Finding Shortest Paths on Surfaces Using Level Sets Propagation

Ron Kimmel, Arnon Amir, and Alfred M. Bruckstein

Abstract—**We present a new algorithm for determining minimal length paths between two regions on a three dimensional surface. The numerical implementation is based on finding equal geodesic distance contours from a given area. These contours are calculated as zero sets of a bivariate function designed to evolve so as to track the equal distance curves on the given surface. The algorithm produces all paths of minimal length between the source and destination areas on the surface given as height values on a rectangular grid.**

Index Terms—**Curve evolution, equal distance contours, geodesic path, numerical algorithms, minimal geodesics.**

I. Introduction

Finding paths of minimal length between two areas on a three dimensional surface is of great importance in many fields such as computer-aided neuroanatomy, robotic motion planning (autonomous vehicle navigation), geophysics, terrain navigation, etc. Paths of minimal Euclidean distance between two points on a surface are usually referred to as *minimal geodesics*.

A new approach for dealing with the problem of finding the minimal distance paths, in which the surface is given as height samples on a rectangular grid, is introduced. As a first step, a distance map from the source area is calculated. The distance map is computed via *equal distance curve* propagation on the surface. Equal distance curves are calculated as the zero sets of a bivariate function evolving in time. This formulation of curve evolution processes is due to Osher and Sethian, [14], [17]. It overcomes some topological and numerical problems encountered in direct implementations of curve evolutions using parametric representations. The implicit representation of the evolving curve produces a stable and accurate numerical scheme for tracing shock waves in fluid dynamics.

The proposed numerical scheme is consistent with the continuous propagation rule. The consistency condition guarantees that the solution converges to the true one as the grid is refined and the time step in the numerical scheme is kept in the right proportion to the grid size. This is known not to be the case in general graph search algorithms that suffer from *digitization bias* due to the *metrication error* when implemented on a grid [10], [9].

The relation between minimal paths, geodesics and equal distance contours may be found in elementary differential geometry textbooks, e.g., [2]. Geodesics are locally shortest paths in the sense that any perturbation of a geodesic curve will increase its length. The minimal length paths between two points are the *minimal geodesics* connecting those points. A simple way of determining minimal geodesics is by constructing a so–called *geodesic polar coordinate system* on the surface around the source area. Using such a coordinate system readily provides the *geodesic circle map*, or the map of equal distance contours on the surface.

In the next section an analytic model for the equal distance contour evolution is discussed. In Section III, a numerical implementation of the analytic propagation is presented. The results of the numerical algorithm are demonstrated by an example in Section IV. We conclude with a discussion of some possible extensions of the algorithm and comment on its complexity in Section V.

Manuscript received Jan. 27, 1993; revised Mar. 13, 1995.

R. Kimmel is with the Dept. of Electrical Engineering, Technion-I.I.T., Haifa 32000, Israel; e-mail: ron@techunix.technion.ac.il.

A. Amir and A.M. Bruckstein are with the Dept. of Computer Science, Technion-I.I.T., Haifa 32000, Israel.

IEEECS Log Number P95067.

II. Equal Geodesic–Distance Contour Propagation

Let us first define a differential equation describing the propagation of equal geodesic–distance contours on a smooth surface starting from a point or a source region on the surface. Given a source area $S \in \mathbf{R}^3$ (S is not necessarily connected), on a surface $S \in \mathbf{R}^3$, let the $3D$ equal distance contour of distance t from S be defined as

$$\{\mathbf{p} \in S \mid d_S(p, S) = t\} = \alpha(*, t),$$

where $d_S(\mathbf{p}, S)$ is the *minimal geodesic distance* determined by the the shortest paths from a point $\mathbf{p}$ to an area S on the surface S.

We shall prove that the $3D$ parametric representations of $\alpha(*, t)$, on S, can be obtained by the equal distance contour propagation

$$\alpha_t = N \times \vec{\mathbf{t}}^{\alpha}, \quad \text{given} \quad \alpha(u, 0) = \alpha(u), \tag{1}$$

where $\vec{\mathbf{t}}^{\alpha}$ is the tangent unit vector to α, and N is the surface normal.

The traces of constant parameter along the curve evolving according to (1) are geodesics, and these geodesics are locally shortest paths. We have the following results:

LEMMA 1. *Define the curve* $\beta(t) = \alpha(u, t)|_{u=u_0}$. *Then, for any* u_0, *the curve* $\beta(t)$ *is a geodesic.*

PROOF. The trace $\beta(t)$ is determined by the evolution of $\alpha_t = N \times \vec{\mathbf{t}}^{\alpha}$, hence $\beta_t = N \times \vec{\mathbf{t}}^{\alpha}$. Since $|\beta_t| = |N \times \vec{\mathbf{t}}^{\alpha}| = 1$, the t parameter is the arclength of β.

In order to show that β is a geodesic we recall the definition of geodesics and prove that

$$\frac{d^2}{dt^2}\beta = \lambda N,$$

where $\lambda(t)$ is a scalar function. A geometric interpretation of this formula is that the second derivative of the curve (its normal direction $k\vec{\mathbf{n}}^{\beta}$) is in the surface normal direction N. To prove the result it is sufficient to verify that

$$\langle \beta_u, \beta_t \rangle = 0 \quad \text{and} \quad \langle \beta_{tt}, \vec{\mathbf{t}}^{\alpha} \rangle = 0,$$

or explicitly that

$$\left\langle \frac{d}{dt}\left[N \times \vec{\mathbf{t}}^{\alpha}\right], N \times \vec{\mathbf{t}}^{\alpha} \right\rangle = 0 \quad \text{and} \quad \left\langle \frac{d}{dt}\left[N \times \vec{\mathbf{t}}^{\alpha}\right], \vec{\mathbf{t}}^{\alpha} \right\rangle = 0$$

which clearly force $\beta_{tt} = \lambda N$. Let us first define k, $\vec{\mathbf{n}}^{\beta}$, and $\vec{\mathbf{t}}^{\beta}$ to be the curvature, normal, and tangent of $\beta(t)$, respectively. Using the Frenet formulas we first have

$$\left\langle \frac{d}{dt}\left[N \times \vec{\mathbf{t}}^{\alpha}\right], N \times \vec{\mathbf{t}}^{\alpha} \right\rangle = \langle \beta_{tt}, \beta_t \rangle = \left\langle k\vec{\mathbf{n}}^{\beta}, \vec{\mathbf{t}}^{\beta} \right\rangle = 0$$

The second expression may be written as follows

$$\left\langle \frac{d}{dt}\left[N \times \vec{\mathbf{t}}^{\alpha}\right], \vec{\mathbf{t}}^{\alpha} \right\rangle = \langle \beta_{tt}, \vec{\mathbf{t}}^{\alpha} \rangle = \left\langle \alpha_{tt}, \frac{\alpha_u}{|\alpha_u|} \right\rangle.$$

Therefore, we should prove that

$$\left\langle \alpha_{tt}, \frac{\alpha_u}{|\alpha_u|} \right\rangle = 0.$$

First note that

$$\frac{d}{dt}\left\langle \alpha_t, \frac{\alpha_u}{|\alpha_u|}\right\rangle = \frac{d}{dt}\left\langle N\times\vec{\mathbf{t}}^{\alpha}, \vec{\mathbf{t}}^{\alpha}\right\rangle = \frac{d}{dt}0 = 0.$$

Using inner product rules we have

$$\frac{d}{dt}\left\langle \alpha_t, \frac{\alpha_u}{|\alpha_u|}\right\rangle = \left\langle \alpha_{ut}, \frac{\alpha_u}{|\alpha_u|}\right\rangle + \left\langle \alpha_t, \left(\frac{\alpha_u}{|\alpha_u|}\right)_t\right\rangle.$$

Therefore, in order to show that

$$\left\langle \alpha_{ut}, \frac{\alpha_u}{|\alpha_u|}\right\rangle = 0,$$

it is enough to show that

$$\left\langle \alpha_t, \left(\frac{\alpha_u}{|\alpha_u|}\right)_t\right\rangle = 0.$$

Define the metric along the curve to be $g \equiv |\alpha_u|$, and compute

$$\frac{d}{dt}\frac{\alpha_u}{|\alpha_u|} = \frac{d}{dt}\vec{\mathbf{t}}^{\alpha} = \frac{\alpha_{ut}g - \alpha_u g_t}{g^2} = \frac{\alpha_{ut}}{g} - \frac{\alpha_u g_t}{g^2}.$$

Using this result we proceed

$$\begin{aligned}\left\langle \alpha_t, \left(\frac{\alpha_u}{|\alpha_u|}\right)_t\right\rangle &= \left\langle \alpha_t, \frac{\alpha_{ut}}{g} - \frac{\alpha_u g_t}{g^2}\right\rangle \\ &= \left\langle \alpha_t, \frac{\alpha_{ut}}{g}\right\rangle - \left\langle \alpha_t, \alpha_u \frac{g_t}{g^2}\right\rangle \\ &= \frac{1}{g}\left\langle \alpha_t, \alpha_{ut}\right\rangle - \frac{g_t}{g^2}\left\langle N\times\vec{\mathbf{t}}^{\alpha}, \vec{\mathbf{t}}^{\alpha}\right\rangle \\ &= \frac{1}{g}\frac{1}{2}\frac{d}{du}\left\langle \alpha_t, \alpha_t\right\rangle - \frac{g_t}{g^2}0 \\ &= \frac{1}{g}\frac{1}{2}\frac{d}{du}1 \\ &= 0.\end{aligned}$$

Therefore, $\left\langle \beta_{tt}, \vec{\mathbf{t}}^{\alpha}\right\rangle = 0$, and this proves that $\beta(t)$ is a geodesic. □

Let $\alpha(u, t)$ be a $3D$ curve propagating on the surface $S \subset R^3$, where u is the parameter and t is the propagation time. Then

LEMMA 2. *The equal distance contour evolution is given by*

$$\alpha_t = N\times\vec{\mathbf{t}}^{\alpha} \quad given \quad \alpha(0).$$

PROOF. As a first step we shall use *Gauss Lemma* to show that the asserted evolution rule formulates *geodesic polar coordinates* when starting from a point. The *geodesic circles* in this coordinates system are the equal distance contours. This result is then generalized to any given initial curve.

Define the tangent plane to a point p on the surface S as $T_p(S)$. Let $\vec{w} \in T_p(S), |\vec{w}| = 1$ be a unit vector indicating a direction from p in the tangent plane. Use a polar coordinate system to define $\vec{w}$ on $T_p(S)$ as $\vec{w}(u) = \sin(u)\hat{x} + \cos(u)\hat{y}$. Define $\gamma^u(t)$ to be the geodesic which starts from p, with $\gamma^u(0) = p$, and $\gamma_t^u(0) = \vec{w}(u)$, where t stands for the arclength.

According to *Gauss Lemma*, see e.g,. [2] p. 287, the radial geodesics $\gamma^u(t)$ together with the geodesic circles $\eta(u, t) = \gamma^u(t)|_{t=\text{const}, u\in S^1}$, form geodesic polar coordinates, in which the geodesic circles are orthogonal to the radial geodesic.

Considering the constant parameter traces along the evolving contours as the radial geodesics $\beta(t) = \gamma^u(t)$, and the contours themselves as the geodesic circles $\alpha(u) = \eta(u)$, we obtain that the equal distance contour evolution rule is given by the asserted equation, when starting from a point. To be more precise, we start from a given infinitesimal geodesic circle around the point.

We now proceed and generalize the result to any given curve $\alpha(u, 0) = \alpha(0)$ on the surface.

Let P be the set of points forming the equal distance contour of distance d from $\alpha(u, 0)$ on the given surface. Propagate an equal distance contour $\eta(v, t)$ starting from any point $p \in P$. Stop the propagation when the equal distance contour first touches $\alpha(u)$, let say at $q = \alpha(u_0) = \eta(v_0, \tau)$. According to the construction $\tau = d$, and therefore, $q = \eta(v_0, d)$. See Fig. 1a.

At q, $\eta(v, d)$ and $\alpha(u)$ osculate, which means that $\vec{\mathbf{t}}^{\alpha}|_{u=u_0}$ is parallel to $\vec{\mathbf{t}}^{\eta}|_{v=v_0}$, see Fig. 1b.

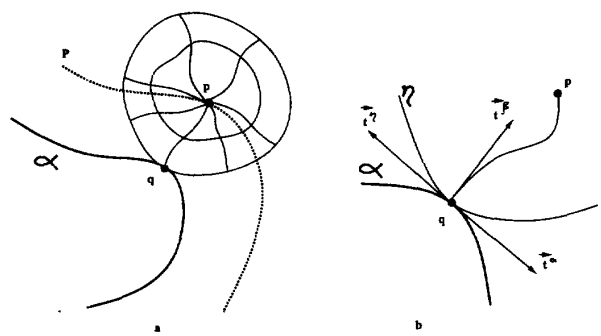

Fig. 1. (a) Equal distance contours are propagated from p, until the contour first touches $\alpha(u)$. (b) The tangent point, observe that at q: $\vec{\mathbf{t}}^{\beta}|_{u=u_0} \| \vec{\mathbf{t}}^{\eta}|_{v=v_0}$.

We have shown earlier that the shortest path from p to q is given by the radial geodesic $\beta(t) = \gamma^{v_0}(t) = \eta(v, t)|_{v=v_0}$, and where $\beta(t)$ and $\eta(v, t)$ are orthogonal along $v = v_0$ (Gauss lemma). Hence, $\vec{\mathbf{t}}^{\beta}|_{t=d} \perp \vec{\mathbf{t}}^{\eta}|_{t=d, v=v_0}$, and therefore,

$$\vec{\mathbf{t}}^{\beta}\Big|_{t=d} \perp \vec{\mathbf{t}}^{\alpha}\Big|_{u=u_0}.$$

We have just proved that the shortest paths from each point in the set P to α, is given by the geodesics starting from $\alpha(u)$ and orthogonal to $\vec{\mathbf{t}}^{\alpha}$. This geodesic is the one obtained via the asserted evolution rule; using the continuity of α, the equal distance contour from $\alpha(u, 0)$, is obtained by the asserted evolution rule, $\alpha_t = N\times\vec{\mathbf{t}}^{\alpha}$. □

Lemma 2 provides the evolution equation of the equal distance contour. Starting from the boundary of the source area $\alpha(0) = \{(x, y, z(x, y))|(x, y, z(x, y)) \in \partial S\}$, it is possible to find the equal distance contour for any desired distance d, by using the evolution equation to calculate $\alpha(u, t)|_{t=d}$. This propagation may be used to build the distance map for each point on the surface.

Implementing the three dimensional curve evolution is quite a complicated task. We are therefore interested in considering the projection of the $3D$ curve on the (x, y)–plane,

$$C(t) = \pi \circ \alpha \equiv \left\{(x, y)\middle|(x, y, z(x, y)) \in \alpha(t)\right\}.$$

A result from the general theory of curve evolution states that the trace of a propagating planar curve may be determined only by its normal velocity [3]. Let us consider the projection of the above evo-

lution on the (x, y)–plane (Fig. 2). The knowledge of how this projected contour behaves allows us to construct a simple, accurate and stable numerical algorithm that can be used to produce these equal distance contours.

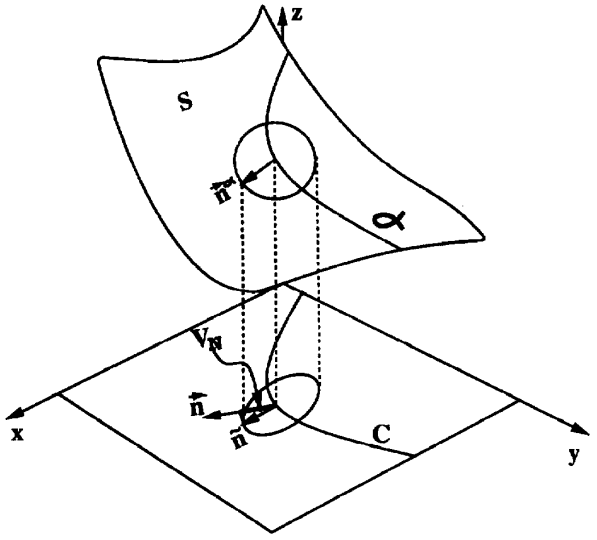

Fig. 2. The projection of an infinitesimal geodesic circle on the surface forms a tilted ellipse on the (x, y)–plane.

In [4] we calculate the planar normal component of the projected velocity of the evolving equal distance contour, $V_N = \langle \vec{n}, \pi \circ (N \times \vec{t}^{\alpha}) \rangle$. Using this velocity we construct a differential equation describing the projected equal distance contour evolution of the form

$$\frac{\partial}{\partial t} C = V_N \vec{n} \quad \text{given} \quad C(0) = \left\{ (x, y) \middle| (x, y, z(x, y)) \in \partial S \right\} \equiv \partial \tilde{S}, \quad (2)$$

where $\vec{n}$ is the planar normal direction, as is the boundary of S, and V_N depends on the surface gradient ($p = \partial z/\partial x$ and $q = \partial z/\partial y$) and $\vec{n}$. Writing the planar normal as its components $\vec{n} = (n_1, n_2)$,

$$V_N = \sqrt{\frac{(1+q^2)n_1^2 + (1+p^2)n_2^2 - (2pq)n_1 n_2}{1+p^2+q^2}}$$

or

$$V_N = \sqrt{an_1^2 + bn_2^2 - cn_1 n_2},$$

where the coefficients a, b, and c depend on the surface gradient and can be computed once at the initialization step.

A. Finding the Minimal Path

The procedure that calculates the equal distance contours allows us to build a Euclidean distance map on the surface, from a given area. Assuming we have reliable distance map procedure in hand, we can construct a simple procedure that finds the minimal path from a source area S to a destination area D (where $S, D \in S$).

Defining $\mathcal{M}_A$ as the distance map of area A as

$$\mathcal{M}_A(x, y) = d_s((x, y, z(x, y)), A).$$

we readily have the following result:

LEMMA 3. *All minimal paths between S and D on S are given by the set $G \subset S$,*

$$G = \{(x, y, z(x, y)) \mid \mathcal{M}_S(x, y) + \mathcal{M}_D(x, y) = g_m\} \quad (3)$$

where $g_m \equiv \min_{(x,y)}(\mathcal{M}_S + \mathcal{M}_D)$ is the global minimum of the sum of the source and destination distance maps.

PROOF.

[$\mathbf{p}_a \in G \Rightarrow \mathbf{p}_a \in$ set of minimal paths] If the point $\mathbf{p}_a$ is in G then $d_s(\mathbf{p}_a, S) + d_s(\mathbf{p}_a, D) = g_m$. Therefore, there exists a path from S to $\mathbf{p}_a$ and from $\mathbf{p}_a$ to D which together form a minimal length path that passes through $\mathbf{p}_a$.

[$\mathbf{p}_a \notin G \Rightarrow \mathbf{p}_a \notin$ any minimal path] If $\mathbf{p}_a \notin G$ then $d_s(\mathbf{p}_a, S) + d_s(\mathbf{p}_a, D) > g_m$. Recalling the d_z definition, all possible paths from S to D which pass through the point $\mathbf{p}_a$ are longer than g_m and, therefore, not minimal. □

Now, we can prove the following result connecting geodesics to the trace of tangential points of the two equal distance contours α_S and α_D, propagating from the source and destination.

LEMMA 4. *The tangential points of $\alpha_S(u, t)$ and $\alpha_D(\tilde{u}, \tilde{t})$ for $\tilde{t} + t = g_m$ generate the minimal paths from point P_1 to point P_2; i.e., lie on a constant parameter $u = u_0(\tilde{u} = \tilde{u}_0)$ of the propagating curve $\alpha_S(u, t)(\alpha_D(\tilde{u}, \tilde{t}))$.*

PROOF. The shortest path built by $d\alpha_S/dt = N \times \vec{t}^{\alpha}$ and tangent at $t = a$ at $u = u_0$ to the path built by $d\alpha_D/d\tilde{t} = N \times \vec{\tilde{t}}$ at $\tilde{t} = b$ at $\tilde{u} = \tilde{u}_0$, is of length $a + b$ (see Lemma 3).

Assume that there is only one minimal path between P_1 and P_2, and assume that the trace of that minimal path passes through two different parameterization points u_0 and u_1. See Fig. 3.

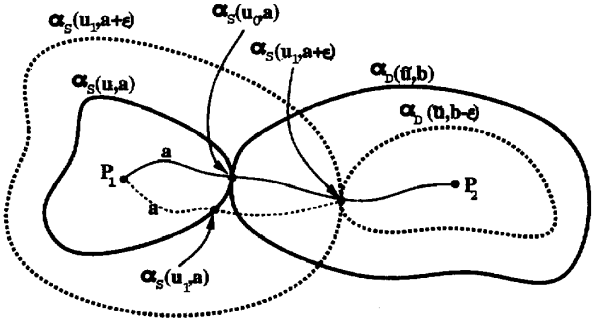

Fig. 3. When considering only one minimal geodesic between two points, the path is the trace of a constant parameter along the evolution, see text.

Let the length of the shortest distance from P_1 to $\alpha_S(a, u_0)$ be a, and the distance from P_2 to $\alpha_D(\tilde{u}_0, b)$ be b. The shortest path is of length $a + b$, and it is given by the path $P_1 \rightsquigarrow \alpha_S(u_0, a) = \alpha_D(\tilde{u}_0, b) \rightsquigarrow P_2$.

Assume there is a parameterization point $u_1 \neq u_0$ through which the minimal path passes at $t = a + \varepsilon$ ($\tilde{t} = b - \varepsilon$). According to that assumption there is a minimal path $P_1 \rightsquigarrow \alpha_S(u_1, a) \rightsquigarrow \alpha_S(u_1, a + \varepsilon) \rightsquigarrow P_2$ of length $a + \varepsilon + (b - \varepsilon)$. However, part of this path, $P_1 \rightsquigarrow \alpha_S(u_1, a)$, is not equal to the original subpath of the minimal path $P_1 \rightsquigarrow \alpha_S(u_0, a)$, and this contradicts the assumption that the minimal path should pass through $\alpha(u_0, a)$, and concludes the proof. □

We also have the following result.

COROLLARY 1. *All minimal paths between S and D which are defined by G* (3), *are minimal geodesics.*

In the next section a numerical scheme based on the level set representation of the evolving planar curve is presented. Note that the shortest paths are minimal value level sets of the function $\mathcal{M}_S + \mathcal{M}_D$. This observation will later be used on to find the minimal paths.

III. THE NUMERICAL APPROXIMATION

When implementing curve evolution equations such as (2) on a digital computer, a number of problems must be solved.

Topological changes may occur while the curve evolves, i.e., the curve may change its topology from one connected curve to two separate evolving curves, or, two curves may merge into one. In [17], [14] some numerical problems which characterize a direct formulation of (2) are described. The problems are caused due to a time varying coordinate system (u, t) of the direct representation (where u is the parameterization, and t - the time). An initial smooth curve can develop curvature singularities. The question is how to continue the evolution after singularities appear. The natural way is to choose the solution which agrees with the *Huygens principle* [16]. Viewing the curve as the front of a burning flame, this solution states that *once a particle is burned, it cannot be reignited* [17]. It can also be proved that from all the *weak* solutions of (2) part the singularities, the one derived from the Huygens principle is unique, and can be obtained by a constraint denoted as the *entropy condition* [14].

Sethian and Osher [17], [14] proposed an algorithm for curve and surface evolution that elegantly solves these problems. As a first step in constructing the algorithm, the curve is embedded in a higher dimensional function. Then, evolution equations for the implicit representation of the curve are solved using numerical techniques derived from hyperbolic conservation laws [12].

A. The Eulerian Formulation

Let the curve $C(t)$ be represented by the zero level set of a smooth Lipschitz continuous function $\phi : \mathbf{R}^2 \times [0, T) \rightarrow \mathbf{R}$, so that ϕ is negative in the interior and positive in the exterior of the zero level set $\phi = 0$. Consider the zero level set defined by $\{X(t) \in \mathbf{R}^2 : \phi(X, t) = 0\}$. We have to find the evolution rule of ϕ, so that the evolving curve $C(t)$ can be represented by the evolving zero level set $X(t)$, i.e., $C(t) \equiv X(t)$. Using the chain rule on $\phi(X(t), t) = 0$ we get $\nabla\phi(X, t) \cdot X_t + \phi_t(X, t) = 0$. Note that for any level set the planar normal can be written as $\vec{\mathbf{n}} = \nabla\phi / \|\nabla\phi\|$. Using this relation in conjunction with the condition equation (2) we obtain

$$\phi_t = -V_N \|\nabla\phi\|, \tag{4}$$

where the curve $C(t)$ is obtained as the zero level set of ϕ. This procedure is known as the Eulerian formulation [17].

This formulation of planar curve evolution processes frees us from the need to take care of the possible topological changes in the propagating curve. The numerical implementation of (4) is based on *monotone* and *conservative* numerical algorithms, derived from hyperbolic conservation laws and the Hamilton–Jacobi "type" equations of the derivatives [14]. For some normal velocities these numerical schemes automatically enforce the *entropy condition*, a condition equivalent to *Huygens principle* [16].

Using the normal component of the velocity, in (4), we get in our case

$$\phi_t = \sqrt{a(x, y)\phi_x^2 + b(x, y)\phi_y^2 - c(x, y)\phi_x\phi_y} \tag{5}$$

This equation describes the propagation rule for the surface ϕ.

B. Finite Difference Approximation

In our implementation, which is motivated by the relation to the Hamilton–Jacobi type equations, we use the following finite difference approximation [14], [15], [13]. Define the *minmod* finite derivative as

$$minmod\{a, b\} = \begin{cases} \text{sign}(a)\min(|a|, |b|) & \text{if } ab > 0 \\ 0 & \text{otherwise} \end{cases}$$

We use this definition to approximate $\phi_x\phi_y$ by

$$\phi_x\phi_y\Big|_{x=i\Delta x, y=j\Delta y} \approx minmod\left(D_x^+\phi_{i,j}, D_x^-\phi_{i,j}\right) minmod\left(D_y^+\phi_{i,j}, D_y^-\phi_{i,j}\right).$$

where $D_x^+\phi_{i,j} \equiv \phi_{i+1,j} - \phi_{i,j}$, $D_x^-\phi_{i,j} \equiv \phi_{i,j} - \phi_{i-1,j}$, $D_y^+\phi_{i,j} \equiv \phi_{i,j+1} - \phi_{i,j}$ and $D_y^-\phi_{i,j} \equiv \phi_{i,j} - \phi_{i,j-1}$, for $\phi_{i,j} \equiv \phi(i\Delta x, j\Delta y, t)$ and $\Delta x = \Delta y = 1$.

A different approximation, that is also motivated by the hyperbolic conservation laws, is used for the squared partial derivatives [15], and is defined as

$$\phi_x^2\Big|_{x=i\Delta x, y=j\Delta y} \approx \left(\max\left(D_x^+\phi_{i,j}, -D_x^-\phi_{i,j}, 0\right)\right)^2$$

$$\phi_y^2\Big|_{x=i\Delta x, y=j\Delta y} \approx \left(\max\left(D_y^+\phi_{i,j}, -D_y^-\phi_{i,j}, 0\right)\right)^2$$

These finite difference approximations yield a first order numerical scheme for the equal distance contours evolution. Using a forward difference approximation in time gives the following numerical scheme for the propagation of the function $\phi_{i,j}^n = \phi(i\Delta x, j\Delta y, n\Delta t)$ on the (x, y) rectangular grid

$$\begin{aligned} \phi_{i,j}^{n+1} = \phi_{i,j}^n + \Delta t \Big[& a_{i,j}\left(\max\left(D_x^+\phi_{i,j}, -D_x^-\phi_{i,j}, 0\right)\right)^2 \\ & + b_{i,j}\left(\max\left(D_y^+\phi_{i,j}, -D_y^-\phi_{i,j}, 0\right)\right)^2 \\ & - c_{i,j}\, minmod\left(D_x^+\phi_{i,j}, D_x^-\phi_{i,j}\right) minmod\left(D_y^+\phi_{i,j}, D_y^-\phi_{i,j}\right)\Big]^{\frac{1}{2}} \end{aligned} \tag{6}$$

which is the finite difference approximation of (5). This numerical scheme is stable and inherently overcomes topological changes in the evolving contour. For higher order accuracy numerical schemes that deal with such *Hamilton Jacobi* type of equations see [12].

C. Initialization

The function $\phi(x, y, 0)$ is actually an implicit representation of $\partial\tilde{S}$ —the projection of the boundary of the source area ∂S to the (x, y)–plane. The first demand for $\phi(x, y, 0)$ is to follow

$$X(0) \equiv \left\{(x, y) \middle| \phi(x, y, 0) = 0\right\} = \left\{(x, y) \middle| (x, y, z(x, y)) \in \partial S\right\} \equiv \partial\tilde{S}.$$

Furthermore, $\phi(x, y, 0)$ should admit smoothness, continuity and be negative in the interior of $\partial\tilde{S}$, and positive in the exterior of $\partial\tilde{S}$. The $(2D)$ planar zero distance contour $\partial\tilde{S}$ is the projection of ∂S on the plane.

There are many ways to initialize $\phi(x, y, 0)$, for examples see [6], [5], [11]. It is possible, for example, to truncate the values of $\phi(x, y, 0)$ using the observation that we are interested in the function behavior only near the relevant contour, (the zero level set). Note that every ϕ function which obeys the demands described earlier is sufficient.

D. Distance Assignment

After the initialization is completed, the ϕ function is propagated according to (6). While propagating the function, our goal is to find the distance of each grid point. A simple way of achieving (first order) accurate results is by interpolating the zero crossings. At every iteration step, for each grid point, check

$$\textbf{If } \left(\phi_{i,j}^{n} \cdot \phi_{i,j}^{n-1} < 0\right) \textbf{ then } Distance_{i,j} = \Delta t\left(n - \frac{\phi_{i,j}^{n}}{\phi_{i,j}^{n} - \phi_{i,j}^{n-1}}\right).$$

Using this procedure, each grid point gets its distance at the "time" when the ϕ functions'–zero level passes through it.

E. Finding Minimal Geodesics

Having $\mathcal{M}_S$ and $\mathcal{M}_D$ on the grid, the minimal geodesic may be found in a simple way. Recalling that $g_m = \min(\mathcal{M}_S + \mathcal{M}_D)$, the projection of the minimal geodesic, G, on to the (x, y)–plane is $\tilde{G} = \left\{(x, y) \middle| \left(\mathcal{M}_S(x, y) + \mathcal{M}_D(x, y)\right) = g_m\right\}$.

The desired minimal geodesics are achieved by applying a contour finder on $\mathcal{M}_S + \mathcal{M}_D$ to find the level set $g_m + \varepsilon$, for some very small ε, and then applying a simple thinning algorithm that operates on the interior of the minimal level set. It is also possible to apply level sets based refinement methods [8], [1].

In [7] we also show how to use level sets of combinations of the geodesic distance maps to solve the 'three point Steiner Problem,' and how to compute Voronoi diagrams on $3D$ surfaces.

IV. Examples and Results

We demonstrate the performance of the algorithm by applying it to a synthetic surface and finding the paths of minimal length given on a mesh of 256 × 256 points. The source and destination areas are points located at $(x, y) = (64, 64)$ and $(x, y) = (192, 192)$. The (black) minimal geodesics connecting the source and the destination on the egg–box surface is presented in Fig. 4a. Fig. 4b present the projection of the equal distance contours from the source point, on which the minimal (plus epsilon) level set is displayed as a smooth black curve and the two minimal geodesics as a chain of pixels. The sum of the distance maps from the source and destination is presented in Fig. 4c and the geodesic distance from the destination point in Fig. 4d. The minimal level set of that sum surface (Fig. 4c) is the minimal geodesics.

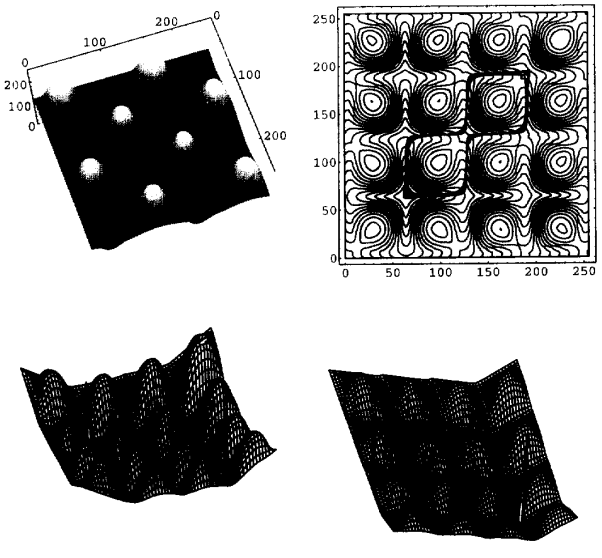

Fig. 4. (a) An "egg–box" surface with two minimal geodesics due to symmetry. (b) Equal geodesic distance contours from the source point, and the minimal level set as black curve. (c) Sum of the two geodesic distance maps. (d) The geodesic distance map from the destination point (192, 192). Artificial singular peaks indicate the source and destination in (c) and (d).

Ways of achieving more accurate results are by increasing the grid resolution, and by decreasing the time step ($\Delta t = 0.21$ in our examples).

V. Concluding Remarks

We have described a numerical method for calculating a geodesic distance map from a given area on a surface, so that topological problems in the propagated equal distance contours are inherently avoided. An algorithm for finding the minimal geodesics between two areas on the surface based on the distance mapping was constructed. The algorithm works on a grid, therefore it is easy to implement the algorithm in parallel using each mesh point as a small calculating device which communicates with its four close neighbors. In each iteration we need to calculate the values of $\phi(x, y, t)$ in those grid points close to the current contour and the rest of the grid points serve as sign holders. This can be exploited to reduce calculation effort. When not considering any possible redundancy, the calculation effort is of order $O(\frac{L}{\Delta t} m \cdot n)$, where L is the length of the shortest geodesic path and $m \cdot n$ is the number of grid points.

It was shown that wavefront propagation methods in fluid dynamics also provide a nice approach to the problem of finding the minimal geodesics.

Acknowledgments

We would like to thank Dr. N. Kiryati for introducing to us the problem of finding the shortest path, and Dr. Guillermo Sapiro, Prof. A. Guez, and Mr. D. Shaked for the many suggestions and discussions on this topic. We also thank the anonymous referee for noting the digitization bias due to metrication error of graph search algorithms, that in some cases (e.g., the plane) invariant to the grid size.

References

[1] D.L. Chopp, "Flow under geodesic curvature," Dept. of Mathematics Report 92-23, UCLA, 1992.

[2] M.P. Do Carmo, *Differential Geometry of Curves and Surfaces*. New Jersey: Prentice-Hall Inc., 1976.

[3] C.L. Epstein and M. Gage, "The curve shortening flow," *Wave Motion: Theory, Modeling, and Computation*. A. Chorin and A. Majda, eds. New York: Springer-Verlag, 1987.

[4] R. Kimmel, A. Amir, and A.M. Bruckstein, "Finding shortest paths on graph surfaces," CIS Report #9301, Technion, Israel, Jan. 1993.

[5] R. Kimmel and A.M. Bruckstein, "Shape from shading via level sets," CIS Report #9209, Technion, Israel, June 1992.

[6] R. Kimmel and A.M. Bruckstein, "Shape offsets via level sets," *CAD*, vol. 25, no. 5, pp. 154–162, Mar. 1993.

[7] R. Kimmel, N. Kiryati, and A.M. Bruckstein, "Distance maps and weighted distance transforms," *J. Mathematical Imaging and Vision, Special Issue on Topology and Geometry in Computer Vision*, to be published 1994.

[8] R. Kimmel and G. Sapiro, "Shortening three dimensional curves via two dimensional flows," *Int'l J. Comp. Math. with App.*, vol. 29, no. 3, pp. 49-62, 1995.

[9] N. Kiryati and G. Székely," Estimating shortest paths and minimal distances on digitized three dimensional surfaces," *Pattern Recognition*, vol. 26, no. 11, pp. 1,623–1,637, 1993.

[10] J.S.B. Mitchell, D. Payton, and D. Keirsey, "Planning and reasoning for autonomous vehicle control," *Int'l J. Intelligent Systems*, vol. 2, pp. 129–198, 1987.

[11] W. Mulder, S.J. Osher, and S.J. Sethian, "Computing interface motion in compressible gas dynamics," *J. Comp. Phys.*, vol. 100, pp. 209–228, 1992.

[12] S. Osher and C.W. Shu, "High–order essentially nonoscillatory schemes for Hamilton–Jacobi equations," *SIAM J. Numerical Analysis*, vol. 28, no. 4, pp. 907–922, Aug. 1991.

[13] S.J. Osher and L.I. Rudin, "Feature-oriented image enhancement using shock filters," *SIAM J. Numerical Analysis*, vol. 27, no. 4, pp. 919–940, Aug. 1990.

[14] S.J. Osher and J.A. Sethian, "Fronts propagating with curvature dependent speed: Algorithms based on Hamilton-Jacobi formulations," *J. Comp. Phys.*, vol. 79, pp. 12–49, 1988.

[15] G. Sapiro, R. Kimmel, D. Shaked, B. Kimia, and A.M. Bruckstein, "Implementing continuous-scale morphology via curve evolution," *Pattern Recognition*, vol. 26, no. 9, pp. 1,363–1,372, 1993.

[16] J.A. Sethian, "Curvature and the evolution of fronts," *Comm. in Math. Phys.*, vol. 101, pp. 487–499, 1985.

[17] J.A. Sethian, "A review of recent numerical algorithms for hypersurfaces moving with curvature dependent speed," *J. Differential Geometry*, vol. 33, pp. 131–161, 1989.

Correction

Correction to "Evaluation of Binarization Methods for Document Images"

Øvind Due Trier and Torfinn Taxt

In the March issue of this transactions, in the above-mentioned correspondence (vol. 17, no. 3., pp. 312-315), the authors made two revisions that were not incorporated into the final version. They were:

1. Ø. D. Trier and T. Taxt are with the University of Oslo, Department of Informatics, P.O. Box 1080 Blindern, N-0316, Norway. E-mail: trier@ifi.uio.no, torfinn@tor.pki.uib.no.

2. In Section II. Binarization Methods, in the first paragraph, line 7, the section "(black or gray)" should be deleted.