigma = u * \eta_\sigma$. Since Q is assumed to be compact in the open image domain Ω, u_σ is well defined as long as σ is smaller than the distance $\text{dist}(Q, \partial\Omega)$.

Then for any admissible curve $\Gamma \in \mathcal{C}$, we introduce the following approximate energy model

$$E_a[\Gamma|u] = \alpha \int_\Gamma ds + \beta \int_\Gamma g(|\nabla u_\sigma(s)|)ds, \tag{7}$$

where $g(t) = \frac{1}{1+\lambda t^2}$ for some fixed parameter $\lambda > 0$. This is in fact precisely the classical active contour model, where $g(|\nabla u_\sigma(s)|)$ is an edge signature distribution, or a *soft* edge indicator [13]. The next theorem explains the relation between E_o and E_a.

Theorem 4.1. *For any admissible curve $\Gamma \in \mathcal{C}$, one has* $\displaystyle\int_\Gamma \frac{1}{1+\lambda|\nabla u_\sigma|^2} ds \longrightarrow \int_{\Gamma_{\rm im}} ds$ *as $\sigma \to 0$.*

Proof. Denote $B_\sigma = \{x \in \Omega \,|\, \mathrm{dist}(x, \partial Q) \le \sigma\}$. Then $u_\sigma = 1$ in $Q \setminus B_\sigma$ and $u_\sigma = 0$ in $Q^c \setminus B_\sigma$. As a result, $\nabla u_\sigma = 0$ and $\frac{1}{1+\lambda|\nabla u_\sigma|^2} = 1$ in $\Omega \setminus B_\sigma$. Since as $\sigma \to 0$, B_σ shrinks to ∂Q, one has

$$\frac{1}{1+\lambda|\nabla u_\sigma|^2} \longrightarrow 1 \text{ as } \sigma \to 0 \text{ pointwise in } \Omega \setminus \partial Q. \tag{8}$$

Since u is a characteristic function, one must have $|\nabla u_\sigma| \to \infty$ as $\sigma \to 0$ along the piecewise smooth edge ∂Q, and

$$\frac{1}{1+\lambda|\nabla u_\sigma|^2} \longrightarrow 0 \text{ as } \sigma \to 0 \text{ on } \partial Q. \tag{9}$$

In combination of (8) and (9), one has

$$\lim_{\sigma\to 0} \frac{1}{1+\lambda|\nabla u_\sigma|^2} = \chi_{\Omega\setminus\partial Q},$$

pointwise on the entire image domain Ω. Since $\Gamma \in \mathcal{C}$ is piecewise smooth, simple closed, and $\Gamma \cap Q^\circ$ is empty, by applying *Lebesgue's dominated convergence theorem* on Γ with respect to its arclength measure $ds|_\Gamma$ [10], one has, as $\sigma \to 0$,

$$\int_\Gamma \frac{1}{1+\lambda|\nabla u_\sigma|^2} ds \longrightarrow \int_\Gamma \chi_{\Omega\setminus\partial Q} ds = \int_{\Gamma\setminus Q} ds = \int_{\Gamma_{\rm im}} ds.$$

This completes the proof. □

Consequently, as $\sigma \to 0$, one has for any admissible contour $\Gamma \in \mathcal{C}$,

$$E_a[\Gamma \mid u] \longrightarrow \alpha \int_{\Gamma_{\rm re}} ds + (\alpha+\beta) \int_{\Gamma_{\rm im}} ds.$$

In particular, the current parameter pair $(\alpha, \alpha+\beta)$ corresponds to the original parameter pair (α, β) in E_o. The approximate energy E_a is, however, computationally much more tractable since the practically difficult decomposition $\Gamma = \Gamma_{\rm re} \cup \Gamma_{\rm im}$ is replaced by much simpler function evaluations of g and u_σ. This technique is well known in the literature of active contours [7, 13].

Furthermore, the active-contour approximation E_a, rigorously established above for binary images, is now readily applicable to general images which could have continuous tones or have been degraded by moderate noises.

Remark 4.2. Due to mollification, 1-D real contours in a mixture configuration are mollified to 2-D stripes. As a result, our later computational results show that the relaxed active-contour model works equally well for contour configurations or object-contour mixtures.

Remark 4.3. It is well known that classical (non-region-based) active contours are often easily trapped in local minima [7]. This no longer becomes an issue herein since illusory contours have been actually modelled as *local* minima, as extensively developed in the preceding sections. That is, it would be bad news if the active contour model E_a never stably stagnates at local minima.

The key question becomes how to compute robustly those local minima that correspond to meaningful illusory contours in cognition. This issue is resolved by the *supervised* level-set computation, which is the main theme of the next section.

5 Supervised level-set-based robust computation

In this section, we derive the level set formulation to the approximate model (7) via classical active contours, and introduce a simple but key step of *supervision* to be able to extract illusory contours as stable local minima.

5.1 Derivation of the Euler-Lagrange equation

The following brief derivation of the Euler-Lagrange equation is canonical (e.g., [7, 13]) and makes the current paper more self-contained.

To derive the formal Euler-Lagrange equation, the candidate contour Γ is assumed to be at least C^2 so that second-order geometric features like curvature κ are well defined. We denote by $\Phi(x) = \Phi(x \mid u, \sigma) = g(|\nabla u_\sigma(x)|)$ for convenience. Then

$$E[\Gamma|u] = \alpha \int_\Gamma ds + \beta \int_\Gamma \Phi(s) ds,$$

where ds denotes the arc-length element of Γ. Since Γ is assumed to be C^2, any smooth perturbation of Γ_r of Γ can be represented as a displacement along the normal $\boldsymbol{n}$:

$$\mathbf{c}_r(s) = \mathbf{c}(s) + r\delta\mathbf{c}(s) = \mathbf{c}(s) + rh(s)\boldsymbol{n}(s),\ r \in (-\varepsilon, \varepsilon), \qquad h = O(1).$$

Then by Frenet's frame formula [7], $\boldsymbol{n}'(s) = -\kappa(s)\boldsymbol{t}(s)$, and

$$\mathbf{c}_r'(s) = (1 - rh(s)\kappa(s))\boldsymbol{t}(s) + rh'(s)\boldsymbol{n}(s),$$

$$|\mathbf{c}_r'(s)| = \sqrt{(1 - rh(s)\kappa(s))^2 + r^2 h'(s)^2}.$$

Suppose $s \in [a, b]$ is the domain of $\mathbf{c}$ and $\tilde{s}$ denotes the arc-length parameter of $\mathbf{c}_r$ so that $d\tilde{s} = |\mathbf{c}_r'(s)|ds$. Then

$$\int_{\Gamma_r} d\tilde{s} = \int_a^b |\mathbf{c}_r'(s)| ds.$$

$$\frac{d}{dr}\int_{\Gamma_r} d\tilde{s} = \int_a^b \frac{\partial}{\partial r}|\mathbf{c}_r'(s)| ds = \int_a^b \frac{-(1 - rh(s)\kappa(s))h(s)\kappa(s) + rh'(s)^2}{\sqrt{(1 - rh(s)\kappa(s))^2 + r^2 h'(s)^2}} ds.$$

$$\begin{aligned} \frac{d}{dr}\bigg|_{r=0} \int_{\Gamma_r} d\tilde{s} &= \int_a^b -h(s)\kappa(s) ds \\ &= \int_\Gamma -\kappa(s)\boldsymbol{n}(s) \cdot h(s)\boldsymbol{n}(s) ds. \end{aligned} \tag{10}$$

On the other hand, for the second term, by noticing that

$$\frac{\partial}{\partial r}\Phi(\mathbf{c}_r(s)) = \nabla\Phi \cdot \frac{\partial}{\partial r}\mathbf{c}_r(s) = \nabla\Phi \cdot h(s)\boldsymbol{n}(s) = h(s)\frac{\partial\Phi}{\partial \boldsymbol{n}}(\mathbf{c}(s)),$$

one has

$$\frac{d}{dr}\int_{\Gamma_r}\Phi d\tilde{s} = \int_a^b \frac{\partial}{\partial r}\Phi(\mathbf{c}_r(s))|\mathbf{c}_r'(s)|ds + \int_a^b \Phi(\mathbf{c}_r(s))\frac{\partial}{\partial r}|\mathbf{c}_r'(s)|ds,$$

$$\frac{d}{dr}\bigg|_{r=0}\int_{\Gamma_r}\Phi d\tilde{s} = \int_a^b h(s)\frac{\partial\Phi}{\partial \boldsymbol{n}}(\mathbf{c}(s))ds - \int_a^b \Phi(\mathbf{c}(s))h(s)\kappa(s)ds. \tag{11}$$

In combination of (10) and (11), one has

$$\begin{aligned}\frac{d}{dr}\bigg|_{r=0} E_a[\Gamma_r|u] = &-\alpha\int_\Gamma \kappa(s)\boldsymbol{n}(s)\cdot\delta\mathbf{c}(s)ds \\ &-\beta\int_\Gamma\left(\Phi(s)\kappa(s)\boldsymbol{n}(s) - \frac{\partial\Phi}{\partial\boldsymbol{n}}(s)\boldsymbol{n}(s)\right)\cdot\delta\mathbf{c}(s)ds.\end{aligned}$$

As a result, after introducing t as the artificial evolution time, we arrive at the gradient-descent equation for $E_a[\Gamma|u]$:

$$\frac{\partial\mathbf{c}}{\partial t} = (\alpha+\beta\Phi)\kappa\boldsymbol{n} - \beta\Phi_{\boldsymbol{n}}\boldsymbol{n}. \tag{12}$$

In particular, if one defines

$$G(x) = G(x \mid \alpha,\beta,\sigma,u) = \alpha + \beta\Phi(x) = \alpha + \beta g(|\nabla u_\sigma(x)|),$$

then (12) simplifies to

$$\frac{\partial\mathbf{c}}{\partial t} = G\kappa\boldsymbol{n} - G_{\boldsymbol{n}}\boldsymbol{n}. \tag{13}$$

This is the canonical "snake" equation in its original form (e.g., [7, 13]).

5.2 Supervised computation via the level-set method

The level-set method of Osher and Sethian [24, 23, 29] is a powerful computational tool to handle curve and interface motions, especially those that intrinsically involve topological changes. The above active contour equation has a simple formulation under the level-set representation.

Suppose the evolving contour $\mathbf{c}(t,s)$ is always identified as the zero-level contour of an evolving level-set function $\phi(t,x)$, which is smooth or at least Lipschitz continuous:

$$\phi(t,\mathbf{c}(t,s)) = 0,\ \forall t \geq 0. \tag{14}$$

Assume for convenience that ϕ is positive in the interior of $\mathbf{c}(t,s)$. By differentiating (14) with respect to t, one has

$$\frac{\partial\phi}{\partial t} + \left\langle \nabla\phi, \frac{\partial\mathbf{c}}{\partial t}\right\rangle = 0.$$

Recall that for an implicit contour, the curvature, and tangent and normal vectors can be represented as follows:

$$\kappa = -\nabla\cdot\left[\frac{\nabla\phi}{|\nabla\phi|}\right], \quad \boldsymbol{t} = \frac{\nabla\phi^\perp}{|\nabla\phi|}, \quad \boldsymbol{n} = \frac{\nabla\phi}{|\nabla\phi|}. \tag{15}$$

Thus the active-contour model (13) allows the following level-set representation:

$$\frac{\partial \phi}{\partial t} + \left\langle \nabla\phi, -G\nabla\cdot\left[\frac{\nabla\phi}{|\nabla\phi|}\right]\frac{\nabla\phi}{|\nabla\phi|} - G_{\boldsymbol{n}}\frac{\nabla\phi}{|\nabla\phi|}\right\rangle = 0.$$

It can be further simplified to

$$\frac{\partial \phi}{\partial t} = |\nabla\phi| G\nabla\cdot\left[\frac{\nabla\phi}{|\nabla\phi|}\right] + |\nabla\phi| G_{\boldsymbol{n}}. \tag{16}$$

Since $\nabla G = G_{\boldsymbol{t}}\boldsymbol{t} + G_{\boldsymbol{n}}\boldsymbol{n}$, one has

$$\nabla G\cdot\nabla\phi = |\nabla\phi|(\nabla G\cdot\boldsymbol{n}) = |\nabla\phi| G_{\boldsymbol{n}}, \tag{17}$$

and consequently,

$$\frac{\partial \phi}{\partial t} = |\nabla\phi|\nabla\cdot\left[G\frac{\nabla\phi}{|\nabla\phi|}\right]. \tag{18}$$

In the variational-PDE literature, this evolutionary partial differential equation is integrated via the Neumann boundary condition $\frac{\partial\phi}{\partial\nu} = 0$ along $\partial\Omega$.

Since in our model, illusory contours have been modelled as *local* minima, the initial contour or the initial configuration of the level-set function becomes important. In the numerical computation, we have chosen the initial function ϕ_0 such that $\phi_0 = 1$ in Ω_ε and -1 otherwise. Here Ω_ε denotes an open sub-domain of Ω such that the Hausdorff distance between Ω and Ω_ε is less than ε. That is, Ω_ε is a slightly contracted version of Ω. As long as the targeted illusory contour is a compact set on Ω, Ω_ε can capture it within for small enough ε.

Following Corollary 2.4 and its subsequent discussion therein, an alternative and more efficient choice for the initial level set function ϕ_0 can be made as follows. Let H denote the convex hull of the given configuration Q, and H_ϵ a slightly expanded (or fattened) version of H. Then one could take:

$$\phi_0(x) = \begin{cases} 1, & x \in H_\epsilon \\ -1, & x \in \Omega\setminus H_\epsilon. \end{cases}$$

In this way, all the *real* edges in ∂Q are efficiently contained within the initial interior ($\phi_0 > 0$), and the subsequent level-set evolution will have the potential to visit each real-edge point along ∂Q - a property necessitated by Corollary 2.4 and its implications. However, this alternative choice demands the computation of the convex hull of a given configuration, which could be expensive for an intricate layout.

Finally, unlike the classical active-contour computation whose goal is to capture *real* boundaries of existent objects, the active-contour model E_a in the present context approximates our original contour model E_o and hence is targeted at capturing illusory contours. There is a side condition which has been stated theoretically throughout the earlier discussion but not yet incorporated into the level-set formulation. In the definition of the admissible class $\mathcal{C}$ of candidate illusory contours, we have required that any admissible contour Γ must lie *outside* the object configuration Q, i.e., that $\Gamma\cap Q^\circ$ (i.e., the *open* interior of Q) is empty.

Classical level-set functions for active contours ideally prefer $\phi > 0$ on Q in order to capture *real* edges and *real* objects, but the above discussion shows that for the computing of illusory contours, ideally one prefers $\phi < 0$ on the object configuration Q. Thus we impose the following supervision condition for the level-set computation:

$$\phi(t,x) \equiv -d_Q(x), \qquad x\in Q^\circ,\ \ t>0,$$

where $d_Q(x) = \text{dist}(x, \partial Q)$ is the absolute distance function. In reality, for computation on a discrete image lattice, the supervision can be conveniently replaced by the less complex one:

$$\phi(t,x) \equiv -1, \qquad x \in Q^\circ, \;\; t > 0.$$

On one hand, it substantially lessens the computational burden without turning to the eikonal equation for $d_Q(x)$, and on the other, it provides a cheaper way to minimizing the second term in Zhu and Chan's model (1) without turning to either the variational formulation or its associated Euler-Lagrange equation. To conclude, the last supervision condition is an efficient tool to reinforce the contour admissibility condition "$\Gamma \cap Q^\circ = \text{empty}$" in the original exact model.

6 Implementation details and computational examples

In this last section, we explain the numerical implementation and discuss several generic numerical examples. The equation and conditions in the preceding section 5.2 can be reiterated by

$$\begin{cases} \dfrac{\partial \phi}{\partial t} = |\nabla \phi| \nabla \cdot \left[G \dfrac{\nabla \phi}{|\nabla \phi|} \right] \text{ for } (t,x) \in (0,\infty) \times \Omega, \\ \phi(0,x) = \phi_0(x), \\ \dfrac{\partial \phi}{\partial \nu} = 0 \text{ for } (t,x) \in (0,\infty) \times \partial\Omega, \\ \phi(t,x) = -1 \text{ for } (t,x) \in (0,\infty) \times Q^\circ, \end{cases} \tag{19}$$

where $G(x) = \alpha + \beta g(|\nabla u_\sigma(x)|)$.

Let $\phi_{i,j}^m$ be the approximation to the value $\phi(t_m, x_i^1, x_j^2)$, where $x_i^1 = i\Delta x_1$, $x_j^2 = j\Delta x_2$ with $x = (x_1, x_2)$, and $t_m = m\Delta t$. For simplicity, we assume the spatial sizes coincide: $\Delta x_1 = \Delta x_2 = 1$. Then,

$$\phi_{i,j}^{m+1} \approx \phi_{i,j}^m + \Delta t |\nabla \phi_{i,j}^m| \nabla \cdot \left[\left(\alpha + \beta g(|(\nabla u_\sigma)_{i,j}^m|) \right) \frac{\nabla \phi_{i,j}^m}{|\nabla \phi_{i,j}^m|_\epsilon} \right], \tag{20}$$

where $|\cdot|_\epsilon$ is defined by

$$|s|_\epsilon := \sqrt{s^2 + \epsilon^2},$$

for some positive $\epsilon \ll 1$. This simple regularization technique helps the computation of the denominator $|\nabla \phi|$ when it is close to be singular (e.g., [26, 31, 30]). We now focus on the spatial discretization of the right hand side at a fixed time t_m, and shall leave out the superscript m for convenience. We have employed central differencing for differential discretization, and mean filtering for interpolation. Write

$$\overrightarrow{V}_{i,j} = (V_{i,j}^1, V_{i,j}^2) = (\alpha + \beta g(|(\nabla u_\sigma)_{i,j}|)) \frac{\nabla \phi_{i,j}}{|\nabla \phi_{i,j}|}.$$

Based on central differencing at the half points of the Cartesian lattice, one has

$$\begin{aligned} \nabla \cdot \overrightarrow{V}_{i,j} &= \frac{\partial}{\partial x} V_{i,j}^1 + \frac{\partial}{\partial y} V_{i,j}^2 \\ &\approx \left(V_{i+\frac{1}{2},j}^1 - V_{i-\frac{1}{2},j}^1 \right) + \left(V_{i,j+\frac{1}{2}}^2 - V_{i,j-\frac{1}{2}}^2 \right). \end{aligned} \tag{21}$$

Thus we need to specify the half-point values in (21). For example, at $(i+\frac{1}{2}, j)$, one has

$$\frac{\partial}{\partial x}\phi_{i+\frac{1}{2},j} \approx \phi_{i+1,j} - \phi_{i,j},$$
$$\frac{\partial}{\partial y}\phi_{i+\frac{1}{2},j} \approx \frac{1}{2}\left(\frac{\phi_{i+1,j+1} - \phi_{i+1,j-1}}{2} + \frac{\phi_{i,j+1} - \phi_{i,j-1}}{2}\right).$$

To satisfy the supervision condition $\phi(t,x) = -1$ for $(t,x) \in (0,\infty) \times Q^\circ$, we reset $\phi_{i,j}^{m+1} = -1$ on Q° after each discrete evolution (20). For more details on numerical discretization, we also refer the reader to [4].

Shown in Fig. 11 through 15 are computational outputs on some generic examples of illusory contours, including the well known classical examples of Kanizsa's triangle (Fig. 11) and square (Fig. 12).

Fig. 13 shows an example of illusory contours from a configuration of contour bundles, as theoretically modelled in Section 3. Due to the mollification effect, the supervised level-set-based active contour model, which has been mainly developed for *object*-based illusory contours, works equally well for *contour*-based illusory contours in terms of computation.

Fig. 14 shows a computed illusory contour emerging from a more complex configuration Q of objects. The small hump remaining in the final output is not due to numerical defection but to the competition between the two terms in the model E_o, or equivalently, between α and β.

Finally, Fig. 15 shows the first-order nature of our proposed model. In order to be able to output smoother illusory shapes, high order geometric features such as the curvature have to be incorporated, as in Zhu and Chan's earlier works [37, 38]. The tradeoff is that comprehensive analytical results as in Section 2 would be extremely difficult to develop, which is well known in the literature of variational modelling involving curvatures [3, 19].

The captions of all the figures also provide further on-site explanation and discussion.

Finally, we make the following remarks on computing before concluding the entire work.

(A) **Stability of the Final Outputs.** With all the concrete examples handy, we now further highlight the local stability of the final outputs from the algorithm, which will also explain why the evolution always stabilizes at those particular configurations and does not shrink further towards vanishing. By both the theoretical prediction in Lemma 2.3 and all the final computational outputs, the imaginary parts Γ_{im}'s consist of straight line segments, which are already the shortest locally and hence stable, and do not generate any shrinking force (or more rigorously, any mean-curvature motion). For any segment γ of the real parts Γ_{re}'s, on the other hand, a detachment from the existing objects will make it part of Γ_{im}'s, and thus generates an energy increment in the net amount of

$$\beta \int_\gamma ds - \alpha \int_\gamma ds = (\beta - \alpha)|\gamma| > 0,$$

since the main structure theorem in Section 2 has required $r = \alpha/\beta \le r_c < 1$. Thus, the observed robust stability can be summarized as: "*the real parts stably lie along the energy trough provided by ∂Q, as well as stably pull the imaginary parts in the same manner as the fingers of a person expand a closed rubber band.*"

(B) **Termination of Time Marching.** Under the level-set implementation, even near reaching illusory equilibria, *individual* values of the level-set function can still keep changing. However,

the real role of the level-set function is only to (redundantly) represent the zero-level curve, or equivalently, its interior

$$(\phi > 0) := \{x \in \Omega \mid \phi(x) > 0\}.$$

Near the targeted equilibria, this set indeed stabilizes and converges to an illusory *shape*. Let $\phi^m(x)$ denote the level-set function at a discrete-time marching step m. Then we have employed the following stopping criterion:

$$\text{terminate the marching at } m+1 \text{ if} \quad \frac{\left|(\phi^{m+1} > 0) \ominus (\phi^m > 0)\right|}{\sqrt{|(\phi^m > 0)|}} \leq \epsilon,$$

where $A \ominus B$ of two sets A and B denotes their symmetric difference, and in the discrete numerical setting, $|A|$ counts the number of pixels in A. The tolerance level ϵ could be, for example, 1%. The square root operator in the denominator is motivated by the isoperimetric problem in geometry, in order to make the denominator roughly proportional to the length of the boundary contour.

(C) **Running Time.** All the above examples have been computed via MATLAB in an ordinary laptop computer with a processor ranging from 1.5 GHz to 2.0 GHz. We have employed the straightforward *global* level-set method [24], instead of the localized (but with more complex data management) fast level-set implementation. For a typical image of size $N = 160 \times 160$ with a moderate configuration as above, for example, the running time towards convergence is between 1 and 2 minutes.

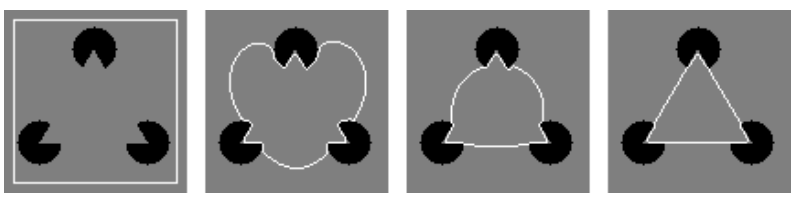


Figure 11: Kanizsa' illusory triangle as a stable local minimum to the proposed contour energy.

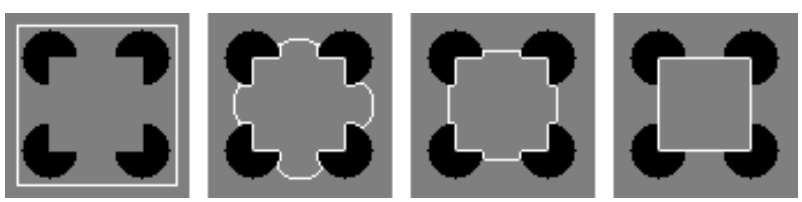


Figure 12: Kanizsa' illusory square as a stable local minimum and the stagnation process.

7 Conclusion

In the current paper, inspired by several recent works on the variational-PDE modeling for illusory contours, a simple mathematical model based on energy minimization has been proposed. The detailed and almost complete structures of the solutions are characterized, and an effective

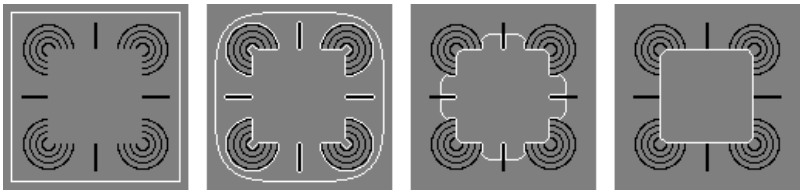


Figure 13: An illusory square emerges from a configuration of *real* contour bundles. The *object*-based supervised level-set computation works equally well for *contour*-based illusory contours due to the mollification operator applied to a given image u.

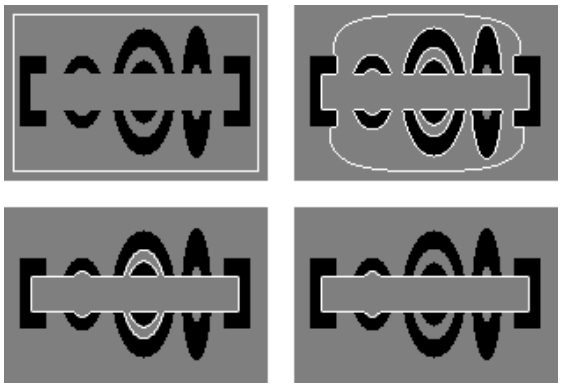


Figure 14: An illusory contour emerges from a complex configuration Q of objects, as captured by the supervised level-set-based active contours. The small hump in the final output (near the left end of the illusory bar) is not due to numerical defection but to the competition between the two weights α and β in the model.

computational algorithm has been proposed based on the asymptotic limit of the model as active contours. Generic numerical examples have been tested.

We must emphasize that the current work has been mainly driven by the spirit of rigorous mathematical modeling and analysis for vision research. By incorporating the vast literature on cognitive and neuronal illusory contours (e.g., [18, 34, 33, 25]), which the current work has not yet fully explored, the proposed model can still be substantially improved.

From the computational and modeling points of view, our future work will focus on further polishing and improving the model by employing other modern tools of image processing and analysis, such as the scale-space theory, the multiresolution framework, as well as the phase-field approach [7, 2, 35].

Acknowledgments

The authors would like to acknowledge the intellectual benefits from the Institute of Mathematics and its Applications (IMA) for her constant support to imaging and vision sciences. Shen also would like to thank Professors Alan Yuille in the Statistics Department at UCLA and Dan Kersten in the Psychology Department at the University of Minnesota for their inspiration and support to his research on cognitive vision.

Without the generous help from our three referees, the quality of this final version of the manuscript could not have been polished so much.

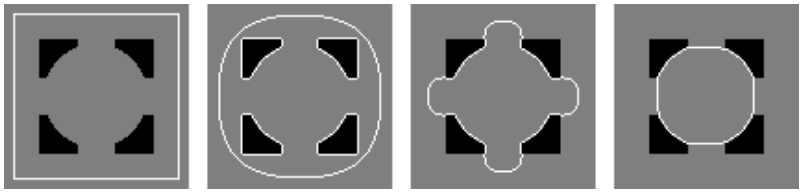


Figure 15: The consequence of first-order geometric modeling: our model can efficiently capture illusory objects and contours in their first-order proximity. To achieve high order fidelity, the model should incorporate at least the curvature information, as in the earlier works of Zhu and Chan [37, 38]. Such high order models are, however, often highly challenging for both rigorous theoretical analysis and efficient numerical computation. In-depth analysis as in Section 2 is often difficult to establish for high order models.

References

[1] E. H. Adelson. Lightness perception and lightness illusions. In M. Gazzaniga, editor, *The New Cognitive Neurosciences*, pages 339–351. MIT press, 2nd edition, 2000.

[2] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing*. Springer-Verlag, 2001.

[3] T. F. Chan, S.-H. Kang, and J. Shen. Euler's elastica and curvature based inpainting. *SIAM J. Appl. Math.*, 63(2):564–592, 2002.

[4] T. F. Chan, S. Osher, and J. Shen. The digital TV filter and nonlinear denoising. *IEEE Trans. Image Process.*, 10(2):231–241, 2001.

[5] T. F. Chan and J. Shen. Nontexture inpainting by curvature driven diffusions (CDD). *J. Visual Comm. Image Rep.*, 12(4):436–449, 2001.

[6] T. F. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math.*, 62(3):1019–1043, 2002.

[7] T. F. Chan and J. Shen. *Image Processing and Analysis: variational, PDE, wavelet, and stochastic methods*. SIAM Publisher, Philadelphia, 2005.

[8] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Process.*, 10(2):266–277, 2001.

[9] S. Esedoglu and J. Shen. Digital inpainting based on the Mumford-Shah-Euler image model. *European J. Appl. Math.*, 13:353–370, 2002.

[10] G. B. Folland. *Real Analysis - Modern Techniques and Their Applications*. John Wiley & Sons, Inc., second edition, 1999.

[11] W. T. Freeman. The generic viewpoint assumption in a framework for visual perception. *Nature*, 368:542–545, 1994.

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

[13] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *Int'l J. Comp. Vision*, 1(4):321–331, 1987.

[14] J. L. Kelley. *General Topology.* Springer Verlag, 1997.

[15] D. Kersten. High-level vision and statistical inference. In M. S. Gazzaniga, editor, *The New Cognitive Neurosciences*, pages 353–363. The MIT Press, Cambridge, MA, USA.

[16] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Conformal curvature flows: from phase transitions to active vision. *Arch. Rational Mech. Anal.*, 134(3):275301, 1996.

[17] D. C. Knill and D. Kersten. Apparent surface curvature affects lightness perception. *Nature*, 351:228–230, 1991.

[18] G. Mather. The role of subjective contours in capture of stereopsis. *Vision Research*, 29(1):143–146, 1989.

[19] D. Mumford. Elastica and computer vision. In C. L. Bajaj, editor, *Algebraic Geometry and its Applications*, pages 491–506. Springer-Verlag, New York, 1994.

[20] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Applied. Math.*, 42:577–685, 1989.

[21] M. Nitzberg, D. Mumford, and T. Shiota. *Filtering, Segmentation, and Depth.* Lecture Notes in Comp. Sci., Vol. 662. Springer-Verlag, Berlin, 1993.

[22] A. V. Oppenheim and A. S. Willsky. *Signals and Systems.* Prentice Hall, 1996.

[23] S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging, Vision and Graphics.* Springer Verlag, 2002.

[24] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys.*, 79(12):12–49, 1988.

[25] S. Petry and G. E. Meyer. *The Perception of Illusory Contours.* Springer-Verlag, New York, 1987.

[26] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[27] A. Sarti, R. Malladi, and J. A. Sethian. Subjective surfaces: A geometric model for boundary completion. *Int'l J. Comp. Vision*, 46(3):201–221, 2002.

[28] J. A. Sethian. A marching level set method for monotonically advancing fronts. *Proc. Nat. Acad. Sci.*, 93(4):1591–1595, 1996.

[29] J. A. Sethian. *Level Set Methods and Fast Marching Methods.* Cambridge University Press, 2nd edition, 1999.

[30] J. Shen. On the foundations of vision modeling I. Weber's law and Weberized TV (total variation) restoration. *Physica D: Nonlinear Phenomena*, 175:241–251, 2003.

[31] J. Shen. Bayesian video dejittering by BV image model. *SIAM J. Appl. Math.*, 64(5):1691–1708, 2004.

[32] J. Shen. On the foundations of vision modeling III. Noncommutative monoids of occlusive preimages. *J. Math. Imag. Vision*, 24:5–17, 2006.

[33] R. von der Heydt and E. Peterhans. Mechanism of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *J. Neurosci.*, 9:1731–1748, 1989.

[34] R. von der Heydt, E. Peterhans, and G. Baumgartner. Illusory contours and cortical neuron responses. *Science*, 224:1260–1262, 1984.

[35] J. Weickert. *Anisotropic Diffusion in Image Processing.* Teubner-Verlag, Stuttgart, Germany, 1998.

[36] L. R. Williams and D. W. Jacobs. Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation*, 9:837–858, 1997.

[37] W. Zhu and T. Chan. Capture illusory contours: A level set based approach. *UCLA CAM Tech. Report*, 03-65, 2003.

[38] W. Zhu and T. Chan. Illusory contours using shape information. *UCLA CAM Tech. Report*, 03-09, 2005.