Minimal surfaces in the roto-translation group with applications to a neuro-biological image completion model

Robert K. Hladky
Department of Mathematics, Dartmouth College, 6188 Bradley Hall Hanover, NH 03755 USA
robert.k.hladky@dartmouth.edu

Scott D. Pauls
Department of Mathematics, Dartmouth College, 6188 Bradley Hall Hanover, NH 03755 USA
scott.d.pauls@dartmouth.edu

We investigate solutions to the minimal surface problem with Dirichlet boundary conditions in the roto-translation group equipped with a subRiemannian metric. By work of G. Citti and A. Sarti, such solutions are amodal completions of occluded visual data when using a model of the first layer of the visual cortex. Using a characterization of smooth minimal surfaces as ruled surfaces, we give a method to compute a minimal spanning surface given fixed boundary data presuming such a surface exists. Moreover, we describe a number of obstructions to existence and uniqueness but also show that under suitable conditions, smooth minimal spanning surfaces with good properties exist. Not only does this provide an explicit realization of the disocclusion process for the neurobiological model, but it also has application to contructing disocclusion algorithms in digital image processing.

Keywords: Minimal surfaces, sub-Riemannian spaces, vision, occlusion

AMS Subject Classification: 53C17,49Q05,92B05

1. Introduction

Starting with the initial work of Hubel and Weisel[19, 20], the past half century has seen deep investigations into the neurological mechanisms in the primary visual cortex (V1), leading to a variety of models of aspects of the function of this area. In this paper, we focus on a model of the function of the long range interaction between the so-called simple cells. Using a now standard modeling of these cells by Gabor filters, Citti and Sarti[7] built on earlier models of Hoffman[18] and Petitot and Tondut[29] to create a model of the simple cell function in V1 via a sub-Riemannian space. This model has several compelling aspects including the incorporation of the Gabor model of simple cells and the derivation of the association fields of Field, Hess and Hayes[11] from the geometric properties of the model. In addition, Citti and Sarti applied their model to the problem of image completion, showing that model performs completion (either modal or amodal) via a solution of the sub-Riemannian

minimal surface problem.

The study of minimal surfaces in sub-Riemannian spaces has recently received a good deal of attention.[4, 5, 8, 9, 12, 13, 17, 22, 24, 26, 27, 30] In the work cited, the various authors explore the existence, uniqueness and properties of minimal surface problem, finding that, at least in specific lower dimensional cases, the minimal surfaces have a rich geometric structure. To make these notions more precise, we focus on the sub-Riemannian space that arises in the Citti-Sarti model, the roto-translation group, $\mathcal{RT}$. This space is homeomorphic to $\mathbb{R}^2 \times \mathbb{S}^1$ and we will use (x, y, θ) as coordinates. We define a sub-Riemannian structure on $\mathcal{RT}$ by distinguishing a horizontal subbundle, $\mathcal{H}$, given by the span of the following two vector fields at each point:

$$X_1 = \cos(\theta)\,\frac{\partial}{\partial x} + \sin(\theta)\,\frac{\partial}{\partial y},\;\; X_2 = \frac{\partial}{\partial \theta}.$$

These two vector fields bracket generate the tangent bundle and form a distribution of contact planes in this three dimensional space. Placing an inner product on $\mathcal{H}$ which makes $\{X_1, X_2\}$ an orthonormal basis for $\mathcal{H}$, we have the standard Carnot-Carathéodory distance on $\mathcal{RT}$:

$$d_{cc}(a, b) = \inf_{\gamma \in \mathscr{A}} \left\{ \int < \gamma', \gamma' >^{\frac{1}{2}} \,\middle|\, \gamma(0) = a, \gamma(1) = b \right\}$$

where $\mathscr{A}$ is the set of absolutely continuous paths whose derivatives, when they exist, are in $\mathcal{H}$.

In the Citti-Sarti model, a greyscale image, $I : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$, has a representation in $\mathcal{RT}$ given by

$$\Sigma = \left\{ (x, y, \theta) \middle| \theta = \arctan\left(-\frac{I_x}{I_y}\right) \right\} \tag{1.1}$$

If a portion of the image is occluded in a domain $\Omega_0 \subset \Omega$, then a completion of the occluded region is constructed by forming a minimal spanning surface. More precisely, if $c \subset \Sigma$ is the curve in $\mathcal{RT}$ associated to $\partial\Omega_0$, then the completion is given by finding the minimal surface in $\mathcal{RT}$ that spans c, i.e. a minimizer of the perimeter measure. For a C^1 surface, Σ, given as a level set of a C^1 function f, the perimeter is given by

$$\mathscr{P}(\Sigma) = \int \sqrt{(X_1 f)^2 + (X_2 f)^2}\, dA \tag{1.2}$$

Moreover we know from the work of either Citti and Sarti or Cheng, Hwang, Malchiodi and Yang[5], that such minimal surfaces satisfy the following partial differential equation:

$$X_1(\nu_1) + X_2(\nu_2) = 0.$$

where ν is the unit horizontal normal to Σ and is given by

$$\nu = \nu_1 X_1 + \nu_2 X_2 = \frac{X_1 u}{\sqrt{(X_1 u)^2 + (X_2 u)^2}}\, X_1 + \frac{X_2 u}{\sqrt{(X_1 u)^2 + (X_2 u)^2}} X_2$$

This differential equation is sometimes written as div $\nu = 0$. In addition to this relationship between the minimal surface problem and a model for biological image reconstruction, Citti and Sarti provide a reinterpretation of a number of existing algorithms for digital inpainting and image completion. In particular, Citti and Sarti[7], and Citti, Manfredini and Sarti[6], examine the variational models of Ambrosio-Masnou[1], elastica models and a variant of the Mumford-Shah functional and find that, under suitable interpretation in the roto-translation group model, minimizing these different functionals is equivalent to solving a geometric problem such as the geodesic problem (in the case of elastica) or the minimal surface problem described above (in the Ambrosio-Masnou model).

In light of this unifying theme in the area of vision and image reconstruction, we explore the minimal surface problem in a class of groups which include the roto-translation group, $\mathcal{RT}$. We note that a number of authors have shown that the smooth minimal surfaces in this setting are ruled surfaces (see Citti-Sarti[7], Cheng et al[5] and, for a more general case, see additional work by the authors[17]).

In Section 4, we use the basic form of the minimal surface equation to explicitly derive the curves that rule smooth minimal surfaces in the roto-translation group. Specifically, we show the following Theorem:

Theorem 1.1. *Suppose S is a non-characteristic surface patch in the roto-translation group. The horizontal minimal surface equation for S*

$$div\ \nu = 0 \tag{1.3}$$

can be written as

$$\nabla_{J\nu} J\nu = 0$$

where ν is the horizontal unit normal. Thus if S satisfies (1.3) *then S is ruled by horizontal ∇-geodesics.*

Here, ∇ denotes the Webster-Tanaka connection on the roto-translation group and $J(\nu) = \nu_2 X_1 - \nu_1 X_2$. We note that below this result is proved for a more general class of spaces, see Theorem 4.2.

Further, we can explicitly calculate such geodesics:

Theorem 1.2. *The horizontal ∇-geodesics in $\mathcal{RT}$ are determined by a choice of basepoint $(x_0, y_0, \theta_0) \in \mathcal{RT}$, a radius R_0, and velocity $\dot{\theta} \in \mathbb{R}$. The ∇-geodesic, $\gamma = (x, y, \theta)$ given by a choice $\{(x_0, y_0, \theta_0), R_0, \dot{\theta}\}$ is*

- *If $\dot{\theta} \neq 0$, $R_0 \neq 0$, then*

$$\begin{aligned} x &= x_c + R\sin\theta \\ y &= y_c - R\cos\theta \\ \theta &= \theta_0 + \dot{\theta}t. \end{aligned}$$

where $R = R_0/\dot{\theta}, x_c = x_0 - R\sin\theta_0$ and $y_c = y_0 + R\cos\theta_0$.

- *If* $\dot{\theta} = 0$, $R_0 \neq 0$, *then*

$$\begin{aligned} x &= x_0 + R_0(\cos\theta_0)t \\ y &= y_0 + R_0(\sin\theta_0)t \\ \theta &= \theta_0. \end{aligned}$$

- *If* $\dot{\theta} \neq 0$, $R_0 = 0$, *then*

$$\begin{aligned} x &= x_0 \\ y &= y_0 \\ \theta &= \theta_0 + \dot{\theta}t. \end{aligned}$$

Qualitatively, this says that smooth minimal surfaces are foliated by circles of varying radii, where we also consider lines to be circles of infinite radii. We note that, under the interpretation of minimal surfaces image completions formed by providing associations between distant areas of V1, this qualitative description matches with notions of *cocircularity* found in the analysis of visual function of the Gestalt school[34, 35], digital image analysis[16] and neuron responses in V1[36].

With this description of smooth minimal surfaces in hand, we turn to an analysis of the so-called *disocclusion problem*, the problem of completing images with occluded or missing pieces with dual goals of understanding the neurobiological process via the Citti-Sarti model and of application to digital image processing. To fix notation, we make the following definitions:

Definition 1.1. Let D be an open subset of $\mathbb{R}^2$ with C^1 boundary. For an image intensity function $I : \overline{D} \subset \mathbb{R}^2 \to \mathbb{R}$, its representation, $\Sigma(I)$, in $\mathcal{RT}$ is given by $(x, y, \theta(x, y))$ where $(x, y) \in \overline{D}$ and θ is given by

$$\frac{\nabla I}{|\nabla I|} = (-\sin(\theta), \cos(\theta))$$

Definition 1.2. Let $D \subset \mathbb{R}^2$ be an open domain with C^1 boundary. Suppose that $I : \mathbb{R}^2 \to \mathbb{R}$ is an image function consider D as a domain of missing or occluded data. We call D an *occlusion domain*.

Let D be a bounded occlusion domain with boundary given by a C^1 simple closed curve, $I : \mathbb{R}^2 \to \mathbb{R}$ be a C^2 image function, β be a parameterization of the boundary by arclength oriented counterclockwise, and $\gamma(t) = (\beta(t), \theta(t)) \subset \mathcal{RT}$. The collection (D, I, γ) is called the *occlusion data* associated to the occlusion problem given by (D, I).

Let Σ be the surface in $\mathcal{RT}$ associated to $I|_{\mathbb{R}^2 \setminus D}$ via Definition 1.1. We say that $\Sigma' \subset \mathcal{RT}$ is a *minimal spanning surface* of Σ if Σ' is a solution to the minimal surface equation and $\partial\Sigma' = \partial\Sigma$.

We note that the function θ is determined by the Citti-Sarti model. Given a point $p = (x_0, y_0, \theta_0) \in (\beta, \theta)$, we use Theorem 1.1 to determine all the points

that are accessible to p via the circular arcs described in Theorem 1.2. Using some trigonometric simplifications, we find the following characterization:

Proposition 1.3. *Given a point $p = (x_0, y_0, \theta_0)$, the set of accessible points, $\{(x, y, \theta)\}$ is given by the implicit equation*

$$(x - x_0, y - y_0) \cdot \left(-\sin\left(\frac{\theta + \theta_0}{2}\right), \cos\left(\frac{\theta + \theta_0}{2}\right)\right) = 0 \tag{1.4}$$

This is an explicit realization of the cocircularity condition described above. We say that a function $u : \mathbb{S}^1 \to \mathbb{S}^1$ is a solution for a set of occlusion data (D, I, γ) if for each $t \in \mathbb{S}^1$, $\gamma(u(t))$ is accessible from $\gamma(t)$.

Our first main result is that there are obstructions to the existence of smooth minimal spanning graphs. The obstruction stems from the existence of *solitary points*, those points, p, whose accessible set consists precisely of p itself.

Theorem 1.3. *Let (D, I, γ) be a set of occlusion data. If $\gamma = (\beta, \theta) \subset \mathcal{RT}$ contains non-Legendrian solitary points, then there does not exist a C^2 minimal spanning surface of γ which is a graph over D.*

Here, a nonLegendrian point of γ is a point where $\gamma' \notin span\{X_1, X_2\}$.

In the somewhat the same direction, we investigate smooth solutions that produce ambiguous image completions when translated back to image functions on D. Specifically, using a solution $u : \mathbb{S}^1 \to \mathbb{S}^1$ to the occlusion data $(D, I, \gamma = (\beta, \theta))$, we construct an image disocclusion via a map $\Pi : \mathbb{S}^1 \times [0, 1] \to \overline{D}$. To properly define Π we need an auxillary function:

$$\vartheta(t, u) = \begin{cases} R(t)(\theta(u(t)) - \theta(t)) \cos(Q(t)) & \text{if } R(t) < \infty \\ (-\dot\beta_2(t), \dot\beta_1(t)) \cdot (\beta(u(t)) - \beta(t)) & \text{if } R(t) = \infty \end{cases} \tag{1.5}$$

Here $R(t)$ is the radius of the rule connecting $\gamma(t)$ to $\gamma(u(t))$ (we let $R(t) = \infty$ if the rule is a straight line) and $Q(t)$ is a *transversality function* defined by $Q(t) = \theta(t) - n_\beta(t)$ where $\dot\beta(t) = (-\sin(n_\beta(t)), \cos(n_\beta(t)))$.

We define the function $\varphi(r)$ by

$$\varphi(r) = \begin{cases} (\theta(u(t)) - \theta(t))r + \theta(t) & \text{if } \vartheta(t, u) \le 0 \\ (\theta(t) - \theta(u(t)))r + \theta(u(t)) & \text{if } \vartheta(t, u) > 0. \end{cases} \tag{1.6}$$

Using these auxillary functions, we can define Π:

Definition 1.4. Then, for t so that $0 < |R(t)| < \infty$, we define $\Pi : \mathbb{S}^1 \times [0, 1] \to \mathbb{R}^2$ by

$$\Pi(t, r) = Rule(t, u(t))(r)$$

where

$$Rule(t, u(t))(r) = \Big(\beta_1(t) + R(t)\big(\sin(\varphi(r)) - \sin(\theta(t))\big), \\ \beta_2(t) - R(t)\big(\cos(\varphi(r)) - \cos(\theta(t))\big)\Big)$$

For points where $R(t) = \infty$, we define

$$Rule(t, u(t))(r) = \begin{cases} (1-r)\beta(t) + r\beta(u(t)) & \text{if } \vartheta(t,u) < 0 \\ (1-r)\beta(u(t)) + r\beta(t) & \text{if } \vartheta(t,u) > 0 \end{cases} \tag{1.7}$$

There are several ways that Π can produce ambiguous or incomplete image functions. Among these are:

(a) The rules may exhibit "backtracking", when $\dot{u} > 0$. This results in a loss of injectivity for Π.
(b) Π may cease to be a local diffeomorphism, also resulting in a loss of injectivity.
(c) Π may not be surjective

In Section 6, we provide a number of calculations that allow us to test for these conditions pointwise. In other words, given a points $p, q \in (\beta, \gamma)$ where q a point accessible to p, we can determine if the solution backtracks, is a local diffeomorphism along the curve connecting the two points, or is locally surjective based only on image data at the two points. This allows us, in Section 7, to outline an algorithm for image disocclusion based on this model.

In Section 8, we restrict to occlusions with circular boundaries and implement the image disocclusion algorithm. We do this for computational reasons as the restriction on the region and boundary allow for significant simplification. We emphasize that a numerical implementation of the general algorithm is not difficult. In the case of circular occlusions, we find characterizations of several cases. These characterizations use the *transversality function* described above which, in the case of a circular boundary, takes the form:

$$Q(t) = \theta(t) - t$$

First, with regard to the obstruction described above, we have:

Theorem 1.4. *Let $(D, I, \gamma = (\beta, \theta))$ be a set of occlusion data with β a circle. Then a necessary condition for the existence of non-Legendrian solitary points is that deg $Q = 0$.*

Moreover, we have an existence theorem:

Theorem 1.5. *Let (D, I, γ) be a set of occlusion data with β a circle. If deg $Q = d \neq 0$ then there are at least $|d|$ solutions for the occlusion data.*

Last, we describe a case where we guarentee the existence of a "good" solution, i.e. one that produces an effective image disocclusion:

Theorem 1.6. *Let $I : \mathbb{R}^2 \to \mathbb{R}$ be a C^2 function and D be a circular occlusion domain. Further, suppose $\gamma \in \mathcal{RT}$ is the θ lift of ∂D and that the occlusion is completely nondegenerate and occludes no critical points of I. If $Q'(t) \neq 0$ for $t \in [0, 2\pi)$ then there exists a minimal spanning surface of γ which maps to D under the map Π and the image of Π lies entirely inside D. Further, an image disocclusion*

derived from the minimal surface exhibits no backtracking. If, in addition, we assume that $Q' < 0$*, then* Π *maps onto* D*.*

An occlusion is completely nondegenerate if if there are a finite number of critical points of I in D and no critical points on the boundary of D.

We end the paper with some explicit examples illustrating the various cases and examples from an implementation of the digital image disocclusion algorithm for circular occlusions.

2. Modeling V1 via the roto-translation group

In this section, we review the basic biological findings describing the function of V1 and describes a mathematical model of V1. Moreover, we describe the connection, provided by Citti and Sarti[7], between minimal surfaces in the model space and solutions to the problem of completing regions of occluded image data.

Over the past several decades, the function and operation of the first layer of the visual cortex, V1, has become increasingly clear. Early research showed that V1 contains so-called simple cells that are sensitive to, among other things, brightness gradients with a particular orientation. These cells are arranged in columns sharing the same orientation preference[19, 20] and the columns are arranged in hypercolumns which represent all possible orientations. This view was further explored and modeled mathematically,[18, 29] where the authors modeled the hypercolumnar cell structure using a contact manifold. The contact model is based on a simplifying assumption that treats each column as a point, ignoring the column structure to focus on the hypercolumn structure. Mathematically, they use the manifold $\mathbb{R}^2 \times \mathbb{S}^1$ to model the hypercolumn structure by placing a circle of directions above each point $(x, y) \in \mathbb{R}^2$. Each point (x, y, θ) represents a column of cells associated to an (x, y) point of retinal data, all of which are attuned to the orientation give by the angle θ. See figure 1 for a schematic of the hypercolumnar structure.

Early assumptions that cortical connectivity should run mostly vertically along the hypercolumns and be severely restricted in horizontal directions, while supported by some research, was contradicted by later evidence which showed that there is "long range horizontal" connectivity in the cortex. These experiments (see for example, Gilbert et al[14]) indicated that horizontal connections are made between cells in different hypercolumns of similar orientation preference. In addition, Field, Hess and Hayes[11]) demonstrated the existence of the so-called "association fields" showing that line segments scattered in the visual field are perceived to have stronger conenctions if they trace a smooth curve which does not vary too much, pointing towards the existence of some similar neural mechanism in the visual cortex. This evidence points towards a geometric structure in this layer where communication between adjacent cells is allowable in certain directions, vertically and between cells in different hypercolumns of similar orientation sensitivity, and vastly restricted in all other directions. We note that this type of situation has been studied in a variety of settings including, for example, control theoretic problems

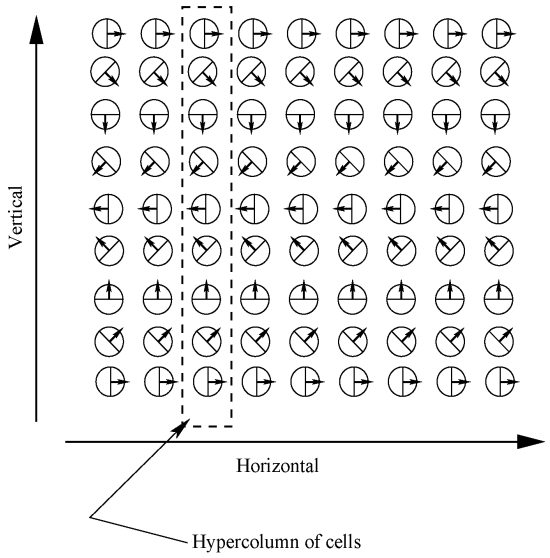


Fig. 1. A schematic of the structure of V1

where the degrees of freedom at a particular point are restricted.

Petitot and Tondut[28, 29] incorporate these biological findings into their model by introducing a contact structure on $\mathbb{R}^2 \times \mathbb{S}^1$ via the one form $\omega = dx - \theta dy$ and introduce a sub-Riemannian metric associated to the contact two-plane distribution to encode the geometry of the model of V1. The plane field given by the kernel of ω, $span\{\partial_\theta, \partial_y + \theta\partial_x\}$, corresponds to the space of allowable directions at each point. Notice that these are precisely the vertical direction and the direction which links cells in different hypercolumns with the same θ value. Citti and Sarti[7] use the following explicit realization of the roto-translation group, $\mathcal{RT}$:

- $\mathcal{RT}$ is diffeomorphic to $\mathbb{R}^2 \times \mathbb{S}^1$ with coordinates (x, y, θ).
- The following three vector fields span the tangent space at each point:

$$\begin{aligned} X_1 &= \cos(\theta)\,\frac{\partial}{\partial x} + \sin(\theta)\,\frac{\partial}{\partial y} \\ X_2 &= \frac{\partial}{\partial \theta} \\ X_3 &= -\sin(\theta)\,\frac{\partial}{\partial x} + \cos(\theta)\,\frac{\partial}{\partial y} \end{aligned} \tag{2.1}$$

We note that $[X_2, X_1] = X_3$ and $[X_2, X_3] = -X_1$.

Definition 2.1. Let D be an open subset of $\mathbb{R}^2$ with C^1 boundary. For an image intensity function $I : \overline{D} \subset \mathbb{R}^2 \to \mathbb{R}$, its representation, $\Sigma(I)$, in $\mathcal{RT}$ is given by

$(x, y, \theta(x, y))$ where $(x, y) \in \overline{D}$ and θ is given by

$$\frac{\nabla I}{|\nabla I|} = (-\sin(\theta), \cos(\theta))$$

We note that this is an explicit realization of the model described above and matches with the biological evidence concerning horizontal connectivity. Moreover, their model provides an explanation for the experimental results of Field, Hess and Hayes[11] concerning the so-called "association fields" used in visual contour integration, further validating the model. The contact subbundle in this presentation is simply $span\{X_1, X_2\}$. It is a direct calculation that this subbundle gives a contact structure and, placing an inner product on this subbundle making $\{X_1, X_2\}$ orthonormal, we have a standard sub-Riemannian metric on $\mathcal{RT}$ (see the next section for a precise definition). One of the main contributions of Citti and Sarti's adaptation of the cortical model is the use of an explicit lifting function that transforms retinal data into a surface in the cortex and allows direct use of the sub-Riemannian structure.

To explore this further, we investigate the application of the model to an image. Using the representation above, we see that the direction given by the angle $(\cos(\theta), \sin(\theta))$ points in a direction perpendicular to the gradient of I, i.e. a direction tangent to the level sets of I. Thus, if nearby points have the same intensity and thereby lie on the same level set of I, their θ representations will be similar. As θ denotes a position in a hypercolumn of cells over the point (x, y), this echoes the biological finding horizontal communication occurs between cells of similar orientation specificity and the property that the representation respects level lines reflects the biological principle encoded in the association fields of Field, Hess and Hayes. In figure 2, we give a schematic of lifting a simple image to $\mathcal{RT}$. Represented are two layers of cells of similar orientation preference but different hypercolumns and two sections of the image that could plausibly be lifted to those layers.

Citti and Sarti[7] use the model of V1 to investigate the problem of filling in an image when a portion of the image is missing due to occlusion or some other factor. To use the model, Citti and Sarti took image data with a portion deleted and used an iteration of a geometric flow and non-maximal suppression to evolve surfaces in $\mathcal{RT}$ and fill in holes. They prove that this mechanism, applied to a surface associated to an image, converges to an infilling given by a solution to the minimal surface equation in $\mathcal{RT}$ satisfying Dirichlet boundary conditions. To state this precisely, we need to make a number of definitions and more precise notation.

3. Notation

In this section, we fix the basic notation used throughout the paper. Let $\mathcal{G}$ be a topologically three dimensional one step graded Lie group. In other words, the Lie algebra of left invariant vector fields $\mathcal{V}$ splits as

$$\mathcal{V} = \mathcal{V}_0 \oplus \mathcal{V}_1, \;\; \dim \mathcal{V}_0 = 2, \;\; \mathcal{V}_1 = [\mathcal{V}_0, \mathcal{V}_0]$$

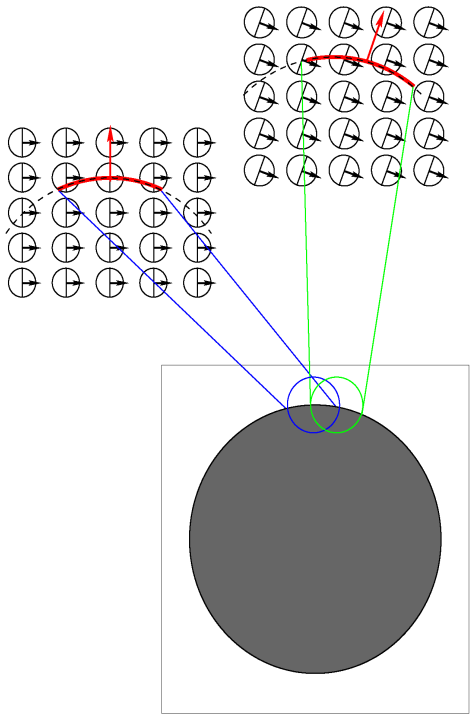


Fig. 2. Example of lifting an image to $\mathcal{RT}$

Moreover, we assume the following

- $\mathcal{G}$ is equipped with a Riemannian metric g, which we sometimes denote in inner product notation by $\langle \cdot, \cdot \rangle$ and which makes the grading orthogonal.
- ∇^{LC} is the Levi-Civita connection associated to g
- $\mathcal{G}$ satisfies the assumption

$$[\mathcal{V}_0, \mathcal{V}_1] \subset \mathcal{V}_0$$

- We define a Carnot-Carathéodory distance on $\mathcal{G}$ by

$$d_{cc}(x,y) = \inf_{\gamma \in \mathcal{A}} \left\{ \int < \gamma', \gamma' >^{\frac{1}{2}} \middle| \gamma(0) = x, \gamma(1) = y \right\}$$

  where $\mathcal{A}$ is the set of all absolutely continuous paths that, where their derivatives are defined, have $\gamma' \in \mathcal{V}$.

Definition 3.1. If $\mathcal{G}$ satifies all of the assumptions above, we say that $\mathcal{G}$ is an admissible sub-Riemannian group.

We note that if an admissible $\mathcal{G}$ is nilpotent, then G is known as a *Carnot group*. We review two special examples.

Example 3.2. *The Heisenberg group.* The topologically three dimensional Heisenberg group, $\mathbb{H}$, is one of the simplest nonabelian nilpotent Lie groups. As a smooth manifold, it is diffeomorphic to $\mathbb{R}^3$. Using the terminology above, we have the Lie algebra given as

$$\mathfrak{h} = \mathcal{V}_0 \oplus \mathcal{V}_1$$

with $\mathcal{V}_0 = span\{X_1, X_2\}$ and $\mathcal{V}_1 = span\{X_3\}$ where there is a single nontrivial bracket operation, $[X_1, X_2] = X_3$. We note that, as $\mathbb{H}$ is nilpotent, it is a Carnot group.

Example 3.3. *The Roto-Translation group.* As the roto-translation group, $\mathcal{RT}$, appears centrally in the model of visual processing in V1, we review its abstract structure. As a smooth manifold, it is diffeomorphic to $\mathbb{R}^2 \times \mathbb{S}^1$ and, using the terminology above, we have the Lie algebra given as

$$\mathfrak{h} = \mathcal{V}_0 \oplus \mathcal{V}_1$$

with $\mathcal{V}_0 = span\{Y_1, Y_2\}$ and $\mathcal{V}_1 = span\{Y_3\}$ where we have the following nontrivial bracket operations, $[Y_1, Y_2] = Y_3, [Y_2, Y_3] = Y_2$. Equation (2.1) in section 2 give a particular presentation of $\mathcal{RT}$. We note that $\mathcal{RT}$ is not a Carnot group as it is not nilpotent.

Given a submanifold $S \subset \mathcal{G}$, the sub-Riemannian geometry of S is determined by the *horizontal normal* to S, N_0, which is simply the projection of the Riemannian normal, N, to $\mathcal{V}_0$, the first layer of the grading. Explicitly, if $\{X_1, X_2, X_3\}$ is a left invariant orthonormal basis of $T\mathcal{G}$ with $\mathcal{V}_0 = span\{X_1, X_2\}$ and S is given as a level set $\varphi = 0$ then we define

Definition 3.4. The horizontal normal to $S \subset \mathcal{G}$ is

$$\begin{aligned} N_0 &= \mathrm{proj}_{\mathcal{V}_0} N \\ &= \mathrm{proj}_{\mathcal{V}_0}((X_1\varphi)\, X_1 + (X_2\,\varphi) X_2 + (X_3\varphi)\, X_3) = (X_1\varphi)\, X_1 + (X_2\varphi)\, X_2 \end{aligned}$$

A point $x \in S$ is said to be characteristic if $N_0(x) = 0$. At noncharacteristic points, we also define the *unit horizontal normal*:

$$\nu = \frac{N_0}{< N_0, N_0 >^{\frac{1}{2}}}$$

Minimal surfaces in Carnot-Carathéodory spaces have been investigated in a number of settings[4, 5, 9, 12, 13, 17, 26, 27]. In particular, Cheng, Hwang, Malchiodi and Yang[5] prove the following Theorem:

Theorem 3.5. *A noncharacteristic C^2 hypersurface in an admissible sub-Riemannian group $\mathcal{G}$ is a critical point of the first variation of area if and only if it satisfies the minimal surface equation:*

$$div\ \nu = X_1\nu_1 + X_2\nu_2 = 0 \tag{3.1}$$

In the literature, the operator $(v_1, v_2) \to X_1 v_1 + X_2 v_2$ is called the horizontal divergence operator. However for admissible groups, or more generally vertically rigid manifolds[17], it is easy to see that this is equivalent to the Riemannian divergence. With this notation in place, we can now precisely describe the use of the roto-translation model in the disocclusion problem.

We will consider images given as functions on the plane, $I : \mathbb{R}^2 \to \mathbb{R}$, where a specified region $D \subset \mathbb{R}^2$ is considered as an occlusion of the image. Let Σ be the surface associated to $I|_{\mathbb{R}^2 \setminus D}$ via Definition 2.1. We first note that Σ is completely noncharacteristic as the surface can be written as $\theta - \theta(x, y) = 0$ and

$$N_0 = -(\cos(\theta(x, y))\theta_x(x, y) + \sin(\theta(x, y))\theta_y(x, y))\ X_1 + X_2$$

Using $\mathcal{RT}$ to model the completion process via an infinite iteration of diffusion and suppression processes, Citti and Sarti prove[7] that, in the limit, the resulting filled surface, Σ', satisfies (3.1). To formalize this, we make a defintion of disocclusion:

Definition 3.6. Let $D \subset \mathbb{R}^2$ be an open domain with C^1 boundary. Suppose that $I : \mathbb{R}^2 \to \mathbb{R}$ is an image function consider D as a domain of missing or occluded data. We call D an *occlusion domain*. Let Σ be the surface in $\mathcal{RT}$ associated to $I|_{\mathbb{R}^2 \setminus D}$ via Definition 2.1. We say that $\Sigma' \subset \mathcal{RT}$ is a *minimal spanning surface* of Σ if Σ' is a solution to (3.1) and $\partial\Sigma' = \partial\Sigma$.

Further,

Definition 3.7. Let $D \subset \mathbb{R}^2$ be an open domain with C^1 boundary and $I : \mathbb{R}^2 \to \mathbb{R}$ be an image intensity function. D is called a *non-degenerate* occlusion if the number of critical points of I lying inside D is finite and θ, as defined in Definition 2.1, can be extended continuously across each critical point. D is called a *completely non-degenerate* occlusion if it is a nondegenerate occlusion and if there are no critical points on the boundary.

The main goal of this paper is to provide a description of such minimal surfaces and to describe method by which they can be constructed in particularly a simple manner.

4. Minimal surfaces in $\mathcal{RT}$

Definition 4.1. A (3 dimensional) pseudohermitian manifold is a triple (M, η, J) consisting of a 3 dimensional manifold M, a non-vanishing 1-form η on M and a smooth bundle map $J\colon H = \mathrm{Ker}(\eta) \to H$ such that $J^2 = -1$. (In higher dimensions there is in addition an integrability condition, see Tanaka[32].)

The structure is said to be *strictly pseudoconvex* if the following bilinear form on TM

$$L(X, Y) = d\eta(X, JY) + \eta(X)\eta(Y)$$

is positive definite. When this is so, the bilinear form is known as the Levi metric. For a strictly pseudoconvex structure, the characteristic vector field is the unique vector field T such that $\eta(T) = 1$ and $T \lrcorner d\eta = 0$.

Lemma 4.1. *If $\mathcal{G}$ is an admissible sub-Riemannian group then there exists a global strictly pseudoconvex pseudohermitian structure $(\eta, J, \mathcal{G})$ with the following properties*

- *span $\mathcal{V}_0 = H$, the contact distribution for η.*
- *$V =$ span $\mathcal{V}_1$ is spanned by the characteristic vector field T for η.*
- *The Levi metric agrees with g on H and is conformal with constant scaling factor to g on V.*

Proof: Let X_1 and X_2 be a left invariant orthonormal frame for $\mathcal{V}_0$ and set $T = [X_1, X_2] \in \mathcal{V}_1$. By left invariance and the bracket generating property of $\mathcal{V}_0$, we see that X_1, X_2, T form a global orthonormal frame for $T\mathcal{G}$. Set η to be the dual 1-form for T with respect to this frame. Then clearly span$\mathcal{V}_0 =$ Ker η and strict pseudoconvexity is immediate as Ker(η) bracket generates at 1-step. The additional bracket condition $[\mathcal{V}_0, \mathcal{V}_1] \subset \mathcal{V}_0$ then implies that $T \lrcorner d\eta = 0$. Thus the first two properties hold automatically regardless of which complex structure J is chosen for H.

Define $J : \mathcal{V}_0 \to \mathcal{V}_0$ by $JX_1 = -X_2$ and $JX_2 = X_1$. Next we extend J to all of $T\mathcal{G}$ by setting $JT = 0$ and declaring J to be linear over $\mathbb{R}$. The Levi metric defined by

$$h(X, Y) = d\eta(X, JY) + \eta(X)\eta(Y) \tag{4.1}$$

then clearly is compatible with g in the required fashion.

■

In this setting, we can now employ the techniques of pseudohermitian and CR geometry. Our key tool is the existence of a canonical connection ∇, derived independently by Webster[33] and Tanaka[32], for any strictly pseudoconvex pseudohermitian geometry. The defining properties of the connection are as follows:

- H, T, η, $d\eta$ and J are all parallel.
- $\mathrm{Tor}(X, Y) = d\eta(X, Y)T$ for $X, Y \in H$.
- $\mathrm{Tor}(JX, T) = J\mathrm{Tor}(X, T) \in H$ for $X \in H$.

Computations using this connection are most easily conducted in the moving frame approach of Cartan, adapted to this setting by Webster[33]. For this technique, we first complexify the contact distribution H and define the space of $(1, 0)$ and $(0, 1)$ vector fields to be the $+i$ and $-i$ eigenspace of J respectively. The $(1, 0)$ vector fields are then spanned by

$$Z = X_2 - iX_1.$$

The vector fields $Z, \overline{Z}$ and T then form an orthonormal (complex) frame for the complexified tangent space. The dual frame will be denoted by $\zeta, \overline{\zeta}$ and η. We

introduce the connection form ω via the identity

$$\nabla Z = \omega \otimes Z \tag{4.2}$$

With respect to our frame the Webster-Tanaka connection and Levi metric can be uniquely computed from the following equations[33]:

$$d\eta = ih\zeta \wedge \overline{\zeta} \tag{4.3}$$
$$dh = \omega h + h\overline{\omega} \tag{4.4}$$
$$d\zeta = \zeta \wedge \omega + \eta \wedge \tau \tag{4.5}$$
$$\tau = 0 \bmod \overline{\zeta} \tag{4.6}$$

The 1-form τ is known as the torsion form.

This connection proves well adapted to many geometric problems. For the study of horizontally minimal surfaces we have the following theorem.

Theorem 4.2. *Suppose S is a non-characteristic surface patch in an admissible group satisfying our structure conditions. The horizontal minimal surface equation for S*

$$div\ \nu = 0 \tag{4.7}$$

can be written as

$$\nabla_{J\nu} J\nu = 0$$

where ν is the horizontal unit normal. Thus if S satisfies (4.7) *then S is ruled by horizontal ∇-geodesics.*

Proof: The volume form for the Levi metric is given by $dV = \eta \wedge d\eta$. By the defining properties of the pseudohermitian structure this is a constant multiple of the Riemannian volume form. From this we immediately see that dV is parallel for ∇ and the divergence operator for dV agrees with the Riemannian divergence. Further it follows that $T \lrcorner \eta \wedge d\eta = d\eta$ and so

$$\text{div } T = 0.$$

Now ν, $J\nu$ and T form a local orthonormal frame for $T\mathcal{G}$. A standard formula in Riemannian geometry (see for example Kobayashi[21]) then yields

$$\text{div } X = \text{trace}(\nabla X + \text{Tor}(X, \cdot)).$$

If X is horizontal then the second part of the trace formula vanishes identically by the defining properties of the Webster-Tanaka connection. Using our particular choice of frame we then see that

$$\begin{aligned} \text{div } \nu &= \langle\, \nabla_\nu \nu\,,\, \nu \,\rangle + \langle\, \nabla_{J\nu} \nu\,,\, J\nu \,\rangle + \langle\, \nabla_T \nu\,,\, \nu \,\rangle \\ &= -\langle\, \nu\,,\, \nabla_{J\nu} J\nu \,\rangle \end{aligned}$$

as the first and third terms vanish because H is parallel and the connection is metric (i.e. $\langle \nabla_U V , W \rangle = -\langle \nabla_U W , V \rangle$) respectively. Thus on a non-characteristic, horizontally minimal surface patch we have

$$\nabla_{J\nu} J\nu = 0 \tag{4.8}$$

everywhere. The integral curves of $J\nu$ are therefore ∇-geodesics. But $J\nu$ spans the intersection of TS with H. Thus the integral curves of $J\nu$ foliate S.

■

Remark 4.3. The divergence form of the mimimal surface equation in $\mathcal{RT}$ was first shown by Citti and Sarti[7] (see section 2.9 of that paper) but, as referenced above, is also a consequence of the more general pseudoherimitian framework of Cheng, Hwang, Malchiodi and Yang[5]. We also note that a version of Theorem 4.2, showing that smooth minimal surfaces are ruled, was first shown in section 2 of the same paper,[5] again in the context of pseudohermitian manifolds. We include the proof of this fact here for completeness and because it facilitates the computations below.

We shall now apply these techniques to the special case to the roto-translation group $\mathcal{RT}$. Here the underlying manifold is $\mathbb{R}^2 \times \mathbb{S}^1$ and $\mathcal{V}_0$ is defined by setting

$$X_1 = \cos\theta \frac{\partial}{\partial x} + \sin\theta \frac{\partial}{\partial y}, \quad X_2 = \frac{\partial}{\partial \theta}$$

and declaring them to be a left-invariant, orthonormal frame for a distribution H. The Riemannian structure by defining the transverse vector field,

$$T = [X_1, X_2] = \sin\theta \frac{\partial}{\partial x} - \cos\theta \frac{\partial}{\partial y}$$

and declaring it to be unit length and orthogonal to X_1 and X_2. The remaining commutation relations can then be explicitly computed as

$$[X_1, T] = 0, \quad [X_2, T] = X_1.$$

When we repeat the construction of Lemma 4.1 we note that g is exactly the Levi metric in this case. The contact form can be explicitly computed as

$$\eta = \sin\theta dx - \cos\theta dy$$

and the dual to the complex vector field $Z = X_2 - iX_1$ is

$$\zeta = \frac{1}{2}\left(d\theta + i\cos\theta dx + i\sin\theta dy\right).$$

Straightforward computations then yield

$$\begin{aligned} d\eta &= \cos\theta d\theta \wedge dx + \sin\theta d\theta \wedge dy = 2i\zeta \wedge \bar{\zeta} \\ d\zeta &= \frac{i}{2} d\theta \wedge (-\sin\theta dx + \cos\theta dy) = -\frac{i}{2}\zeta \wedge \eta + \frac{i}{2}\eta \wedge \bar{\zeta}. \end{aligned}$$

The first identity also follows from the fact that the pseudohermitian structure was explicitly constructed to ensure that X_1 and X_2 were orthonormal. Since, using (4.3) and the form of $d\eta$, $h = 2$ we can immediately deduce from the 2nd Webster identity (4.4) that the connection form ω is pure imaginary. Thus we can deduce from (4.2) that

$$\omega = -\frac{i}{2}\eta, \quad \tau = \frac{i}{2}\bar{\zeta}.$$

This implies that for the frame X_1, X_2, T the only non-trivial covariant derivatives are in the T direction. By examining the real and imaginary parts of the equation

$$\nabla_T Z = -\frac{i}{2}Z$$

we see $\nabla_T X_1 = \frac{1}{2}X_2$ and $\nabla_T X_2 = -\frac{1}{2}X_1$.

With this initial compuations, we can compute the horizontal ∇-geodeiscs explicitly.

Theorem 4.4. *The horizontal ∇-geodesics in $\mathcal{RT}$ are determined by a choice of basepoint $(x_0, y_0, \theta_0) \in \mathcal{RT}$, a radius R_0, and velocity $\dot{\theta} \in \mathbb{R}$. The ∇-geodesic, $\gamma = (x, y, \theta)$ given by a choice $\{(x_0, y_0, \theta_0), R_0, \dot{\theta}\}$ is*

- *If $\dot{\theta} \neq 0$, $R_0 \neq 0$, then*

$$\begin{aligned} x &= x_c + R\sin\theta \\ y &= y_c - R\cos\theta \\ \theta &= \theta_0 + \dot{\theta}t. \end{aligned}$$

  where $R = R_0/\dot{\theta}, x_c = x_0 - R\sin\theta_0$ and $y_c = y_0 + R\cos\theta_0$.
- *If $\dot{\theta} = 0$, $R_0 \neq 0$, then*

$$\begin{aligned} x &= x_0 + R_0(\cos\theta_0)t \\ y &= y_0 + R_0(\sin\theta_0)t \\ \theta &= \theta_0. \end{aligned}$$

- *If $\dot{\theta} \neq 0$, $R_0 = 0$, then*

$$\begin{aligned} x &= x_0 \\ y &= y_0 \\ \theta &= \theta_0 + \dot{\theta}t. \end{aligned}$$

Proof: Consider a curve $\gamma = (x, y, \theta)$. Thus

$$\dot{\gamma} = (\dot{x}, \dot{y}, \dot{\theta}) = (\dot{x}\cos\theta + \dot{y}\sin\theta)X_1 + \dot{\theta}X_2 + (\dot{x}\sin\theta - \dot{y}\cos\theta)T.$$

Thus if γ is a purely horizontal curve, we must have

$$\dot{x}\sin\theta - \dot{y}\cos\theta = 0.$$

Under this assumption, we claim $\nabla_{\dot\gamma}\dot\gamma = 0$ if and only if both $\dot\theta$ and $\dot x\cos\theta + \dot y\sin\theta$ are constant. Computing, we have

$$D_t\dot\gamma = (\dot x\cos\theta + \dot y\sin\theta)_t X_1 + \ddot\theta X_2 + (\dot x\cos\theta + \dot y\sin\theta)D_t X_1 + \dot\theta D_t X_2$$

Since $D_t X_i = \nabla_{\dot\gamma} X_i$ and the only nontrivial covariant derivatives of X_i are in the T direction, the fact the $\dot\gamma$ is horizontal implies the claim. Denoting $\dot x\cos\theta + \dot y\sin\theta$ by R_0, we can then solve the equation

$$\begin{pmatrix}\sin\theta & -\cos\theta\\ \cos\theta & \sin\theta\end{pmatrix}\begin{pmatrix}\dot x\\ \dot y\end{pmatrix} = \begin{pmatrix}0\\ R_0\end{pmatrix}$$

to obtain $\dot x = R_0\cos\theta$, $\dot y = R_0\sin\theta$. The cases listed above are now immediate. ∎

Note that the second and third cases arises from the first as we take the limit as $\dot\theta \to 0$ and $R_0 \to 0$ respectively. In the sequel we shall refer to the horizontal ∇-geodesics (and connected subsets of them) as rules.

Next, we begin the investigation of the existence and properties of minimal ruled spanning surfaces of curves in $\mathcal{RT}$. As we require the solution to the occlusion problem to be a surface ruled by ∇-geodesics, we first look at the set of points which can be connected to a given point by ∇-geodesics. With this in mind, we make the following definition:

Definition 4.5. For a point $p \in \mathcal{RT}$ we define the accessible set $\mathscr{A}(p)$ to be the collection of points that can be connected to p by a single, horizontal ∇-geodesic.

Proposition 4.6. *Given a point $p = (x_0, y_0, \theta_0)$, the set of accessible points, $\{(x, y, \theta)\}$ is given by the implicit equation*

$$(x - x_0, y - y_0)\cdot\left(-\sin\left(\frac{\theta+\theta_0}{2}\right), \cos\left(\frac{\theta+\theta_0}{2}\right)\right) = 0 \tag{4.9}$$

Proof: When the connecting ∇-geodesic is a straight line this is immediate. Suppose that (x_0, y_0, θ_0) and (x, y, θ) are connected by a ∇-geodesic given by

$$c(t) = (x_c + R\sin(\dot\theta t + \theta_0), y_c - R\cos(\dot\theta t + \theta_0), \dot\theta t + \theta_0)$$

for fixed real numbers $\{x_c, y_c, R, \dot\theta\}$ as given in Theorem 4.4 so that $c(0) = (x_0, y_0, \theta_0)$ and $c(t_1) = (x, y, \theta)$. In particular, we will use the shorthand that $\dot\theta t_1 + \theta_0 = \theta$. Then, $x - x_0 = R(\sin(\theta) - \sin(\theta_0))$ and $y - y_0 = R(\cos(\theta_0) - \cos(\theta))$. If, for example, $x \neq x_0$, we have that

$$\frac{y - y_0}{x - x_0} = \frac{\cos(\theta_0) - \cos(\theta)}{\sin(\theta) - \sin(\theta_0)} \tag{4.10}$$

To rewrite this, we use the tangent half angle formulae, $\tan\left(\frac{\varphi}{2}\right) = \frac{\sin(\varphi)}{1+\cos(\varphi)} = \frac{1-\cos(\varphi)}{\sin(\varphi)}$, to see that,

$$\tan\left(\frac{\theta+\theta_0}{2}\right) = \frac{\cos\theta_0 - \cos\theta}{\sin\theta - \sin\theta_0}. \tag{4.11}$$

Thus, (4.10), reduces to

$$\frac{y-y_0}{x-x_0} = \tan\left(\frac{\theta+\theta_0}{2}\right)$$

Similarly, if $y \neq y_0$ we have

$$\frac{x-x_0}{y-y_0} = \cot\left(\frac{\theta+\theta_0}{2}\right)$$

Rewriting these yields (4.9). ∎

This provides a description of the accessible set of p:

Lemma 4.2. *Every accessible set $\mathscr{A}(p)$ is the image of an embedding of the Möbius strip into $\mathcal{RT}$.*

Proof: Consider the ∇-exponential map at p restricted to the horizontal distribution. From our explicit description of the horizontal geodesics passing through p we note that if $p = (x_0, y_0, \theta_0)$ then

$$\begin{aligned}\exp_p(aX_1 + bX_2) = \big(&x_0 + a/b(\sin(\theta_0+b) - \sin\theta_0),\\ &y_0 + a/b(\cos\theta_0 - \cos(\theta_0+b)), \theta_0 + b\big)\end{aligned} \tag{4.12}$$

at least when $b \neq 0$. When $b = 0$ we instead get

$$\exp_p(aX_1) = (x_0 + a\cos\theta_0, y_0 + a\sin\theta_0, \theta_0). \tag{4.13}$$

If $\exp(a,b) = \exp(a',b')$ we must therefore have that $b = b' + 2k\pi$ (with neither being 0) and $a'/b' = a/b$. This later can be summarized as the points (a,b) and (a',b') must lie on the same line through the origin with $b = b' + 2k\pi$. Therefore the exponential map is bijective from $\mathbb{R} \times [-\pi, \pi]$ to $\mathscr{A}(p)$ provided that the sides of the strip are identified via $(a, -\pi) \sim (-a, \pi)$. This provides an explicit embedding of the Möbius strip into $\mathcal{RT}$ with image $\mathscr{A}(p)$.

∎

For each point on a curve γ, $\mathscr{A}(\gamma(t))$ may contain many points of γ or very few. Of most interest to the question of building spanning surfaces are the points $\gamma(t)$ where $\gamma \cap \mathscr{A}(\gamma(t)) = \{\gamma(t)\}$ - i.e. the points that only connect to themselves. These points give constraints on the formation of a minimal ruled spanning surface. To help understand these points, we make the following definition:

Definition 4.7. Given an embedded curve γ and a point $p \in \gamma$ we define

$$\mathscr{A}(p, \gamma) = \gamma \cap \mathscr{A}(p),$$

the points in γ accessible to p. A point p such that $\mathscr{A}(p,\gamma) = \{p\}$ is called an solitary point of γ. The solitary points of γ will be denoted $\mathscr{I}(\gamma)$. A point $p = \gamma(t)$ such that $\dot\gamma \in \mathcal{V}_0$ is called a Legendrian point of γ. The Legendrian points of γ will be denoted $\mathscr{L}(\gamma)$. We also define the orthogonal points of γ, denoted $\mathscr{O}(\gamma)$ to be where $\dot\gamma \in \operatorname{span}\{X_2, X_3\}$.

As seen in the definition, there are two types of solitary points, the Legendrian points and the non-Legendrian points. We remark that for a Legendrian point p, the candidate rule passing through p is tangent to the curve γ and one can use this as a starting place for building a ruled minimal spanning surface. To investigate the structure of the set of solitary points further we prove the following lemma.

Lemma 4.3. *For any embedded curve γ, the set $\mathscr{I}(\gamma) - (\mathscr{L}(\gamma) \cap \mathscr{I}(\gamma))$ is open.*

Proof: For a point $p = (x_0, y_0, \theta_0)$, it follows that the accessible set $\mathscr{A}(p)$ can be parametrized by $F_p(s, \varphi)$

$$F_p \colon (s, \varphi) \in \mathbb{R} \times [0, 2\pi] \mapsto \left(x_0 + s\cos\left(\frac{\theta+\varphi}{2}\right), y_0 + s\sin\left(\frac{\theta+\varphi}{2}\right), \varphi \right) \in \mathcal{RT}.$$

Therefore we can define a map $D_t(s, \varphi)$ to be the minimal Euclidean distance between $\gamma(t)$ and $F_{\gamma(t)}(s, \varphi)$. This map is then continuous in all 3 variables.

At a non-Legendrian $p = \gamma(t)$ the curve γ intersects $\mathscr{A}(p)$ transversely at p. Thus the zero of D_t occuring at $(0, \theta(t))$ is isolated. Since the Legendrian points of γ occur exactly where $\dot{\gamma}$ is perpendicular to X_3, we see that $\mathscr{L}(\gamma)$ is closed as a subset of γ. Thus, as p is non-Legendrian, there is an open neighborhood of p in γ consisting only of non-Legendrian points.

Let $p = \gamma(t_0)$ be both solitary and non-Legendrian. Now suppose there is a sequence of non-Legendrian, non-solitary points $q_i = \gamma(t_i)$ converging to p. Non-solitary implies that there must be a second sequence $\tilde{q}_i$ of points which correspond to zeros of the respective D_{t_i} functions. By compactness, a subsequence $\tilde{q}_i$ must converge to some point $\tilde{q}$. Now $\tilde{q}$ cannot equal p as this would violate the isolated zero of D_{t_0} condition. But by continuity, $\tilde{q}$ must correspond to a zero of D_{t_0} violating the condition that p is solitary. This is a contradiction.

■

For the occlusion problem we shall work exclusively with curves that occur as the boundary of a smooth, simply connected bounded region D lifted by the contour direction field of an intensity function $I : \mathbb{R}^2 \to \mathbb{R}$. The boundary ∂D is can be viewed as the image of an embedding $\beta : \mathbb{S}^1 \to \mathbb{R}^2$. Away from critical points of I, we can define the function $\theta : \mathbb{S}^1 \to \mathbb{S}^1$ by

$$\theta(t) = \arctan\left(-\frac{I_x \circ \beta(t)}{I_y \circ \beta(t)} \right) \tag{4.14}$$

where at each point we choose the branch of arctan which makes θ continuous. As we use this setup often we make the following definition:

Definition 4.8. Let D be a bounded occlusion domain with boundary given by a C^1 simple closed curve, $I : \mathbb{R}^2 \to \mathbb{R}$ be a C^2 image function, β be a parameterization of the boundary by arclength oriented counterclockwise, and $\gamma(t) = (\beta(t), \theta(t)) \subset \mathcal{RT}$. The collection (D, I, γ) is called the *occlusion data* associated to the occlusion problem given by (D, I).

We define the normal angle function for β by

$$\beta'(t) = (-\sin n_\beta(t), \cos n_\beta(t)) \tag{4.15}$$

Using this notation, we point out that, under the assumption that β is oriented counterclockwise, the vector $(\cos(n_\beta(t)), \sin(n_\beta(t))$ is the outward pointing normal to β.

Using these, we have:

Definition 4.9. Let $I : \mathbb{R}^2 \to \mathbb{R}$ be a smooth image intensity function, $D \subset \mathbb{R}^2$ an open domain with C^1 boundary and $\beta : \mathbb{S}^1 \to \mathbb{R}$ a parameterization of ∂D. With $\theta(t), \gamma(t)$ and $n_\beta(t)$ defined as above, we define

$$Q(t) = \theta(t) - n_\beta(t).$$

to be the transversality function for γ.

Note that Q measures the angle between the curve β and the level set of I at a point of intersection.

Lemma 4.4. *For any lift γ associated to a completely non-degenerate occlusion problem,*

(a) *$\gamma(t)$ is Legendrian if and only if $Q(t) = \frac{\pi}{2} + k\pi$ for some integer k.*
(b) *$\gamma(t)$ is orthogonal if and only if $Q(t) = k\pi$ for some integer k.*
(c) *$\mathscr{L}(\gamma)$ is non-empty, in fact there must be at least 2 Legendrian points.*

Proof: For the first item, we note that, by definition, γ is Legendrian if $\gamma' \in \mathcal{V}_0$. Computing, we have

$$\begin{aligned}\langle \gamma'(t), X_3 \rangle &= \beta'(t) \cdot \left(\frac{\nabla I}{|\nabla I|}(\beta(t)) \right) \\ &= (-\sin(n_\beta(t)), \cos(n_\beta(t))) \cdot (-\sin(\theta(t)), \cos(\theta(t))) \\ &= (\cos(\theta(t) - n_\beta(t))) = \cos(Q(t))\end{aligned}$$

and so, if we assume $\gamma(t)$ is Legendrian, we have that, equivalently, $\beta'(t) \cdot \nabla I = 0$ or $Q(t) = \theta(t) - n_\beta(t) = \frac{\pi}{2} + k\pi$. The second item, concerning orthogonal points, follows from a similar computation of $\langle \gamma'(t), X_1 \rangle$.

The last item follows from the observation that

$$\int_{\partial D} \nabla I \cdot d\vec{r} = \int_0^{2\pi} |\nabla I(\beta(t))| \cos\alpha = 0$$

where α is the angle between ∇I and β'. Since the first two terms of the integral are strictly positive we must have $\cos\alpha$ taking both positive and negative values. In particular, this implies there are at least two points in $[0, 2\pi)$ where $\cos(\alpha) = 0$. By the computation at the outset of the proof, these two points are Legendrian points.

■

We record one more fact about θ:

Lemma 4.5.

$$\dot{\theta}(t) = -\frac{1}{|\nabla I|}\begin{pmatrix} \frac{I_y}{|\nabla I|} \\ -\frac{I_x}{|\nabla I|} \end{pmatrix}^t \mathscr{H} \begin{pmatrix} \beta_1'(t) \\ \beta_2'(t) \end{pmatrix} \tag{4.16}$$

where $\mathscr{H}$ is the Hessian of I. Letting

$$\kappa_1 = \begin{pmatrix} \frac{I_y}{|\nabla I|} \\ -\frac{I_x}{|\nabla I|} \end{pmatrix}^t \mathscr{H} \begin{pmatrix} \frac{I_y}{|\nabla I|} \\ -\frac{I_x}{|\nabla I|} \end{pmatrix}$$

and

$$\kappa_2 = \begin{pmatrix} \frac{I_y}{|\nabla I|} \\ \frac{-I_x}{|\nabla I|} \end{pmatrix}^t \mathscr{H} \begin{pmatrix} \frac{I_x}{|\nabla I|} \\ \frac{I_y}{|\nabla I|} \end{pmatrix}$$

we have that

$$\dot{\theta}(t) = -\frac{1}{|\nabla I|}(\kappa_1 \sin(Q(t)) + \kappa_2 \cos(Q(t)))$$

In particular,

$$|\dot{\theta}(t)| \leq \frac{2}{|\nabla I|}\kappa_{max}(t)$$

where $\kappa_{max}(t)$ is the maximum of the absolute value of the eigenvalues of $\mathscr{H}$ at $\beta(t)$.

Proof: (4.16) follows from direct computation and is left to the reader. To see the simplification, we rewrite $\mathscr{H}$ in terms of the orthonormal basis

$$\left\{ \begin{pmatrix} \frac{I_y}{|\nabla I|} \\ -\frac{I_x}{|\nabla I|} \end{pmatrix}, \begin{pmatrix} \frac{I_x}{|\nabla I|} \\ \frac{I_y}{|\nabla I|} \end{pmatrix} \right\}$$

Then, (4.16) becomes

$$-\frac{1}{|\nabla I|}\begin{pmatrix} 1 \\ 0 \end{pmatrix}^t \begin{pmatrix} \kappa_1 & \kappa_2 \\ \star & \star \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = -\frac{1}{|\nabla I|}(\kappa_1 a + \kappa_2 b)$$

where

$$a = \dot{\beta} \cdot \left(\frac{I_y}{|\nabla I|}, -\frac{I_x}{|\nabla I|} \right)$$

and

$$b = \dot{\beta} \cdot \left(\frac{I_x}{|\nabla I|}, \frac{I_y}{|\nabla I|} \right)$$

Computing a using the definition of $n_\beta, \theta(t)$ and Q yields:

$$\begin{aligned} a &= (-\sin(n_\beta(t)), \cos(n_\beta(t))) \cdot (\cos(\theta(t)), \sin(\theta(t))) \\ &= \sin(\theta(t) - n_\beta(t)) = \sin(Q(t)) \end{aligned}$$

Similarly, $b = \cos(Q(t))$ and the result follows.

■

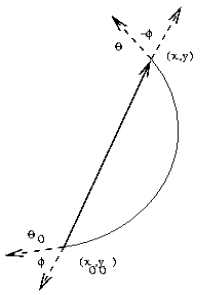


Fig. 3. Cocircularity condition

5. Cocircularity

We give an interpretation of the accesibility condition that has broad application in the study of image processing and recognition. An examination of (2.1) shows that the value of θ at a point $(x, y) \in \mathbb{R}^2$ is given by the angle the tangent vector to the level set of I passing through the point (x, y) makes with the x-axis. To interpret the condition of Proposition 4.6 we make a definition:

Definition 5.1. Given two points $(x_0, y_0), (x, y) \in \mathbb{R}^2$ and tangent angles θ_0, θ associated $(x_0, y_0), (x, y)$ respectively, we say that $((x_0, y_0), \theta_0)$ and $((x, y), \theta)$ are cocircular if there exists a circle C with $(x_0, y_0), (x, y) \in C$ and the tangent vectors to C at (x_0, y_0) and (x, y) are given by $(\cos(\theta_0), \sin(\theta_0))$ and $(\cos(\theta), \sin(\theta))$ respectively (see figure 3).

From Proposition 4.6, we see that the minimal surfaces precisely reflect the cocircularity condition. Equation (4.9) says that the vector $(x, y) - (x_0, y_0)$ points in the same direction as the difference of the vectors $(\cos(\theta), \sin(\theta))$ and $(\cos(\theta_0), \sin(\theta_0))$. Thus, the two triples $((x_0, y_0), \theta_0)$ and $((x, y), \theta)$ are cocircular if and only if the difference of the tangent vectors of the connecting circular arc point in the same direction as the line connecting (x_0, y_0) to (x, y). Note that collinearity arises in the special case when $\theta = \theta_0$.

The notion of cocircularity arises in a variety of contexts, notably within the Gestalt framework of good continuation of curves. In this setting, Gestalt psychologists identified cocircularity (and its special case colinearity) as one of the elements of the rules of good continuation.[34, 35]

Moreover, this notion has been used in variety of problems in digital image processing. For example, Parent and Zucker[16] introduced a cocircularity defect as a measure of curvature and variants of this idea have been used in a variety of applications (see for example, the work of Ben-Shahar and Zucker[2, 3] and that of Kimia, Frankel, Pompescu[23]).

In addition, there is strong statistical evidence (see Sigman, Cecchi, Gilbert and Magnasco[31]) that natural images obey cocircularity conditions and there is

biological evidence (see the work of Zhou, Samonds, Bernard and Bonds[36]) that cells in the primary visual cortex respond to both colinear and cocircular stimuli. This provides additional evidence of the suitability of the Citti-Sarti model.

6. Constructing minimal surfaces

The accessibility condition (4.9) allows us to explicitly construct C^2 disocclusions (when they exist).

Definition 6.1. Let (D, I, γ) be a set of occlusion data. We say that a function $u(t) : \mathbb{S}^1 \to \mathbb{S}^1$ satisfying either:

$$(\beta(u(t)) - \beta(t)) \cdot \left(-\sin\left(\frac{\theta(u(t)) + \theta(t)}{2}\right), \cos\left(\frac{\theta(u(t)) + \theta(t)}{2}\right)\right) = 0 \qquad (6.1)$$

or (equivalently):

$$(\beta(u(t)) - \beta(t)) \cdot (\cos(\theta(u)) - \cos(\theta(t)), \sin(\theta(u)) - \sin(\theta(t))) = 0 \qquad (6.2)$$

where $\theta(t)$ is the function defined in (4.14), is a solution for the occlusion data (D, I, γ).

Note that the regularity properties of u depend on those of β and I. Under the assumptions, that $I \in C^{k+1}, \beta \in C^k, \theta \in C^k$, we have that $u \in C^k$.

Remark 6.2. We note that (6.1) and (6.2) always have a solution, namely $u(t) = t$. This solution corresponds to a cyclindrical surface in $\mathcal{RT}$ that connects $\beta(t)$ to itself via a straight line in the ∂_θ direction. Given our application to the problem of completing occlusions, we will disregard this solution as it does not provide a solution to the disocclusion problem. Consequently, we will focus on other possible solutions of (6.1) or (6.2) and point out that, *a priori*, there may not exist other such solutions.

In fact, we will disregard solutions that have significant pieces where $u(t) = t$:

Definition 6.1. A solution to a set of occlusion data (D, I, γ) is called *non-viable* if there exists a connected interval J so that $u(t) = t$ for all $t \in J$. A solution without this property is called *viable*. Similarly, minimal surfaces associated to these solutions are called non-viable and viable respectively.

This provides us with an obstruction to viable solutions:

Theorem 6.3. *Let (D, I, γ) be a set of occlusion data. If $\gamma = (\beta, \theta) \subset \mathcal{RT}$ contains non-Legendrian solitary points, then there does not exist a viable C^2 minimal spanning surface for this occlusion problem.*

Proof: By Definition 4.7, for a solitary point corresponding to the parameter t, the only possible value for $u(t)$ is t itself. Moreover, Lemma 4.3 shows that non-Legendrian solitary points form an open set on the curve. Thus, under the hypotheses of this Theorem, there exists an open interval where any solution to (6.1) or (6.2) satisfies $u(t) = t$. ∎

Remark 6.4. Note that, in particular, this implies that if there exist non-Legendrian solitary points, there cannot exist a smooth minimal spanning graph. This is Theorem 1.3.

Proposition 6.5. *If (x_0, y_0, θ_0) is connected to (x_1, y_1, θ_1) with $\theta_0 \neq \theta_1$ then, the connecting rule must have the parametrization $(x_c + R\sin\varphi, y_c - R\cos\varphi, \varphi)$ where $x_c = x_0 - R\sin(\theta_0), y_c = y_0 + R\cos(\theta_0)$ and if $x_0 \neq x_1$ we have*

$$R = \frac{x_1 - x_0}{\sin\theta_1 - \sin\theta_0} \tag{6.3}$$

otherwise

$$R = \frac{y_1 - y_0}{\cos\theta_0 - \cos\theta_1} \tag{6.4}$$

Proof: As (x_0, y_0, θ_0) can be connected to (x_1, y_1, θ_1) with $\theta_0 \neq \theta_1$, by Theorem 4.4 the connecting rule must have the parametrization $(x_c + R\sin\varphi, y_c - R\cos\varphi, \varphi)$. Thus we need only solve the matrix equation

$$\begin{pmatrix} 1 & 0 & \sin\theta_0 \\ 0 & 1 & -\cos\theta_0 \\ 1 & 0 & \sin\theta_1 \\ 0 & 1 & -\cos\theta_1 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ R \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ x_1 \\ y_1 \end{pmatrix} \tag{6.5}$$

where the fact that the points are connectable guarantees the existence of a solution. For example, elementary methods yield that if $x_1 \neq x_0$,

$$R = \frac{x_1 - x_0}{\sin\theta_1 - \sin\theta_0}, \quad x_c = x_0 - R\sin\theta_0, \quad y_c = y_0 + R\cos\theta_0. \tag{6.6}$$

If $x_1 = x_0$, the second statment follows from a similar computation.

This of course yields two separate connecting rule segments depending on whether φ transverses $\mathbb{S}^1$ clockwise or counterclockwise.

∎

Definition 6.6. If (x_0, y_0, θ_0) is connected to (x_1, y_1, θ_1) with $\theta_0 = \theta_1$ then the connecting rule is a horizontal straight line and we define $R = \infty$.

Definition 6.7. Let u be a solution for the occlusion data (D, I, γ). We define the function $R(t)$ to be R restricted to $\beta(t)$. In other words, $R(t)$ is characterized by

$$\begin{pmatrix} 1 & 0 & \sin\theta(t) \\ 0 & 1 & -\cos\theta(t) \\ 1 & 0 & \sin\theta(u(t)) \\ 0 & 1 & -\cos\theta(u(t)) \end{pmatrix} \begin{pmatrix} \beta_1(t) - R(t)\sin(t) \\ \beta_2(t) + R(t)\cos(t) \\ R(t) \end{pmatrix} = \begin{pmatrix} \beta_1(t) \\ \beta_2(t) \\ \beta_1(u(t)) \\ \beta_2(u(t)) \end{pmatrix} \tag{6.7}$$

We note that as, under our assumption, u, β, θ are differentiable, $R(t)$ is differentiable in neighborhoods where it is finite.

Lemma 6.1. *Let u be a solution for the occlusion data (D, I, γ). If $R(t) = 0$ for a fixed t, then $u(t) = t$.*

Proof: The statement follows immediately from substituting $R(t) = 0$ in (6.7). ∎

We note that $R'(t)$ may be computed explicitly from this characterization entirely in terms of boundary data at $\gamma(t)$ and $\gamma(u(t))$.

The balance of this section is used to explore the behavior of solutions u for a set of occlusion data (D, I, γ) near points of interest. In particular, we will focus on points where $u(t) = t$ as they provide significant information concerning the construction of a minimal spanning surface. In particular, solitary points must satisfy this condition for any solution u.

We first compute the derivative of u in terms of the information on the bounding curve.

Lemma 6.2. *If $u : \mathbb{S}^1 \to \mathbb{S}^1$ is a solution for a set of occlusion data (D, I, γ), then*

$$\dot{u} = \frac{\cos\left(Q(t) + \frac{\Delta\theta}{2}\right) - \frac{\varepsilon(t)}{2}\dot{\theta}(t)}{\cos\left(Q(u) - \frac{\Delta\theta}{2}\right) + \frac{\varepsilon(t)}{2}\dot{\theta}(u)} \tag{6.8}$$

where $\Delta\theta = \theta(u) - \theta(t)$,

$$\varepsilon(t) = (\beta(u) - \beta(t)) \cdot \left(-\cos\left(\frac{\theta(u) + \theta(t)}{2}\right), -\sin\left(\frac{\theta(u) + \theta(t)}{2}\right)\right)$$

and we use the notational convienience of denoting $u(t)$ by u.

Proof: Implicitly differentiating (6.1) yields

$$\begin{aligned} 0 = {} & (\dot{u}\dot{\beta}(u) - \dot{\beta(t)}) \cdot \left(-\sin\left(\frac{\theta(u) + \theta(t)}{2}\right), \cos\left(\frac{\theta(u) + \theta(t)}{2}\right)\right) \\ & + (\beta(u) - \beta(t)) \cdot \left(-\cos\left(\frac{\theta(u) + \theta(t)}{2}\right), -\sin\left(\frac{\theta(u) + \theta(t)}{2}\right)\right) \left(\frac{1}{2}(\dot{\theta}(u)\dot{u} + \dot{\theta}(t)\right) \end{aligned} \tag{6.9}$$

We start by examining the first term on the right hand side, which we will denote $\boxed{\text{A}}$. Note that

$$\begin{aligned} \dot{\beta}(s) \cdot \left(-\sin\left(\frac{\theta(u) + \theta(t)}{2}\right), \cos\left(\frac{\theta(u) + \theta(t)}{2}\right)\right) = {} & \\ & \left(\sin(n_\beta(s))\sin\left(\frac{\theta(u) + \theta(t)}{2}\right) + \cos(n_\beta(s))\cos\left(\frac{\theta(u) + \theta(t)}{2}\right)\right) \\ & = \cos(\theta(s) - n_\beta(s)) \\ & = \cos(Q(s)) \end{aligned}$$

Here, we used the definition of n_β from (4.15) and Definition 4.9. Using this, the first term becomes

$$\boxed{\text{A}} = \dot{u}\cos\left(Q(u) - \frac{\Delta\theta}{2}\right) - \cos\left(Q(t) + \frac{\Delta\theta}{2}\right) \tag{6.10}$$

where $\Delta\theta = \theta(u) - \theta(t)$.

Turning to the second term, which we denote $\boxed{\text{B}}$, we note that (6.1) implies $\beta(u) - \beta(t)$ is parallel to the vector

$$\left(\cos\left(\frac{\theta(u)+\theta(t)}{2}\right), \sin\left(\frac{\theta(u)+\theta(t)}{2}\right)\right)$$

Letting $\varepsilon(t) = (\beta(u) - \beta(t)) \cdot \left(-\cos\left(\frac{\theta(u)+\theta(t)}{2}\right), -\sin\left(\frac{\theta(u)+\theta(t)}{2}\right)\right)$, we have

$$\boxed{\text{B}} = \frac{\varepsilon(t)}{2}(\dot{\theta}(u)\dot{u} + \dot{\theta}(t)) \tag{6.11}$$

Using (6.10) and (6.11), we have

$$\begin{aligned} 0 = \boxed{\text{A}} + \boxed{\text{B}} &= \dot{u}\cos\left(Q(u) - \frac{\Delta\theta}{2}\right) - \cos\left(Q(t) + \frac{\Delta\theta}{2}\right) \\ &\quad + \frac{\varepsilon(t)}{2}(\dot{\theta}(u)\dot{u} + \dot{\theta}(t)) \end{aligned}$$

Solving the second term of the product above for $\dot{u}$ yields the desired result.

■

Remark 6.8. We can read some qualitative information from this expression for $\dot{u}$. In particular, if $|\beta(u) - \beta(t)|$ is small then the behavior of u is governed, in large part by Q and $\Delta\theta$. However, when $|\beta(u) - \beta(t)| >> 0$ (as may happen when the occluded region is very large), then the behavior of u is governed by the behavior of $\dot{\theta}$.

The following corollary follows from work in the proof of the the previous lemma.

Corollary 6.9. *Let (D, U, γ) be a set of occlusion data and let u be a solution for this occlusion data. Then, it there exists a t_0 so that $u(t_0) = t_0$ and $\gamma(t_0)$ is not Legendrian, then $u'(t_0) = 1$.*

Proof: Equation (6.9) in the proof of the previous lemma shows that if $u(t_0) = t_0$ than either $\dot{u} = 1$ or

$$\dot{\beta}(t_0) \cdot (-\sin(\theta(t_0)), -\cos(\theta(t_0))) = 0$$

Using (4.15) and trigonometric simpliciation, we see that this is equivalent to

$$\cos(Q(t_0)) = 0$$

And, we conclude that $Q(t) = \frac{\pi}{2} + k\pi$ and the result follows from Lemma 4.4. ■

7. Image disocclusions

In the previous sections, we have focused on describing the properties of disocclusions in $\mathcal{RT}$ - the smooth minimal spanning surfaces for fixed boundaries. The Citti-Sarti model provides a link between minimal surfaces and *image* disocclusions.

To study the problem of image disocclusion, we will construct a map which associates a minimal surface to an infilling of a planar region. Let (D, I, γ) be a set of occlusion data and let $u : \mathbb{S}^1 \to \mathbb{S}^1$ be a one to one solution for this data. We define a number of useful functions. Recall that the radius function at a generic parameter t is given by

$$R(t) = \frac{\beta_1(u(t)) - \beta_1(t)}{\sin(\theta(u(t))) - \sin(\theta(t))} = \frac{\beta_2(u(t)) - \beta_2(t)}{\cos(\theta(u(t))) - \cos(\theta(t))} \tag{7.1}$$

We note that $R(t) \to \pm\infty$ as $t \to t_0$ is equivalent to $\theta(t_0) = \theta(u(t_0))$ where $u(t_0) \neq t_0$. We define a function

$$\vartheta(t, u) = \begin{cases} R(t)(\theta(u(t)) - \theta(t)) \cos(Q(t)) & \text{if } R(t) < \infty \\ (-\dot\beta_2(t), \dot\beta_1(t)) \cdot (\beta(u(t)) - \beta(t)) & \text{if } R(t) = \infty \end{cases} \tag{7.2}$$

Finally, the function $\varphi(r)$ is defined by

$$\varphi(r) = \begin{cases} (\theta(u(t)) - \theta(t))r + \theta(t) & \text{if } \vartheta(t, u) \leq 0 \\ (\theta(t) - \theta(u(t)))r + \theta(u(t)) & \text{if } \vartheta(t, u) > 0. \end{cases} \tag{7.3}$$

Thus the function $\varphi(r)$ linearly interpolates between $\theta(t)$ and $\theta(u)$ in the direction required to keep the ruling internal to the occlusion domain.

Using these auxillary functions, we can make the following definition.

Definition 7.1. Then, for t so that $0 < |R(t)| < \infty$, we define $\Pi : \mathbb{S}^1 \times [0, 1] \to \mathbb{R}^2$ by

$$\Pi(t, r) = \mathit{Rule}(t, u(t))(r)$$

where

$$\mathit{Rule}(t, u(t))(r) = \Big(\beta_1(t) + R(t)\big(\sin(\varphi(r)) - \sin(\theta(t))\big), \\ \beta_2(t) - R(t)\big(\cos(\varphi(r)) - \cos(\theta(t))\big)\Big)$$

For points where $R(t) \to \pm\infty$ as $t \to t_0$, we may extend the definition of Π via a limit. Using (7.1) and Definition 7.1, in a case where $\vartheta(t, u) > 0$ for $t > t_0$ yields

$$\lim_{t \to t_0^{\#}} \mathit{Rule}(t, u)(r) = \big((1 - r)\beta_1(t_0) + r\beta_1(u(t_0)), (1 - r)\beta_2(t_0) + r\beta_2(u(t_0)\big) \tag{7.4}$$

The other case (when $\vartheta(t, u) < 0$) yields a similar formula:

$$\lim_{t \to t_0^{\#}} \mathit{Rule}(t, u)(r) = \big((1 - r)\beta_1(u(t_0)) + r\beta_1(t_0), (1 - r)\beta_2(u(t_0)) + r\beta_2(t_0)\big) \tag{7.5}$$

In each of these cases we use # to denote the appropriate one-sided limit. For open intervals on which $R(t) = \infty$, we define

$$Rule(t, u(t))(r) = \begin{cases} (1-r)\beta(t) + r\beta(u(t)) & \text{if } \vartheta(t,u) < 0 \\ (1-r)\beta(u(t)) + r\beta(t) & \text{if } \vartheta(t,u) > 0 \end{cases} \tag{7.6}$$

Recalling that the inward pointing normal to β is given by $(-\cos(n_\beta(t)), -\sin(n_\beta(t)))$, we see that the definition above forces the curve $Rule(t, u(t))$ to move into the interior of D. For the case when $R = \infty$, this is immediate from the definition and for the cases where $0 < |R| < \infty$, we see that

$$\frac{\partial \Pi}{\partial r} \cdot (-\cos(n_\beta(t)), -\sin(n_\beta(t))) = -R(t)\varphi'(r)\cos(Q(t))$$

Computation shows that the choice of $\varphi(r)$, depending on the sign of $\vartheta(t, u)$, makes this quantity positive if $\vartheta(t, u) \neq 0$. Of course, depending on the shape of the curve β, $Rule(t, u)$ may not lie entirely inside D.

We observe that the choice of $\varphi(r)$ changes in three possible places:

(a) For t_0 where $R(t_0) = 0$
(b) For t_0 where $R(t) = \infty$ (or, equivalently, where $\theta(u(t_0)) = \theta(t_0)$ but $u(t_0) \neq t_0$).
(c) At Legendrian points, i.e. where $\cos(Q(t_0)) = 0$.

Lemma 7.1. *Suppose I is an interval so that for all $t_0 \in I$, $\gamma(t_0)$ is not Legendrian, $R(t_0) \neq 0$ and $R(t) \neq \infty$. Then, Π restricted to $I \times [0, 1]$ is differentiable.*

Proof: On any subinterval which excludes parameter values where $\theta(t) = \theta(u(t))$, we have a single choice for $\varphi(r)$ and $\beta(t), R(t), \theta(t), u(t)$ are all differentiable, so is Π. ∎

We now have the language to precisely define an image disocclusion.

Definition 7.2. Let u be a solution for a set of occlusion data (D, I, γ) and Σ be a C^2 smooth minimal surface spanning γ associated to u. An *image disocclusion* of I on D associated to Σ is a function $\tilde{I} : D \to \mathbb{R}$ depending on functions f_t and F where

(a) For the rule connecting t to $u(t)$, $Rule(t, u(t))$, f_t is a function $Rule(t, u(t)) \to \mathbb{R}$ with the property that $f_t(Rule(t, u(t))(0)) = I(\beta(t)), f_t(Rule(t, u(t))(1)) = I(\beta(u(t)))$. Further, as $Rule(t, u(t))$ and $Rule(u(t), t)$ coincide as point sets, we require the $f_t, f_{u(t)}$ coincide as well.
(b) For each $p \in \Sigma$, let $\{t_i\}$ be the list of parameter values with the property that $p \in Rule(t_i, u(t_i))$ for all i. Then choose $F(p)$ to be a function that depends only on $\{f_{t_i}(p)\}$. Further, if $\{t_i\}$ contains a single value t_0, we require that $F(p) = f_{t_0}(p)$.

and, for p in the interior of D, $\tilde{I}(p) = F(p)$ and $\tilde{I}|_{\partial D} = I|_{\partial D}$

The best possible outcome of the construction of an image disocclusion happens when each $p \in \Sigma$ has a unique rule containing it. Then, up to a choice of the functions f_t, the ruled surface provides a unique and unambiguous completion of the function I on D which is compatible with the values of I on ∂D. Needless to say, this does not always happen nor should it. There are numerous examples of modal and amodal completions that produce conflicting or ambiguous data. For example, the Kanisza fish and the cross occlusions discussed by Citti and Sarti[7] provide examples where there are multiple equally plausible completions. We wish to explore how this type of behavior can be detected using the machinery of the last sections. Ambiguous data potentially arises when the projections of two segments of rules to the xy-plane intersect (i.e. $\Pi(t_1, r_1) = \Pi(t_2, r_2)$). One way this can occur is if the matching function u *backtracks* when transversing the curve γ.

Definition 7.3. Let u be a solution to (6.1) for a curve $\gamma = (\beta, \theta)$. We say that u backtracks is there exist two parameter values $t_1 < t_2$ so that $u(t_1) < u(t_2)$.

As an immediate consequence of the definition and the mean value theorem, we have

Proposition 7.4. *Let u be a nontrivial solution to* (6.1) *for a curve $\gamma = (\beta, \theta)$. Then u is backtracking if and only if there exist parameter values where $\dot{u} > 0$*

Another way in which ambiguous data can arise is if Π ceases to be a local diffeomorphism. Calculating the Jacobian of the map Π yields the following characterization of this situation.

Proposition 7.5. *For t values where $R(t)$ is finite, the map Π is a local diffeomorphism at (t, r) if*

$$0 \neq R(t)\frac{\partial \varphi(r)}{\partial r}\Big(\cos(Q(t)) + R'(t)(\cos(\varphi(r) - \theta(t)) - 1) - R(t)\left(\frac{\partial \varphi(r)}{\partial r}\sin(\varphi(r) - \theta(t))\right)\Big).$$

On open intervals of t values where $R = \infty$ and Π is given by (7.6), *Π is a local diffeomorphism if*

$$0 \neq (1-r)((\beta(u) - \beta(t)) \cdot (\dot{\beta}_2(t), -\dot{\beta}_1(t)) + r\dot{u}(t)((\beta(u) - \beta(t)) \cdot (\dot{\beta}_2(u), -\dot{\beta}_1(u))$$

or

$$0 \neq r((\beta(u) - \beta(t)) \cdot (\dot{\beta}_2(t), -\dot{\beta}_1(t)) + (1-r)\dot{u}(t)((\beta(u) - \beta(t)) \cdot (\dot{\beta}_2(u), -\dot{\beta}_1(u))$$

We note that at $r = 0$, the first quantity is

$$R(t)\frac{\partial \varphi(r)}{\partial r}\cos(Q(t)) = \vartheta(t, u)$$

which is nonzero away from Legendrian points and points where $R(t) = 0$. The second quantity is

$$(\beta(u) - \beta(t)) \cdot (\dot{\beta}_2(t), -\dot{\beta}_1(t))$$

which is zero at Legendrian points, and the third quantity is

$$-\dot{u}(t)((\beta(u)-\beta(t))\cdot(\dot{\beta}_2(u),-\dot{\beta}_1(u))$$

which is zero at Legendrian points and places where $\dot{u}=0$. Thus, away from the union of such points, we have that the map is a local diffeomorphism in a neighborhood of β.

Lastly, ambiguous data may arise if the rule, when projected to the xy-plane, has portions outside of the occlusion domain or if Π is not a global diffeomorphism. These two cases depend heavily on the shape and structure of β and θ.

In another direction, we may not have a satisfactory image disocclusion if there is *too little* data. In particular, this may happen if Π does not map onto the domain D.

Proposition 7.6. *Let u be a solution for a set of occlusion data (D,I,γ). Suppose there is an interval $I=(t_1,t_2)$ so that*

(a) *γ_I has no Legendrian points.*
(b) *$\dot{u}|_I<0$*
(c) *$\beta|_I\cap\beta|_{u(I)}=\emptyset$*
(d) *$Rule(t,u(t))\subset\overline{D}$ for all $t\in\bar{I}$.*

If D_0 is the domain bounded by $\beta|_I,\beta|_{u(I)},Rule(t_1,u(t_1)),Rule(t_2,u(t_2))$, then Π maps $\bar{I}\times[0,1]$ onto D_0.

Proof: We first consider the case where for all $t\in I$, $R(t)$ is finite. We note that the third condition in the Theorem excludes any points where $u(t)=t$. Thus, we cannot have solitary points of any kind. Hence, by Lemma 6.1, there are no $t\in I$ so that $R(t)=0$.

Suppose that $x\in D_0$ is not in the image of Π but that there are points x_i arbitrarily close to x so that the x_i are in the image of Π. The continuity given by Lemma 7.1 immediately yields a contradiction.

Suppose now that there exists $x\in D_0,\varepsilon>0$ so that $B(x,\varepsilon)$ is not covered by the image under Π of $[t_1,t_2]\times[0,1]$. Assume further that ε is maximal so that there exist $t_3<t_4\in[t_1,t_2]$ with $\{p_i\}=Rule(t_i,u(t_i))\cap\overline{B}(x,\varepsilon)$ for $i=3,4$. We note that there may only be a single such intersection. Indeed, if there were two intersections, either both are tangential or at least one intersection is transverse. If both are tangential, $Rule(t_i,u(t_i))\subset\partial B(x,\varepsilon)$ and we must have $\partial B(x,\varepsilon)$ is tangent to β and hence $\gamma(t_i)$ is Legendrian in violation of the first hypothesis. If at least one intersection is transverse, then $Rule(t_i,u(t_i))$ would pass into the interior of $B(x,\varepsilon)$, violating our assumption on the existence of such a ball. Let $r(t_i)$ be the number so that $p_i=\Pi(t_i,r(t_i))$. Now, for the two pairs of points, $\{p_i\}$ and $\{\beta(u(t_i))\}$, we may connect each p_i to one of the $\{\beta(u(t_i))\}$ with a line $\mathcal{C}_i$ so that $\mathcal{C}_3\cap\mathcal{C}_4=\emptyset$. We construct a curve $\mathcal{C}$ as the union of

(a) The line joining p_2 and p_3 through the interior of $B(x,\varepsilon)$.

(b) The line segments $\mathcal{C}_3$ and $\mathcal{C}_4$.

By the assumption that $Rule(t, u(t)) \subset D$ and $\dot{u} < 0$ (so that $\beta|_{u([t_3,t_4])}$ has endpoints $\{\beta(u(t_i))\}$), we have that the region D_1 bounded by $\mathcal{C}$ and $\beta|_{u(I)}$ is homeomorphic to a disk. Now, for each $t \in I$, we have that $\Pi(t, 0) \in \overline{D} \setminus D_1$ and $\Pi(t, 1) = \beta(u(t)) \in D_1$. Thus, there exists $r(t)$ so that $\Pi(t, r(t))$ is the smallest value of r for which $\Pi(t, r(t)) \in D_1$. Since Π is differentiable and $\mathcal{C} \cup \beta|_{u(I)}$ is piecewise smooth, we have that $r(t)$ is at least continuous. Moreover, our assumption that $Rule(t, u(t)) \cap B(x, \varepsilon) = \emptyset$ implies that for each $t \in [t_3, t_4]$, $\Pi(t, r(t))$ lies in either $\mathcal{C}_3$ or $\mathcal{C}_4$. However, as $(t, r(t))$ is a connected set and Π is continuous, we must have that the image of $(t, r(t))$ for $t \in [t_3, t_4]$ is connected. Since $\Pi(t_3, r(t_3)) \in \mathcal{C}_3$, $\Pi(t_4, r(t_4)) \in \mathcal{C}_4$ and $\mathcal{C}_3 \cup \mathcal{C}_4$ is disconnected, we reach a contradiction.

To extend this result to parameter values where $R(t) = \infty$, we first note that we may decompose I into open subintervals where either R is finite or $R = \infty$ and so that the closure of the union of all of these intervals contains I. For any subinterval where $R = \infty$, all of the rules are straight lines and the result follows easily from the monotonicity assumption (that $\dot{u} < 0$).

∎

Definition 7.7. Given (D, I, γ) a set of occlusion data, we say that D has an infilling if Π maps $\mathbb{S}^1 \times [0, 1]$ onto D.

We note that given a fixed occlusion domain, D, with boundary β and associated functions θ and Q, we may use Proposition 7.4, Proposition 7.5 and Proposition 7.6 to explicitly test potential solutions to the image disocclusion problem. For example, if, for a fixed t, we find a potential value for $u(t)$ (e.g. through numerical methods), these three conditions may be tested entirely in terms of data prescribed on the boundary *without having an explicit formulation for the whole function* u. Thus, this provides a method for constructing effective disocclusions pointwise and is the basis for all the algorithms derived from the analysis of minimal surfaces. As we will see in Section 10, we can use this construction to implement a disocclusion scheme for digital images. While, in Section 10, we focus on circular occlusion domains (using the results of the next section), these methods can be used to construct and test disocclusions for more general occlusion domains following this general outline:

(a) Consider a greyscale image that is represented by a smooth function, I.
(b) Fix an occlusion boundary, a simple closed C^1 curve, $\beta(t)$, oriented counterclockwise.
(c) Compute $\theta(t)$ and $Q(t)$.
(d) Test $\gamma = (\beta, \theta)$ for non-Legendrian solitary points. If they exist, the algorithm fails.
(e) For each t, numerically solve for all possible values of $u(t)$ using (6.1) or (6.2).
(f) Assess the suitability of each potential $u(t)$ for

   i. Backtracking using Proposition 7.4 and Lemma 6.2

ii. Local diffeomorphism using Proposition 7.5 and the formulae for $Q(t)$, $R(t)$ and $R'(t)$ using Definition 4.9 and (7.1).

(g) Construct one or more possible solution functions $u(t)$. At this stage, one may choose to ignore solutions that are found unsuitable in the previous step.
(h) Construct all rules, $Rule(t, u(t))$, using the data from the function u and the formulae in Theorem 4.4.
(i) For each solution u, with a set of identified values t of incompatibility, choose functions f_t and F as in Definition 7.2.
(j) Construct an image disocclusion using Definition 7.2 and the map Π.

If we are concerned with applying this algorithm to the disocclusion problem for digital images, the first step may *a priori* present problems; all digital data is, by definition, given as discrete data and is not smooth. There are two reasonable solutions to this problem. First, one could smooth the image using convolution with a Gaussian kernel to create a smooth function with which to work. Second, one can calculate discrete derivative as a substitute. We note that the algorithm above relies heavily on the linearity of the derivative (i.e. the fact that a directional derivative can be computed using a dot product with the gradient) and standard difference quotient methods will thus introduce significant error. However, this is easily fixed by using the discrete derivatives of Farid and Simoncelli[10] who generate filters for computing derivatives that precisely reflect the desired linearity. If one uses such derivatives and a quantization of the various variables, then the algorithm above can be followed exactly without smoothness of the image function.

The main difficulty in implementation of this algorithm arises in from the choice of the f_t and F and in the resolution of the situation in (d) where the algorithm fails.

8. Occluded Disks

To explore the framework presented in the last sections, we will retrict our attention to a specific class of occlusions, those where β is a circle, to better understand the various possibilities described in the results above as well as to give a more concrete description of the infilling construction. We first note that there is an ambiguity inherent in the definition of θ associated with the choice of the branch of the arctangent function: Given a point (x, y) and a contour passing through this point at angle θ, it is unclear whether to lift it to (x, y, θ) or $(x, y, \theta + \pi)$. Accordingly, for a point $p = (x, y, \theta) \in \mathcal{RT}$ we shall define its conjugate point to be $\overline{p} = (x, y, \theta + \pi)$ and frequently consider conjugate lifts γ and $\overline{\gamma}$ simultaneously.

Any circular lift γ can be expressed parametrically in standard form as

$$\gamma(t) = \big(\beta(t), \theta(t)\big) = \big(x_0 + \rho\cos t, y_0 + \rho\sin t, \theta(t)\big). \tag{8.1}$$

When γ is understood, we shall frequently refer to a point of $p \in \gamma$ simply by its parameter value with this parametrization. With this parametrization understood,

we have $n_\beta(t) = t$ and we can simplify the transversality function Q to

$$Q(t) = \theta(t) - t. \tag{8.2}$$

Lemma 8.1. *For a circular lift $\gamma = (x, y, \theta)$, the non-trivial part of $\mathscr{A}(\gamma(t), \gamma)$ is given implicitly by*

$$Q(t) + Q(u) = (2k+1)\pi, \qquad k \in \mathbb{Z} \tag{8.3}$$

and $\mathscr{A}(\gamma(t), \overline{\gamma})$ is given implicitly by

$$Q(t) + Q(u) = 2k\pi, \qquad k \in \mathbb{Z}. \tag{8.4}$$

Proof: From the work of Proposition 4.6 we see that $\gamma(u) \in \mathscr{A}(t, \gamma)$ if and only if

$$\tan\left(\frac{\theta(u)+\theta(t)}{2}\right) = \frac{y(u)-y(t)}{x(u)-x(t)} = -\frac{\sin u - \sin t}{\cos t - \cos u}.$$

Applying the trigonometric identity (4.11) we see that this is equivalent to

$$\tan\left(\frac{\theta(u)+\theta(t)}{2}\right) = -\cot\left(\frac{u+t}{2}\right).$$

From this it follows that

$$\frac{\theta(u)+\theta(t)}{2} = \frac{u+t-\pi}{2} \mod 2\pi.$$

The first result is then immediate. A virtually identical arguments yields the second part also.

∎

Lemma 8.2. *Conjugation twist-commutes with the exponential map in the sense that.*

$$exp_{\overline{p}}(a, b) = \overline{exp_p(-a, b)}.$$

Proof: This follows from direct computation from (4.12) and (4.13). ∎

Corollary 8.1. *If $\overline{q} \in \mathscr{A}(p)$ then $q \in \mathscr{A}(\overline{p})$. Furthermore the projections to $\mathbb{R}^2$ of the connecting rules match precisely.*

Therefore when constructing an image disocclusion we may consider how points in γ can be connected to γ as well as to $\overline{\gamma}$. We shall say that parameter values t and u are connected, if there either $\gamma(u), \overline{\gamma}(u) \in \mathscr{A}(\gamma(t))$.

Corollary 8.2. *Legendrian points connect only to Legendrian points. Orthogonal points connect only to orthogonal points.*

Furthermore, any rule connecting distinct Legendrian points projects to the bounding circle ∂D.

Proof: The first part follows from Lemma 8.1 and Lemma 4.4. The second part follows from the elementary observation that two circles which intersect tangentially at distinct points must be the same.

■

Corollary 8.3. $\overline{\mathscr{L}(\gamma)} = \mathscr{L}(\overline{\gamma})$, $\overline{\mathscr{O}(\gamma)} = \mathscr{O}(\overline{\gamma})$.

This follows immediately from the characterization of Legendrian and orthogonal points in Lemma 4.4.

Corollary 8.4. *On an implicit plot of all points (t, u) such that there exists a rule connecting $\gamma(t)$ and $\gamma(u)$, any transverse crossing of the leading diagonal $u = t$ occurs at either a Legendrian or an orthogonal point.*

Proof: A transverse crossing of the leading diagonal at t implies the existence of a sequences of points t_j, u_j both converging to t such that t_j is connected to u_j by a non-trivial rule. By Lemma 8.1, this implies $Q(t_j) + Q(u_j) = k\pi$ for some integer k. By continuity, this implies $Q(t) = \frac{k}{2}\pi$. The result then follows from Lemma 8.1. ■

As much of our previous analysis depended on knowledge of the derivative of u, we compute it again in this restricted case. As it turns out, the assumption of circular boundary simplifies the calculation significantly. If we implicitly differentiate either (8.3) or (8.4) with respect to t, we see that

$$u'(t) = \frac{\theta'(t) - 1}{1 - \theta'(u)} = -\frac{Q'(t)}{Q'(u)}. \tag{8.5}$$

In particular, these implicit plots fail to be graphs over the t-axis precisely when either $Q'(t) = 0$ or $Q'(u) = 0$. The presence of such points is a necessary condition for obstructions to the existence of a monotone function $u(t)$. The next Theorem follows immediately from (8.5):

Theorem 8.5. *Let (D, I, γ) be a set of occlusion data with β a circle. Suppose $Q'(t) \neq 0$ for all t. If u is a solution to* (8.3) *or* (8.4)*, then u is monotone decreasing.*

Of primary importance is the degree of the map $Q\colon \mathbb{S}^1 \to \mathbb{S}^1$. Since connections are made by implicitly solving the equation

$$Q(t) + Q(u) = \pi \pmod{2\pi} \tag{8.6}$$

the size of the set $\mathscr{A}(\gamma, t)$ is intimately related to deg Q.

For the occlusion problem, the degree of Q is directly related to the critical point theory of the intensity function $I(x, y)$. If the occlusion is completely non-degenerate, then there is a well-defined map

$$I_\theta = \frac{\nabla I}{|\nabla I|}\colon \partial D \cong \mathbb{S}^1 \to \mathbb{S}^1.$$

The continuous lift θ is then given by $I_\theta + \pi/2$. Hence

$$\deg Q = \deg I_\theta - 1. \tag{8.7}$$

Theorem 8.6. *Let (D, I, γ) be a set of occlusion data with β a circle. Then a necessary condition for the existence of non-Legendrian solitary points is that deg $Q = 0$.*

Proof: Non-Legendrian solitary points do not occur in isolation but as open sets by Lemma 4.3. Since Q is continuous, the presence of non-Legendrian solitary points implies, using (8.3) and (8.4), that the image of Q is contained in a narrow (width $< \pi$) band. In particular, this implies that deg $Q = 0$.

■

Remark 8.7. Part of the content of this Theorem is that if deg $Q \neq 0$ then the obstruction of Theorem 6.3 does not occur. To explore this condition further, we shall suppose that $I(x, y)$ has at most one (nearby) critical point p which if it exists is contained in the interior of the occluded region D. In this instance it follows from the definitions that

$$\deg Q = \text{index}_p \nabla I - 1$$

if p exists. If there is no critical point in the interior then I_θ extends to a continuous function on the interior disc. Standard results in algebraic topology then imply that deg $I_\theta = 0$ and so deg $Q = -1$.

Theorem 8.8. *Let (D, I, γ) be a set of occlusion data with β a circle. If deg $Q = d \neq 0$ then there are at least $|d|$ solutions to* (8.3) *or* (8.4).

Proof: If deg $Q = d$ then for a regular value, t_0, of Q, we have that $\#Q^{-1}(t_0) \geq |d|$. So, the equation $Q(u)+Q(t) = (2k+1)\pi$ may be rewritten as $Q(u) = 2(k+1)\pi - Q(t)$ and $Q(u) + Q(t) = 2k\pi$ may be rewritten as $Q(u) = 2k\pi - Q(t)$. For a fixed t and k so that $k\pi - Q(t) \in [0, 2\pi)$ and is a regular value of Q, we then have at least $|d|$ values for u which satisfy one of these equation. Since, by the implicit function theorem, any solution u is C^1, we see that there are at least $|d|$ solutions to (8.3) and (8.4). ■

Lemma 8.3. *If $Q' < 0$ everywhere then the limiting rules at any Legendrian point are external to the occluded disc.*

Proof: Suppose not. Then without loss of generality we may rotate and reflect the image data to match Figure 4, where we are assuming that the curves are oriented to the counter-clockwise direction. Elementary arguments then show that the angles marked are indeed $Q(t) - \pi/2$ and $\pi/2 - Q(u)$ and that these must therefore both be positive. However at the Legendrian point we must have $Q = \pi/2$. This clearly violates the condition that $Q' < 0$.

■

Remark 8.9. We remark that if Q' is positive at a Legendrian point, then the circle tangent to the occlusion boundary may indeed lie inside the occluded circle.

Theorem 8.10. *Let $I : \mathbb{R}^2 \to \mathbb{R}$ be a C^2 function and D be a circular occlusion domain. Further, suppose $\gamma \in \mathcal{RT}$ is the θ lift of ∂D and that the occlusion is completely nondegenerate and occludes no critical points of I. If $Q'(t) \neq 0$ for $t \in [0, 2\pi)$ then there exists a minimal spanning surface of γ which maps to D under*

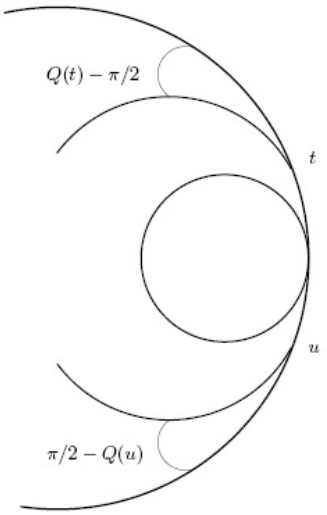


Fig. 4. Legendre Gap.

the map Π *and the image of* Π *lies entirely inside* D*. Further, an image disocclusion derived from the minimal surface exhibits no backtracking. If, in addition, we assume that* $Q' < 0$*, then* Π *maps onto* D*.*

Proof: If the occlusion is completely nondegenerate and occludes no critical points then (8.7) shows that *deg* $Q = -1$. Thus, by Theorem 8.8, there exists a solution u to $Q(t) + Q(u) = (2k+1)\pi$ or $Q(t) + Q(u) = 2k\pi$. Since $Q' \neq 0$, we have that the function $u(t)$ is monotone by (8.5). Thus, $u : \mathbb{S}^1 \to \mathbb{S}^1$ is one to one and onto and, for each t, there is a unique point connecting to to t given by $u(t)$. The lack of backtracking is a direct consequence of Proposition 7.4. Further, $Q' \neq 0$ also implies that every Legendrian point is isolated. The matrix equation (6.7) determines the rules joining t to $u(t)$ and the definition of $Rule(t, u(t))(r)$ forces $\frac{\partial}{\partial r} Rule(t, u(t))(r)$ to point into the interior of D. Since ∂D is a circle and two distinct circles may intersect in at most two points, we see that $\Pi(t, u)|_{[0,1]}$ lies entirely inside $\overline{D}$.

To show the surjectivity under the assumption that $Q' < 0$, we wish to use Proposition 7.6. Consider a nonLegendrian point on γ and let $I = (t_1, t_2)$ be the maximal interval so that $\gamma(I)$ contains this point and also does not contain any Legendrian points. Let $J = u(I) = (u(t_2), u(t_1))$ (the order follows from monotonicity of u). We claim that $I \cap J = \emptyset$. To see this, suppose there is a parameter value in the intersection. Then, we are in one of two cases: that there exists $t_3 \in [u(t_2), t_2)$ or $t_4 \in (t_1, u(t_1)]$ (here we consider the intervals as subsets of $\mathbb{S}^1$). In either case, the fact that u is monotone decreasing implies that there exists a parameter value t_5 so that $u(t_5) = t_5$. By Corollary 6.9, we know that $\gamma(t_5)$ is either Legendrian or that $\dot{u}(t_5) = 1$. By construction of I and the fact that $\dot{u} < 0$, we reach a contradiction.

Now, we may repeat this construction, creating intervals $\{I_1, \ldots, I_k\}$ and $J_l = u(I_l)$ so that $\bigcup_l I_l \cup J_l$ is the complement of the parameter values of Legendrian points in $[0, 2\pi)$. Now, by construction and the discussion above, each of the pairs I_l, J_l satisfy the hypotheses of Proposition 7.6 and hence $\Pi|_{I_l \times [0,1]}$ maps surjectively onto a region $D_l \subset \bar{D}$ bounded by $\beta|_{\bar{I}_l}, \beta|_{\bar{J}_l}$ and $\Pi|_{\partial I \times [0,1]}$. Letting $D_0 = \cup_{i=1}^{k} D_i$, we have that ∂D_0 is the union of rules passing through Legendrian points. We examine two cases. First, if we have a Legendrian point corresponding to a parameter t_6 with $u(t_6) \neq t_6$ then Corollary 8.2 implies that $\gamma(u(t_6))$ is also Legendrian and $Rule(t_6, u(t_6)) \subset \beta$. Second, if we have a Legendrian point corresponding to a parameter t_6 with $u(t_6) = t_6$, then the further assumption that $Q'(t) < 0$ shows that, by Lemma 8.3, $Rule(t_6, t_6) \cap \beta = \{\beta(t_6)\}$. Thus, in either case, the rules composing ∂D_0 are subsets of β. Hence, we conclude that Π is surjective. ∎

9. Examples

Next, we will present a series of explicit examples designed to show that if the behavior of I near the occluded region is relatively tame, then we can construct a minimal surface that spans the occluded region. All of the remaining examples presented show different types of pathology:

(a) Case 2 shows the simplest type of failure of a spanning surface to be a graph: when Q' has zeros. In this case, we have "backtracking" of rules which causes the resulting surface derived from the implicitly defined function u to give an immersed rather than embedded surface.
(b) Case 3 shows an instance where no smooth minimal surface exists due to the presence of non-Legendrian solitary points. As pointed out above, this behavior seems to be characteristic of occluded critical points of I. Moreover, this behavior further indicates that restricting to smooth spanning surfaces, while computationally effective, will not solve any possible minimal surface problem with Dirichlet conditions.
(c) Case 4 shows that with higher degree there are potentially nonuniqueness issues.

The pathologies outlined above, coupled with the discussion of an example using connections from γ to $\overline{\gamma}$, lead us to several conclusions. First, the restriction to smooth spanning surfaces, while sufficient for many types of problems (such as those of Theorem 8.10) is likely insufficient for more complicated areas of an image. Second, the different types of pathologies suggest that to amodally complete a given occlusion, the "best" completion is likely to come from knitting together various pieces of several different solutions (i.e. from different branches of the curve defining u or from pieces connecting γ to $\overline{\gamma}$). Again, this points towards the neccesity of a more sophisticated mechanism. However, we point out that this is consistent with the simulation data found by Citti and Sarti[7] showing that several different possible completions are present at the same time in $\mathcal{RT}$ after using their diffusion method. If this model of minimal surface completion is accurate in reflecting the completion

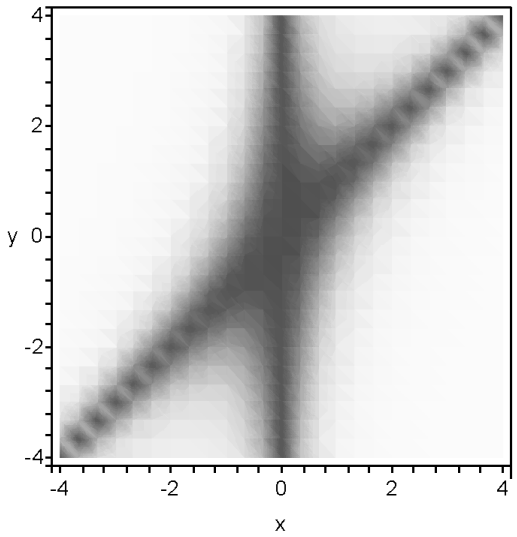


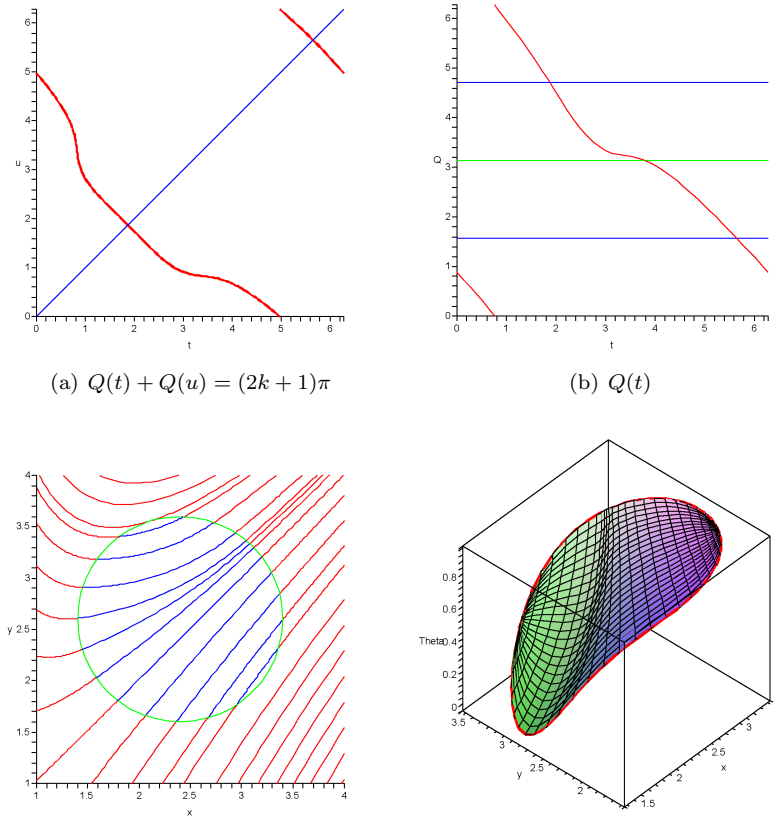
Fig. 5. Intensity plot for $I = (1 + x^2(x - y)^2)^{-1}$.

mechanism in V1, this ambiguity stemming from multiple (partial) solutions may be resolved by the input and feedback from other layers fo the visual cortex. In particular, we note that our algorithm often produces connections between level sets of different "heights" thus creating a completion which is not ideal from the point of view of matching like intensities within the image. This is consistent with the model of the visual cortex present in section 2 as the representation of the image in $\mathcal{RT}$ does not carry information about the intensity of the image, but only information about the level sets of the image itself. One expects that with additional input such as color/intensity information, the best possible completion could be picked out of the possibilities.

In these examples we will deal primarily with completely nondegenerate curves.

9.1. *Case 1: deg* $Q = -1$, $Q' \neq 0$

As a basic example, we consider the following example: set the intensity function as $I(x, y) = (1+x^2(y-x)^2)^{-1}$ representing the image of an angled cross, see Figure 5. The region to be occluded is $(x-2.4)^2+(y-2.6)^2 < 1$, which is over just one branch of the cross. While formally I has critical points everywhere along the lines $y = x$ and $x = 0$, we can replace I with $x(y-x)$ without altering the underlying structure of the level sets. With this simplification, we have a completely non-degenerate occlusion with no occluded critical points. Thus deg $Q = -1$ (and hence surjective onto $\mathbb{S}^1$). The implicit plot of (**??**) is shown in Figure 6. In this instance we see that Q is one-to-one from $\mathbb{S}^1$ to $\mathbb{S}^1$ and so there is only one branch of (**??**), which spans



(a) $Q(t) + Q(u) = (2k+1)\pi$ (b) $Q(t)$

(c) Contour Completion (d) Minimal Spanning Surface

Fig. 6. $I = (1 + x^2(x-y)^2)^{-1}$, Center=$(2.4, 2.6)$, Radius=1, Connecting γ to γ.

the entire range. Moreover, direct calculation shows that $Q'(t) \neq 0$ for all $t \in [0, 2\pi]$. The boundary lift γ has exactly 2 Legendrian points, represented in Figure 6 (b) by the intersection of the curve $Q(t)$ with the blue lines $Q = \pi/2$ and $Q = 3\pi/2$. The intersections with the lines $Q = 0, 2\pi$ and $Q = \pi$ correspond to the orthogonal points.

Since Legendrian points can only connect to other Legendrian points (by Corollary 8.2), we see from the implicit plot of (**??**) that the Legendrian points for γ are solitary and correspond to the intersections with the leading diagonal in Figure 6 (a). Thus we can pick either and construct connecting rules by tracing the sole branch of the implicit plot until we reach the other Legendrian point. By symmetry every point on the curve has now been connected to another and we can build a surface ruled by the ∇-geodesic segments that project into the interior of the occluded region. The projections of these segments provide a contour completion through

the occluded region. In Figure 6 (c) and (d) we show the full contour completion together with the associated minimal lift in the roto-translation group. We note that it is possible to show directly that the surface constructed is a graph over the occluded region.

9.2. *Case 2: deg $Q = -1$, Q' has zeros*

The last example worked so well because not only was deg $Q = -1$, but also because Q was injective. This was because the directions of the underlying contours were relatively uniform. If we move the occluded region in closer to the center of the cross, we lose this uniformity and we find that Q depsite having degree -1 is no longer injective. Examining Figure 7 confirms that there are points where $Q' = 0$ (this can also be confirmed by direct calculation). This is represented by the failure of the implicit plot of (**??**) to be a graph over either u or t. When we follow the program laid out earlier for constructing minimal spanning surfaces, we find that some points have multiple connections. The surface then connects to some parts of the curve as a ridge. Note this surface has the "backtracking" feature of Proposition 7.4 as $\dot{u}$ takes some positive values.

9.3. *Case 3: deg $Q = 0$*

If deg Q = 0, Theorem 6.3 and Theorem 8.6 show that there are possible obstructions to even local existence of spanning surfaces for γ. We note that in the situation of a single critcal point of I being occluded, the index of ∇I must be 1. In other words, we must be occluding a local maximum or minimum.

A simple example is to consider the circular lift

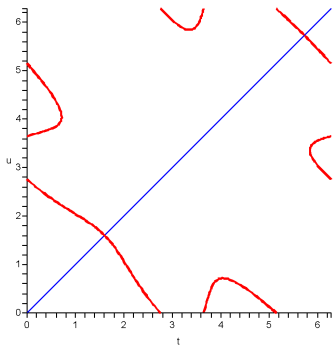
$$\gamma(t) = (\cos t, \sin t, t). \tag{9.1}$$

Here $Q(t) = 0$ everywhere and so every point is orthogonal with outward pointing orientation. The set of (t, u) that satisfy (**??**) is therefore empty. Every point is therefore solitary and non-Legendrian and there is no non-characteristic minimal surface that spans even a part of γ.

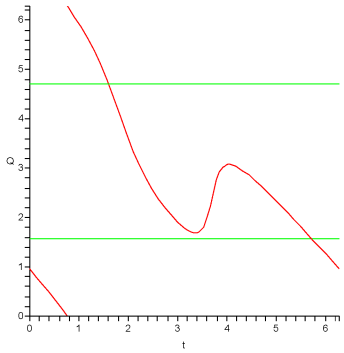
It is clear that the lift (9.1) cannot occur from an occlusion problem as it would require the vector field ∇I to be rotational and hence non-conservative. However gaps in the implicit plot are characteristic of occluded maxima and minima, at least in the absence of symmetry. See Figure 8 for an explicit example where $I = (1+x^2+0.9y^2)^{-1}$ and the circle occludes a local maximum of the function. In figure 8 (a), we see that there are two gaps where there are no connections between $\gamma(t)$ and any other point on the curve. The nature of the gap as part of the minimal spanning surface is shown in the remaining graphs.

9.4. *Case 4: $|\deg Q| > 1$*

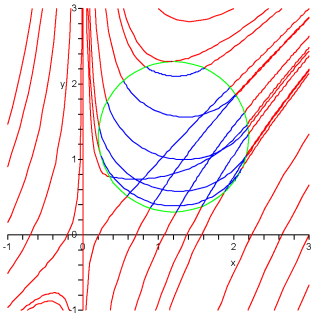
When Q has large degree, the phenomena of overlapping contours and immersed, discontinuous spanning surfaces occurs naturally even when we are considering only



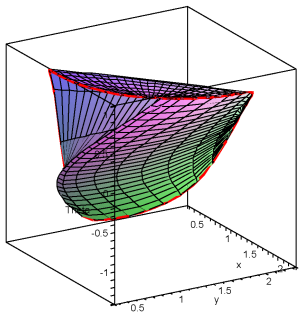
(a) $Q(t) + Q(u) = (2k+1)\pi$



(b) $Q(t)$



(c) Contour Completion



(d) Minimal Spanning Surface

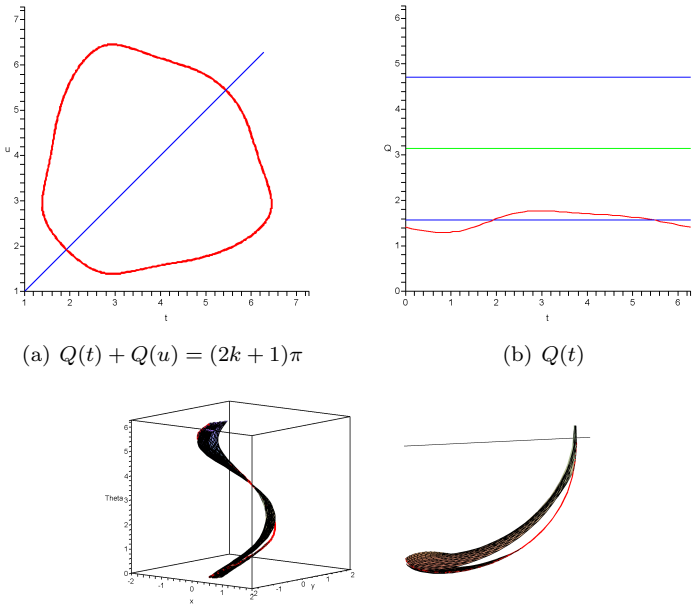
Fig. 7. $I = (1 + x^2(x-y)^2)^{-1}$, Center=$(1.2, 1, 4)$, Radius=1, Connecting γ to γ.

rules connecting γ to itself. In Figure 9 we return to the intensity function $I = (1 + x^2(x-y)^2)^{-1}$, but with the occluded region shifted to have center $(0.1, -0.3)$ and radius 1. Since we are now occluding the saddle point (of $x(y-x)$) at $(0,0)$, the degree of Q is -2.

In this instance we see that Q is now a monotone two-to-one function from $\mathbb{S}^1$ to $\mathbb{S}^1$. This corresponds to there now being two branches of the implicit plot of (**??**) in Figure 9 (a). For convenience of reference we shall refer to the highlighted branch as branch I and the other as branch II. In both contour completions there is overlap as Legendrian points are crossed while transversing the branches.

9.5. *Connecting γ to $\overline{\gamma}$*

For the problem of image disocclusion, we may consider also rules connecting the lift γ to its conjugate lift $\overline{\gamma}$. To illustrate this we return to our original example with



(a) $Q(t) + Q(u) = (2k+1)\pi$ (b) $Q(t)$

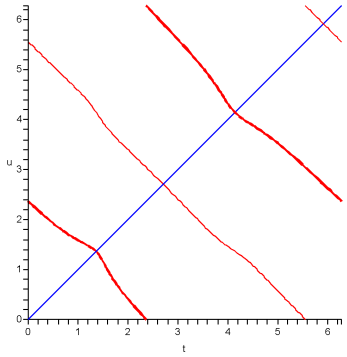
(c) Contour Completion (d) Close Up

Fig. 8. $I = (1 + x^2 + 0.9y^2)^{-1}$, Center=$(0.1, -0.3)$, Radius=1, Connecting γ to γ.

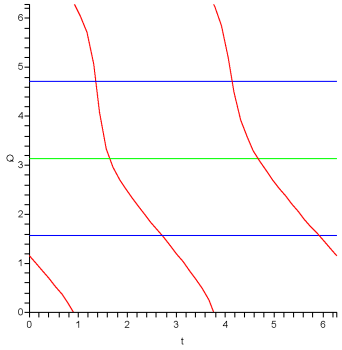
the occluded region being the unit disk centred at $(2.4, 2.6)$. We follow the same basic program, but instead focus on the orthogonal points as our start and finish locations. Unfortunately, this introduces a pathology into the construction of our contour completion and minimal spanning surfaces. To progress from one orthogonal point to the other along a branch of the implicit plot of (8.4) it is necessary to pass through a Legendrian point. The effect of this is to switch which segment of the connecting rules projects to the interior of the occluded region. As is seen in Figure 10 (b) this causes overlaps of the contour completion. If we look at the minimal surface formed from these internal rule segments, in order to completely span γ we must continue further along the branch until we return to the original orthogonal point. This produces the discontinuous self-intersecting surface of Figure 10 (c) . From the perspective of minimal spanning surfaces, it is more natural to allow rule segments that project outside the occluded region. As is shown in Figure 10 (d), this yields a smooth immersed surface between γ and $\overline{\gamma}$, but it still self-intersects.

10. Implementation

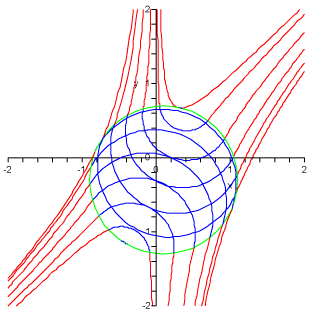
In this section, we present the results (in figures 11 - 15) of a implementation of the scheme described previously to the problem of disocclusion for circular occlusion of digital images. We begin by summarizing the procedure in this special case:



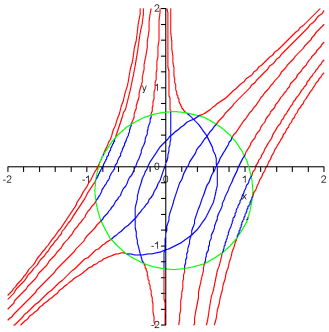
(a) $Q(t) + Q(u) = (2k+1)\pi$



(b) $Q(t)$



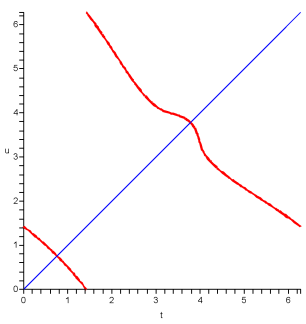
(c) Contour Completion I



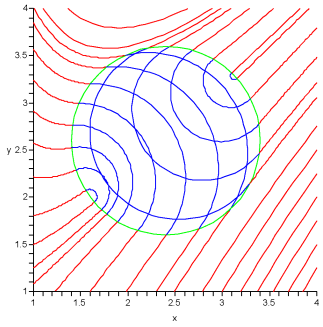
(d) Contour Completion II

Fig. 9. $I = (1 + x^2(x-y)^2)^{-1}$, Center=$(0.1, -0.3)$, Radius=1, Connecting γ to γ.

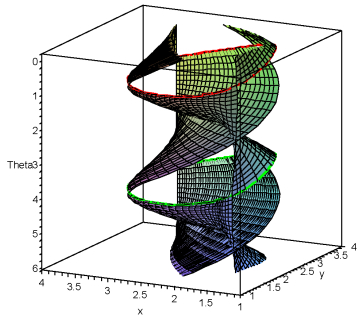
(a) Consider a greyscale image that was represented by a smooth function, I. We note that a standard digital image could be used if it were first smoothed via a Gaussian convolution.
(b) Fix a circle β bounding an occlusion domain D.
(c) Compute θ, β.
(d) Compute Q' and deg Q and determine suitability of the algorithm. We require $Q' \neq 0$ and deg $Q \neq 0$. If $|\deg Q| > 1$, we may repeat this procedure for all possible solutions u.
(e) For each t, test each possible $u(t)$ for backtracking and the failure of local diffeomorphism.
(f) If the algorithm is suitable, construct all rules, $Rule(t, u(t))$, using the data from the function u and the formulae in Theorem 4.4.
(g) For each solution u, with a set of identified values t from step (e) of incompat-



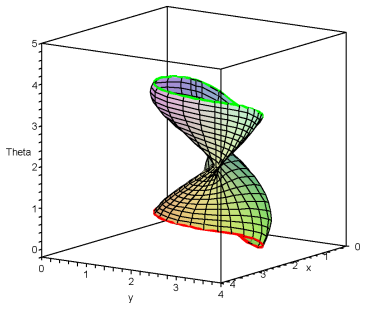
(a) $Q(t) + Q(u) = 2k\pi$



(b) Contour Completion



(c) Minimal Spanning Surface I



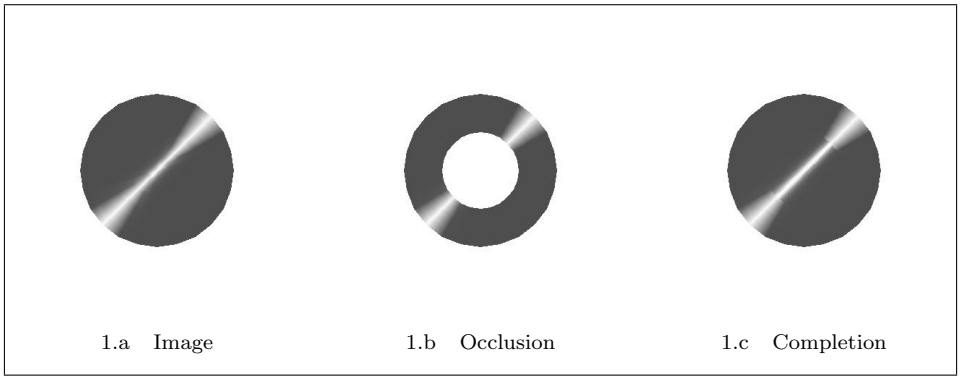
(d) Minimal Spanning Surface II

Fig. 10. $I = (1 + x^2(x - y)^2)^{-1}$, Center=$(2.4, 2.6)$, Radius=1, Connecting γ to $\overline{\gamma}$.

ibility, choose functions f_t and F as in Definition 7.2.

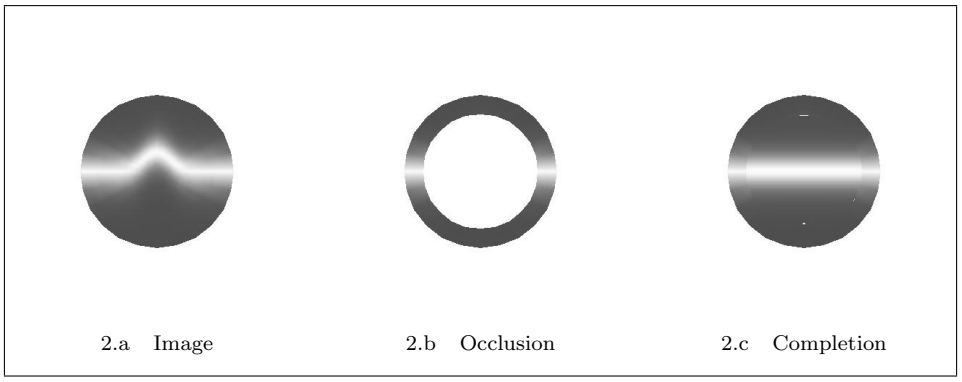
(h) Construct an image disocclusion using Definition 7.2 and the map Π.

We note that if, in step d, the algorithm is found to be suitable, it will always produce an image and steps (c-f) have a combined computational complexity which is linear in area of the occluded region. We note that if the algorithm is unsuitable (e.g. if deg $Q = 0$ or Q' has regions of both signs), then the ambiguity inherent in the completion requires additional decision to provide for complete automation (steps (g-h)). In the figures that follow, we choose f_t to be a linear interpolation of the the values of I at the endpoints. This choice is not crucial and many other plausible choices exist. We emphasize that, in this implementation, we do not require that the rules connect point with a shared intensity value. This is in keeping with a consequence of the Citti-Sarti model, that only level sets of the intensity function are relevant to the model. As, once the image is represented as a surface in $\mathcal{RT}$,



1.a Image 1.b Occlusion 1.c Completion

Fig. 11. Occlusion of a simple linear image



2.a Image 2.b Occlusion 2.c Completion

Fig. 12. Occlusion of a complicated portion of an image.

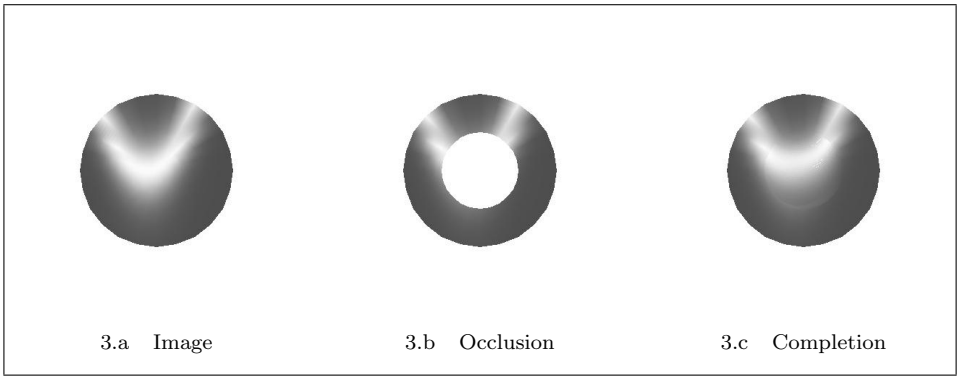
there is no way to determine if two points lie on the same level set with global information, we allow the completion algorithm to make a "mistake" of this type as it thus better reflects the qualities of the model. We note that if we impose the requirement that intensities match, the algorithm fails much more often. Last, we only present examples where there is a unique point associated to each point in D so there is not a need to specify F.

Acknowledgment

Both authors are partially supported by NSF grant DMS-0306752.

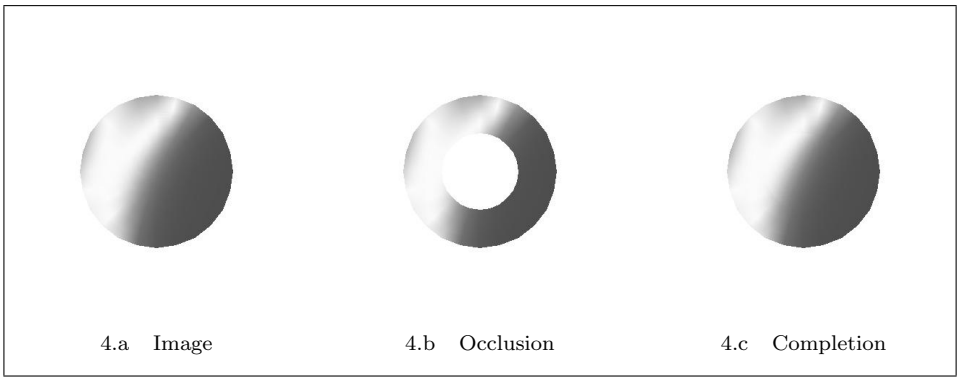
References

1. L. Ambrosio and S. Masnou. A direct variational approach to a problem arising in image reconstruction. *Interfaces and Free Boundaries*. To appear.
2. O. Ben-Shahar and S.W. Zucker The Perceptual Organization of Texture Flow: A



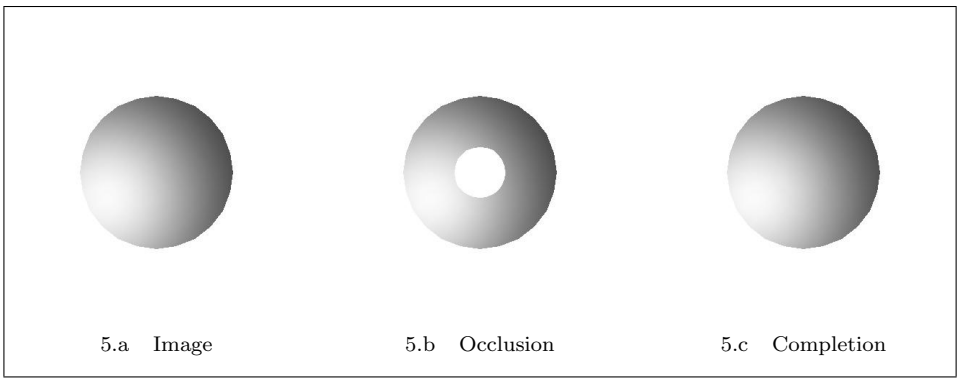
3.a Image 3.b Occlusion 3.c Completion

Fig. 13. Occlusion of a curve



4.a Image 4.b Occlusion 4.c Completion

Fig. 14. Occlusion of a convcave region



5.a Image 5.b Occlusion 5.c Completion

Fig. 15. Occlusion of a convex region

Contextual Inference Approach *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 25(4):401-417, 2003.
3. O. Ben-Shahar and S.W. Zucker Hue Geometry and Horizontal Connections *Neural Networks*, Special Issue on Vision and Brain, To appear.
4. Jih-Hsin Cheng and Jenn-Fang Hwang. Properly embedded and immersed minimal surfaces in the Heisenberg group. *Bull. Austral. Math. Soc.* 70 (2004), no. 3, 507–520
5. Jih-Hsin Cheng, Jenn-Fang Hwang, Andrea Malchiodi, and Paul Yang. Minimal surfaces in pseudohermitian geometry. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) 4 (2005), no. 1, 129–177
6. G. Citti, M. Manfredini, and A. Sarti. Neuronal oscillation in the visual cortex: Gamma-convergence to the Riemannian Mumford-Shah functional. *SIAM Jornal of Mathematical Analysis*, 35(6):1394 – 1419, 2003.
7. G. Citti and A. Sarti. A cortical based model of perceptual completion in the roto-translation space. 2004. Preprint.
8. Daniel Cole. *On minimal surfaces in Martinet-type spaces.* PhD thesis, Dartmouth College, 2005.
9. D. Danielli, N. Garofalo, and D.-M. Nhieu. Minimal surfaces, surfaces of constant mean curvature and isoperimetry in Carnot groups. August, 2001. Preprint.
10. H. Farid and E. P. Simoncelli. Differentiation of Multi-Dimensional Signals. *IEEE Transactions on Image Processing*, 13(4):496–508, 2004.
11. Field, Hess and Hayes Contour Integration by the Human Visual System: Evidence for a Local 'Association Field. *Vision Research* 33-2, 173-193, 1993.
12. Nicola Garofalo and Duy-Minh Nhieu. Isoperimetric and Sobolev inequalities for Carnot-Carathéodory spaces and the existence of minimal surfaces. *Comm. Pure Appl. Math.*, 49(10):1081–1144, 1996.
13. Nicola Garofalo and Scott D. Pauls. The Bernstein problem in the Heisenberg group. 2003. Submitted.
14. C.D. Gilbert, A. Das, M. Ito, and G. Westheimer. Spatial integration and cortical dynamics. *Proceedings of the National Academy of Sciences USA*, 93:615–622.
15. Robert Hladky and Scott D. Pauls. A disocclusion algorithm based on a model of the visual cortex. 2005. In preparation.
16. Parent, P., and Zucker, S.W. Trace Inference, Curvature Consistency, and Curve Detection *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(8): 823-839, 1989.
17. Robert K. Hladky and Scott D. Pauls. Constant mean curvature surfaces in sub-riemannian spaces. 2005. Preprint.
18. William C. Hoffman. The visual cortex is a contact bundle. *Appl. Math. Comput.*, 32(2-3):137–167, 1989. Mathematical biology.
19. D. H. Hubel and T. N. Weisel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160:106–154, 1962.
20. D. H. Hubel and T. N. Weisel. Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. London (Biol)*, 198:1–59, 1977.
21. S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry.* John Wiley & Sons, Inc., 1963.
22. G. P. Leonardi and S. Masnou. On the isoperimetric problem in the Heisenberg group $\mathbf{H^n}$. Preprint, 2002.
23. B. Kimia, I. Frankel and A. Popescu Euler Spiral for Shape Completion *International Journal of Computer Vision*, 54(1/2/3), 159182, 2003
24. G. P. Leonardi and S. Rigot. Isoperimetric sets on Carnot groups. *Houston J. Math.*, 29(3):609–637 (electronic), 2003.
25. M. Nitzberg, D. Mumford, and T. Shiota. *Filtering, segmentation and depth*, volume

662 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1993.
26. Scott D. Pauls. H-minimal graphs of low regularity in the Heisenberg group. *Comm. Math. Helv.* to appear.
27. Scott D. Pauls. Minimal surfaces in the Heisenberg group. *Geom. Ded.*, 104:201–231, 2004.
28. J. Petitot. The neurogeometry of pinwheels as a sub-Riemannian contact structure. *J. Physiology*, 97:265–309, 2003.
29. J. Petitot and Y. Tondut. Vers une neuro-geometrie. fibrations corticales, structures de contact et contours subjectifs modaux. *Mathematiques, Informatique et Sciences Humaine, EHESS, Paris*, 145:5–101, 1998.
30. Manuel Ritoré and César Rosales. Rotationally invariant hypersurfaces with constant mean curvature in the Heisenberg group $\mathbb{H}^n$. 2005. Preprint.
31. M. Sigman, G. Cecchi, C. Gilbert and M. Magnasco On a common circle: Natural scenes and Gestalt rules *PNAS* 98(4):1935-1940, 2001.
32. N. Tanaka. *A differential geometric study on strongly pseudoconvex manifolds*. Kinokuniya Book-Store Co., Ltd., 1975.
33. S.M. Webster. Pseudo-Hermitian structures on a real hypersurface. *J. Differential Geometry*, 13:25–41, 1978.
34. M. Wertheimer. Untersuchungen zur Lehre von der Gestalt *Psychologische Forschung* 4:301–350 ,1923. (condensed and translated as "Laws of organization in perceptual forms", in A Source Book of Gestalt Psychology Ed.W D Ellis (1938, New York: Harcourt Brace) pp 71–88).
35. M. Wertheimer. Ueber Gestalttheorie, Lecture before the Kant Gesellschaft. (reprinted in translation in A Source Book of Gestalt Psychology Ed.W D Ellis (1938, New York: Harcourt, Brace) pp 1–11).
36. Z. Zhou, J. Samonds, M. Bernard and A. Bonds Synchronous activity in cat visual cortex encodes collinear and cocircular contours *J. Vision*, 5(8):675a, 2005