International Journal of Computer Vision, 5:3, 219–238 (1990)

Potentials, Valleys, and Dynamic Global Coverings*

CHANTAL DAVID AND STEVEN W. ZUCKER
McGill University, Research Centre for Intelligent Machines and Department of Electrical Engineering, 3480 University St., Montréal, Québec, Canada

Abstract
We present a new approach to effect the transition between local and global representations. It is based on the notion of a covering, or a collection of objects whose union is equivalent to the full one. The mathematics of computing global coverings are developed in the context of curve detection, where an intermediate representation (the tangent field) provides a reliable local description of curve structure. This local information is put together globally in the form of a potential distribution. The elements of the covering are then short curves, each of which evolves in parallel to seek the valleys of the potential distribution. The initial curve positions are also derived from the tangent field, and their evolution is governed by variational principles. When stationary configurations are achieved, the global dynamic covering is defined by the union of the local dynamic curves.

1 Introduction

One of the key questions in computational vision is how to effect the transition from local representations to global ones. For example, edge operators give an indication of local edge position and orientation; how can they be threaded together into a global contour? A popular solution is commonly called "edge following" or contour tracing [Ballard and Brown 1982; Levine 1985] and it works as follows: from a starting edge position, move in the direction indicated by the orientation until the next edge point is encountered, then move in that direction, and so on until the final point is reached. But such algorithms are inherently sequential, and result in myriad "garden paths" if any errors or noise are present. Global parameters, such as starting and ending points; total number of edge points; or total contour length, aid in their application, but are rarely available (except through user interaction, e.g., Kass, Witkin, and Terzopoulos [1987]). To make matters worse, inferring global parameters, such as the total length of a curve, raises problems as difficult as global curve detection itself! Other methods, more sophisticated than simple edge following, have been developed in an attempt to apply global constraints; these include minimizing properties such as contrast variation along the curve, total curvature [Montanari 1971; Martelli 1976], smoothness across scale space [Lowe 1988], energy within plate and membrane models [Blake and Zisserman 1987; Terzopoulos 1986], or applying top-down information [Tsotsos 1987; Draper et al. 1989]. However, few of the objects in our visual world are made from bent plates, and verifying problems generally arise when verifying such constraints are valid globally. It is almost as if the cart is being put before the horse, in that (hypothetical) global information is being used to force local decisions.

In an attempt to turn this situation around, we introduce a new class of algorithms for synthesizing contours from local representations of their differential structure. These algorithms are inherently parallel and are natural for vision applications. The mathematical idea behind these new algorithms is to compute properties of global structures indirectly by computing a *covering* of them. Loosely speaking, a covering of an object is a collection of different objects whose union is equivalent to it. In our case, we shall define a covering of a global curve that consists of a collection of short, overlapping curves. Each of these derives from a local representation of image information (which we call a tangent field) and then evolves like the "snakes" of Kass et al. [1988] according to a potential distribution also derived from the tangent field. But unlike standard "snakes," ours are local objects; it is their union that comprises the cover.

Each element of the cover evolves dynamically and in parallel, and no external global information is required. It is all available by construction in the potential distribution. For these reasons we refer to our representation of global curves as *dynamic coverings*.

The idea of computing dynamic coverings developed in a biologically oriented project of curve detection, a brief overview of which is provided next (see also [Zucker et al. 1989]). Within this context, the key to computing dynamic coverings is to impose two representations between the image and global curves: a tangent field and a potential distribution. The tangent field (defined in section 2) represents the local properties reliably, and the potential distribution puts them together globally. These intermediate representations thus break the enormous gulf between images and global contours into more manageable "chunks," and represent another key difference between our approach and other recent ones [Kass et al. 1988; Blake and Zisserman 1987]. We infer the tangent field from the image, and then build the potential distribution and snakes from it; other researchers attempt to build the potential distribution directly from the image. We can actually calculate the generators for the potential distribution from the equivalence class of curves that map into each tangent field entry.

This article is organized as follows. Following the overview of our approach to curve detection (section 2), we develop the preliminary notions of potential distributions (section 3) and deformable curves (section 4). We then get to the heart of the article in section 5, where the precise definition of a global covering is developed. Examples of the dynamic computation of global coverings are finally shown in section 6, for both artificial and natural images.

2 Overview of the Algorithm

The algorithm for curve detection consists of two distinct stages, the first leading to a coarse description of the local structure of curves, and the second to a much finer, global one [Zucker et al. 1988].

2.1 Stage 1: Inferring the Tangent Field

Orientation selection is the inference of a local description of a curve everywhere along it. Formally, this amounts to inferring the trace of the curve, or the set of points (in the image) through which the curve passes, its quantized tangent and curvature at those points, and the discontinuities [Zucker 1985]. Through all this work, we will refer to such information as the *tangent field*.

This first stage of orientation selection is in turn modeled as a two-step process:

Step 1.1: *Initial Measurement* of the local fit at each point to a model or orientation and curvature. The form of the initial measurements is biologically motivated, and a model of endstopped simple cells, which encode both the orientation and curvature estimates [Dobbins et al. 1987] is used to perform them. The local measurements are quantized into discrete classes of orientation and curvature. However, since the local measurements are inherently inaccurate, we require

Step 1.2: *Interpretation* of the local measurements such that they become globally consistent. Curvature, which relates neighboring tangents, is used to define a natural functional to be minimized in order to attain consistency [Parent and Zucker 1989; Iverson and Zucker 1987]. Relaxation labeling [Hummel and Zucker 1983] provides the formal framework for this.

A complete treatment of this first stage can be found in Iverson [1988].

At the end of the first stage, the tangent field gives the discrete trace of the curves in an image. The two steps above are necessary to guarantee that this trace contains reliable information. More precisely, the *tangent field* is a set of n 3-tuples,

$$\{((x_i, y_i), \theta_i, \kappa_i) \mid (x_i, y_i) \in I, \theta_i \in \Theta, \kappa_i \in \mathcal{K}\}$$

$i = 1, \ldots, n$, where n is the number of trace points in the image I, (x_i, y_i) denotes the (quantized) grid coordinate of the ith trace point indexed over the image I, and θ_i and κ_i are its quantized orientation and curvature, respectively. Typically, Θ consists of eight orientation classes, and $\mathcal{K}$ of five curvature classes (see figure 1a). Thus the tangent field summarizes the local differential properties of curves in what Zucker [1987] described as a qualitative manner.

Two topological properties of the tangent field are important: (1) each curve in the image is represented as a connected (in the discrete sense) set of tangent-field entries; and (2) discontinuities, intersections, and bifurcations are represented as multiple tangent-field entries at identical coordinates. To elaborate on (1), recall that, except on the border, each point (x, y) of

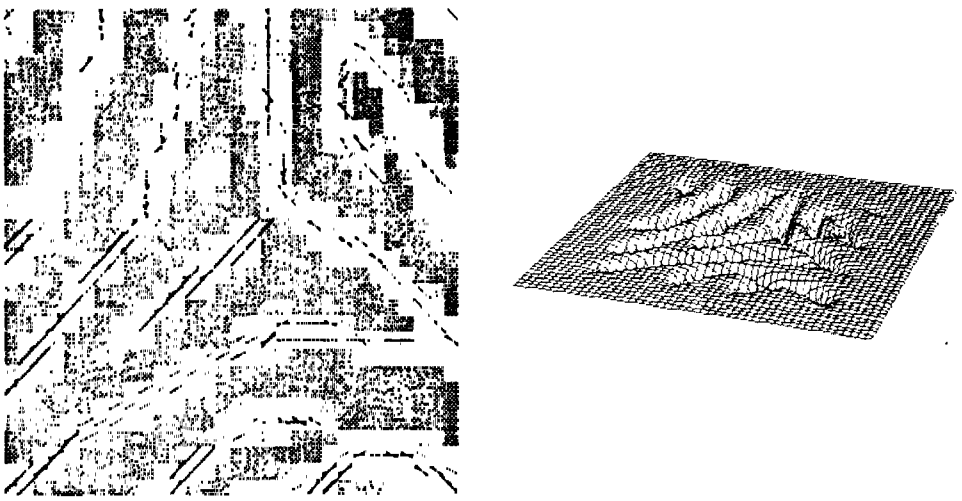

Fig. 1. Illustration of the two intermediate structures between the image and the global curves: the tangent field and the **potential distribution**. The image is a section of a fingerprint image, where each individual pixel is shown as a little block. (a) The tangent field, superimposed onto the original image. Each tangent is represented as a unit-length straight line in the tangent direction (curvature is not displayed). (b) The potential distribution, computed from the discrete tangent field by pointwise summation of elongated 2D Gaussians. Notice the presence of smooth valleys, corresponding to the curves in the image.

a (digital) image has four horizontal and vertical neighbors, and four diagonal neighbors; these eight points are called the (3×3) *neighborhood* of (x, y). A *chain* between two points (x_1, y_1) and (x_2, y_2) is a sequence of points where the first one is (x_1, y_1), the last one is (x_2, y_2), and each point of the sequence is in the neighborhood of the preceding one. A set of points of an image is then said to be connected when there exists a chain between any two points of the set [Rosenfeld and Kak 1982]. For an elaboration of (2), see [Zucker et al. 1988].

2.2 *Stage 2: Inferring a Covering of the Curve*

Since the tangent is the first derivative of a curve (with respect to arc length), the global curve can be recovered as an integral through the tangent field. Such a view typically leads to sequential recovery (numerical integration) algorithms (e.g., contour following). But these algorithms require starting points and some amount of topological structure (i.e., which tangent point follows which), and there are immense difficulties inherent in obtaining these parameters directly from the image. Moreover, because we are interested in biologically plausible models, we shall require that all the algorithms be parallel.

The key idea behind our approach is to recover the global curve by computing a *covering* that is sufficient for obtaining any global properties of the curve required for subsequent visual processing. The elements of the covering are short curves, initially born at each (discrete) tangent location with unit length, but which then evolve according to a potential distribution constructed from the tangent field. Thus the algorithm that we are proposing can be viewed, mathematically, as approximating the computation of integral curves through a direction field in a parallel manner with very nonstandard initial conditions (provided, of course, that multiple entries at the same position are properly interpreted).

Again, there are two conceptually distinct steps to stage 2 of the algorithm:

Step 2.1: *Constructing the Potential Distribution* from the discrete tangent field. Assuming that each tangent in the tangent field gives a coarse hypothesis about the behavior of the curve in its neighborhood, for each tangent-field entry we represent this class of hypotheses by an elongated (in the tangent direction) 2D Gaussian. This weighting function indicates the (equivalence) class of curves that project to each tangent-field entry in extending the discrete tangent in both length and width. The potential distribution is the pointwise summation of these n (one per tangent-field entry) Gaussians. The

key feature of the potential distribution is the presence of smooth valleys, constructed by summation of nearby Gaussians. These valleys indicate, with theoretical continuum precision (although the implementation is limited to machine precision), the expected positions of the curves in the picture (see figure 1b). The construction of the potential distribution thus effects a transition from a local to a global representation of structure and shape.

Step 2.2: *Curve Dynamics.* The valleys of the potential distribution are located by dynamic unit-length curves, born at each tangent position in the tangent field. They evolve according to a variational scheme that depends on curve properties (tension and rigidity) as well as on the potential distribution. The evolution takes two forms: (i) a migration in position to achieve smooth coverings at subpixel accuracy; and (ii) a growth in length such that nearby covering elements overlap; see figure 2. At the end of the migration process, we shall have that each valley is completely covered by this collection of curves, and each curve will have deformed itself to fit the exact shape of the valley. We call this collection of curves the *global covering* of the curves in the image.

The dynamic curves of the second stage are a generalization of "snakes" [Kass et al. 1988]. The key differences between our approach and theirs are twofold: (i) our scheme is parallel, so questions about total length, boundary points, etc., are handled implicitly; and (ii) our scheme is constructive, with the potential distribution computed directly from the tangent field, so questions about features, degree of image blurring, etc., are handled explicitly. We do not need a user to interactively supply information as did, for example, Kass et al., nor must we—to avoid smoothing across discontinuities—distribute penalties globally as, for example, Blake and Zisserman were required to do. Imposing stable intermediate structures—the tangent field and the potential distribution (see figure 1)—between the image and the global curves was the key, and the mathematical notion of coverings provided the unifying structure.

We now proceed to develop the notions of potential distribution and dynamic curves, before putting them together into global dynamic coverings in section 5.

3 The Potential Distribution

The local descriptors in the tangent field are synthesized into a global description represented as a potential distribution. In this section, we first articulate this construction process, then define a mathematical notion of a valley in the potential distribution. And, finally, since valleys are intended to be the locus of points that represent (the trace of) a curve, we show that, for simple potential distributions, these valleys do indeed correspond to the global curves.

3.1 Constructing the Potential Distribution

The tangent field, consisting of discrete trace, discrete tangent, and discrete curvature, gives, at each trace point, a coarse estimate of the local behavior of curves. Since each tangent-field entry corresponds not to a single curve in the world, but rather represents the equivalence class of possible curves that would project onto this tangent-field entry, we represent this (local) equivalence class of possibilities by the (Gaussian-weighted) Wiener measure over them. The Wiener measure (a Gaussian distribution) arises because the class of continuous (but not necessarily differentiable) functions is equivalent to the sample functions of a Brownian motion, and the Gaussian weighting along it is an approximation to the diffusion of this measure in time [Doob 1984]. The result, then, is a two-dimensional Gaussian distribution; and, since each entry in the tangent field is considered independently, the total potential distribution is the pointwise summation of them.

More formally, the set of generators of the potential field is the set of 2D-Gaussians $\{G_i : i = 1, \ldots, n\}$. For each trace point i, G_i is obtained by the multiplication of two 1D-Gaussians—one in the direction of the quantized tangent at this point, and another (with smaller deviation) in the perpendicular direction. Thus, at a trace point $i = (x_i, y_i)$, with tangent θ_i

$$\begin{aligned} G_i(x, y) &= -(K_E e^{-(f_i(x,y)-x_i)^2/\sigma_E^2})(K_B e^{-(g_i(x,y)-y_i)^2/\sigma_B^2}) \\ &= -K_E K_B e^{-((f_i(x,y)-x_i)^2/\sigma_E^2)} e^{-((g_i(x,y)-y_i)^2/\sigma_B^2)} \end{aligned} \tag{1}$$

where the functions $f_i(x, y)$ and $g_i(x, y)$, which rotate the x and y axes to the axes indicated by the θ_i and $\theta_i + \pi/2$ directions respectively, are given by

$$f_i(x, y) = x_i + (x - x_i) \cos \theta_i + (y - y_i) \sin \theta_i$$

$$g_i(x, y) = y_i - (x - x_i) \sin \theta_i + (y - y_i) \cos \theta_i$$

The parameter σ_E, which is the standard deviation of the 1D-Gaussian in the θ_i-direction, controls how far the local information given by the tangent θ_i is extended

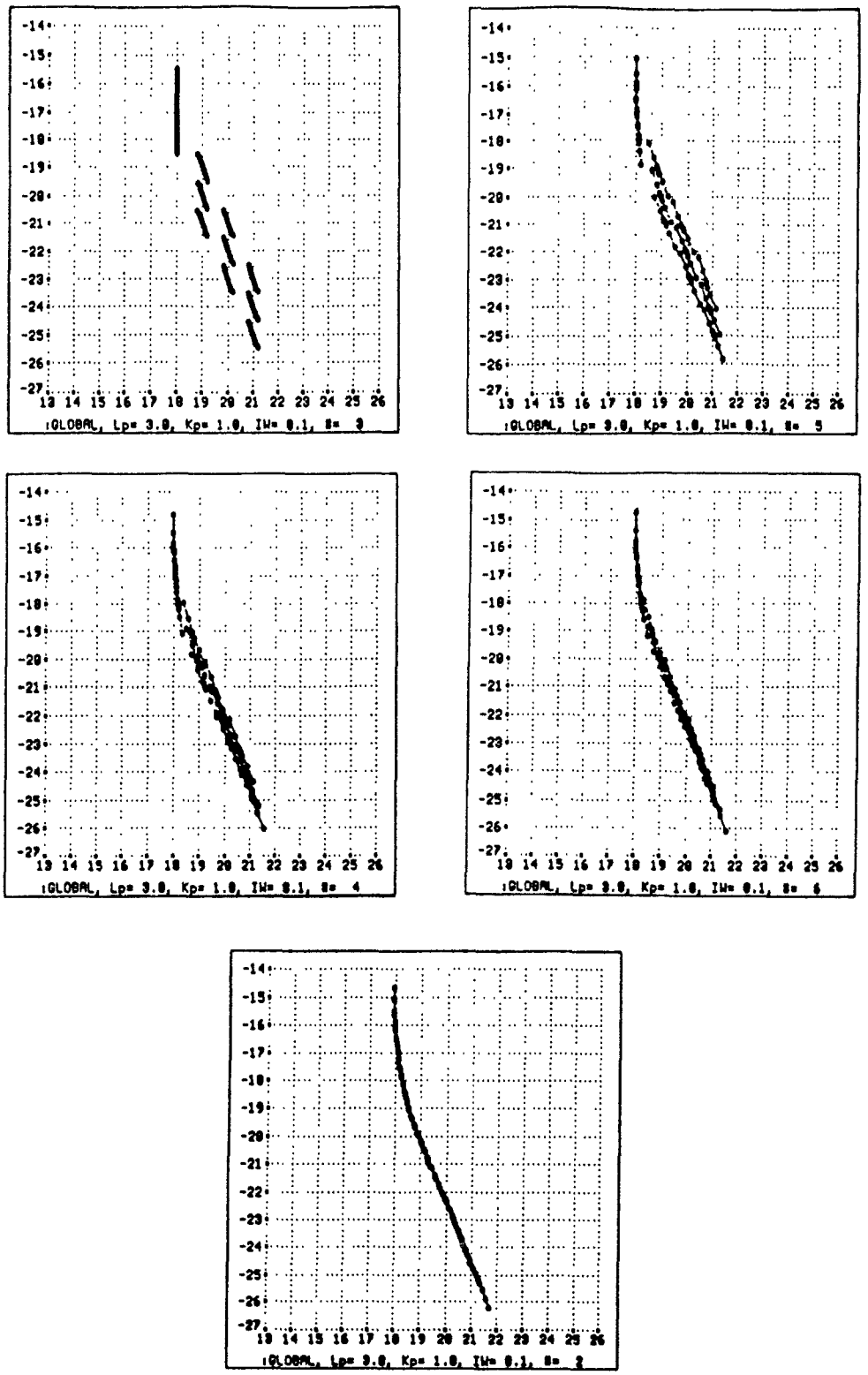

Fig. 2. Illustration of how the collection of short dynamic curves evolve to cover the locus of points comprising the valleys in the potential distribution. In the usual order, the covering elements migrate from their initial positions (top left) to reach the bottom of the valley indicating the edge (bottom). Note the two aspects of the movement: a migration, and a length increase.

from the trace point (x_i, y_i). The parameter σ_B, which is the standard deviation of the 1D-Gaussian in the $(\theta_i + \pi/2)$-direction, controls the width over which the extended information is valid. For these reasons, we shall refer to the parameters σ_E and σ_B as the *extension parameter* and the *blurring parameter*, respectively, and K_E and K_B are (positive) weights associated with each of the 1D-Gaussians. Figure 3 presents a plot of the resulting 2D-Gaussian.

The *(tangent-field) potential distribution*, U_{TF}, is then the pointwise summation of the G_i:

$$U_{TF}(x, y) = \sum_{i=1}^{n} G_i(x, y)$$

It should be noted that each generator G_i, as defined, is a function of the tangent estimate alone, with the curvature estimate negligible. The curvature could be used if curved, 2D Gaussians G_i' were defined for each tangent-field entry i. G_i' would then be a Gaussian along a curved axis of curvature κ_i in the θ_i direction, multiplied by another Gaussian perpendicular to the curved axis. However, the potential distribution U'_{TF} obtained by pointwise summation of these G_i', which is

$$U'_{TF}(x, y) = \sum_{i=1}^{n} G_i'(x, y)$$

would be very similar to U_{TF} for the following reasons. First, the curvature is only retrieved in a very coarsely quantized form by the first stage of the curve-detection process, so the generators G_i' would be only weakly tuned to curvature. Second, extending and summing the tangent information implicitly gives curvature, which is a relation between neighboring tangents. Thus, theoretically there appears to be no need to use curved Gaussians, and experimentally, potential distributions computed with straight Gaussians yielded smooth curved valleys. Figure 4 shows examples of potential distributions and of these valleys, where one can see that the result is a smooth "landscape," in which the jaggies due to sampling, noise, and quantization have been removed.[1]

Recall from the introduction that the key feature of the potential distribution is the location of the valleys, since it is to the valleys that the covering elements migrate. Thus we can now state the claims behind this article: *(i) the valleys in the potential distribution correspond to smooth curves*, and *(ii) the set of quantized tangents of these smooth curves is precisely the tangent field used to construct the distribution*. Moreover, over all the possible sets of curves that would project onto this tangent field, the valleys of the potential distribution give the most probable ones, in the sense of Wiener measure.

The rest of this section is devoted to an analysis of the potential distribution characteristics, in order to prove the above claims. This analysis results in a design criterion for the blurring parameter, which insures that separate curves remain separate.

The analysis is conducted for C^1 (i.e., continuous with continuous first derivatives) and nonintersecting curves, however some of the tangent fields used in the experiments contain intersecting curves and corners (in

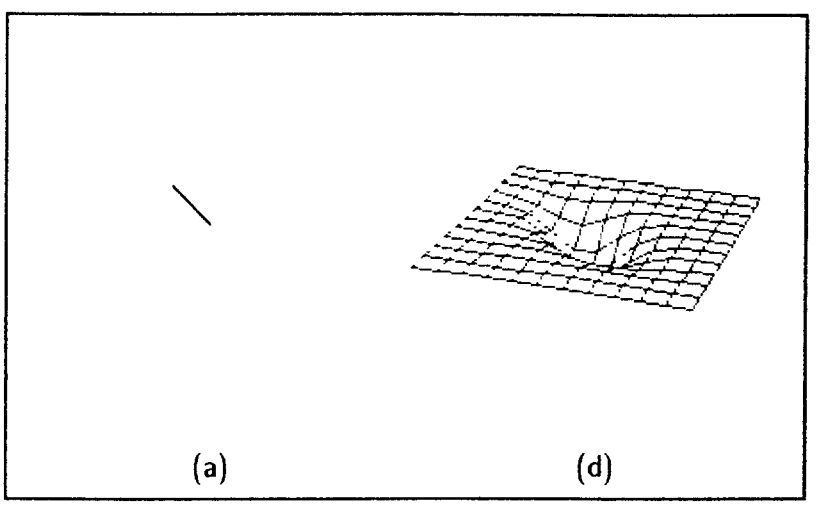

Fig. 3. 3D plot of $G_i(x, y)$. Note that the resulting Gaussian is elongated (in the θ_i-direction) since $\sigma_E > \sigma_B$.

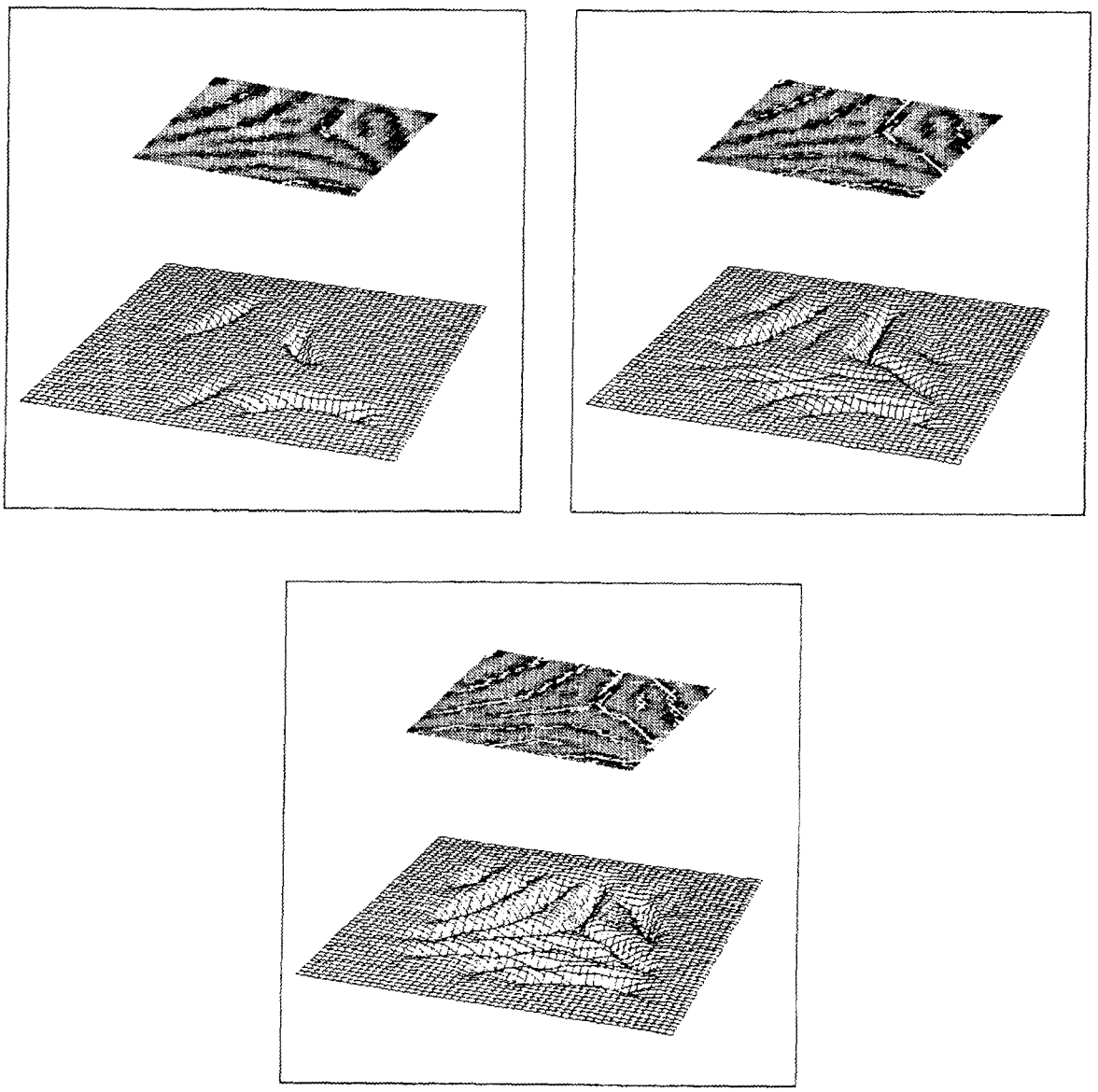

Fig. 4. Examples of potential distributions. The image is the fingerprint image shown in figure 1. It is displayed in reverse contrast, with the tangent field shown in white. Three potential distributions are shown, computed from first a few, then more, and finally all the tangent field entries.

particular, see figure 8). Other research is currently in progress to deal with intersecting curves and tangent discontinuities by extending the potential distribution to a branched potential distribution; this research will be reported elsewhere.[2]

3.2 Valleys in the Potential Distribution

We now introduce the concept of a valley, the locus of points toward which the elements of our covering migrate. To begin, given a direction **v**, making an angle ϕ with the positive x-axis, we define a *ϕ-valley* of the potential distribution $U_{TF}(x, y)$ to be a connected component of the set

$$\{(x, y) : \nabla_{\phi+\pi/2} U_{TF}(x, y) = 0, \quad \nabla^2_{\phi+\pi/2} U_{TF}(x, y) > 0\}$$

where $\nabla_\theta f(x, y)$ and $\nabla^2_\theta f(x, y)$ denote respectively the first and second directional derivative of f in the direction given by the angle θ. More generally, a *valley*

of the potential distribution $U_{TF}(x, y)$ is a connected subset of the set

$$\{(x, y) : \nabla_{\phi+\pi/2} U_{TF}(x, y) = 0, \quad \nabla^2_{\phi+\pi/2} U_{TF}(x, y) > 0\}$$

where $\phi = \phi(x, y)$, that is, the angle ϕ can vary with the coordinates x and y. A ϕ-valley is then a "straight" valley in the direction ϕ.

The idea behind the concept of valleys is that of *directional minima*. While local minima are points of a valley, points of a valley are not (necessarily) local minima of the potential distribution. The condition for a point (x, y) to be a local minimum of U_{TF} is that $\nabla_\theta U_{TF}(x, y) = 0$ and $\nabla^2_\theta U_{TF}(x, y) > 0$ *for all directions* θ. The definition of a valley is thus weaker than the definition of a local minimum, and points in the valleys are minima with regard to a single direction, namely the direction given at each point by the angle $\phi(x, y)$.

3.3 Analysis of Potential Distributions

The analysis of the two simplest potential distributions, namely a single Gaussian, or perfectly aligned Gaussians, is simple and shows that the valleys are ϕ-valleys (where ϕ is the direction of the Gaussian) corresponding to the most probable smooth curve indicated by the tangent field. The formal treatment of those cases can be found in [David and Zucker 1989].

In order to achieve this result for a more general potential distribution, we consider the case of two parallel Gaussians (this also leads to a criterion on the blurring parameter of the Gaussian weighting functions which insures curve separation).

Observe that, when the potential distribution is generated by several Gaussians that are in general position, we require the potential distribution to be such that: (1) nearby tangent-field entries that lie along the trace of a unique curve must form a unique valley, corresponding to the most probable position of the curve, and (2) nearby tangent-field entries belonging to two nearby curves must not be blurred together, and must give rise to two separate valleys.

These two requirements are sufficient to insure curve separation, since the curves are found by a migration of covering elements, born at tangent-field entry positions, over the potential distribution (see section 2.2). Then, in the case of tangent-field entries belonging to different curves, the migration will stop as soon as the covering elements each drop into their surrounding valley. Hence, possible additional valleys between the two positions will never be found if two valleys exist sufficiently close to each one of the original positions.

We now concentrate on the second observation above, namely that nearby curves give rise to distinct valleys. This leads to our design criterion for σ_B. Again consider only two interacting Gaussians, but from a tangent field consisting of two parallel and vertical (i.e., $\theta_1 = \theta_2 = \pi/2$) tangents, with identical y-coordinates. Then, $U_{TF} = G_1(x, y) + G_2(x, y)$, and using $\theta_1 = \theta_2 = \pi/2$, $y_1 = y_2$ in equation 1, we get

$$\begin{aligned} G_1(x, y) &= -K_E K_B e^{-((y-y_1)^2/\sigma_E^2)} e^{-((x-x_1)^2/\sigma_B^2)} \\ G_2(x, y) &= -K_E K_B e^{-((y-y_1)^2/\sigma_E^2)} e^{-((x-x_2)^2/\sigma_B^2)} \end{aligned} \tag{2}$$

(see figure 5).

We now look for the $(\pi/2)$-valleys of U_{TF} in this particular case. From now on, and without loss of generality, we assume $x_1 < x_2$. The critical points (for the $(\pi/2)$-valleys) are those where $\nabla_\pi U_{TF} = 0$, or equivalently, those where $(\partial/\partial x) U_{TF}(x, y) = 0$. Computing the partial derivative, we obtain

$$\frac{2K_E K_B e^{-(y-y_1)^2/\sigma_E^2}}{\sigma_B^2} \left[(x - x_1) e^{-(x-x_1)^2/\sigma_B^2} + (x - x_2) e^{-(x-x_2)^2/\sigma_B^2}\right] = 0$$

or, since first term of the product is always strictly positive,

$$(x - x_1) e^{-(x-x_1)^2/\sigma_B^2} + (x - x_2) e^{-(x-x_2)^2/\sigma_B^2} = 0 \tag{3}$$

Then, by positivity of the exponential function, any solution x of equation (3) is in the open interval (x_1, x_2). Also, equation (3) depends only on $\Delta x = x_2 - x_1$ and σ_B. Without loss of generality, set $x_1 = 0$; then $x_2 = \Delta x$ and $F(x_1, x_2, \sigma_B)$ becomes

$$x e^{-x^2/\sigma_B^2} + (x - \Delta x) e^{-(x-\Delta x)^2/\sigma_B^2} \tag{4}$$

which depends only on Δx and σ_B, and which we denote $F(x)$ for short.

Observe that the points (x, y) belonging to the $(\pi/2)$-valleys of U_{TF} are entirely characterized by the function $F(x)$.[3] More precisely, it can be shown that (x_0, y_0) belongs to a $(\pi/2)$-valley of U_{TF} if and only if $F(x_0) = 0$ and $(\partial/\partial x) F(x_0) > 0$.

A condition to separate cases (b) and (c) of figure 5 relates the distance Δx between the two Gaussian operators and the blurring parameter σ_B. More precisely, it can be deduced from the last result [David and Zucker

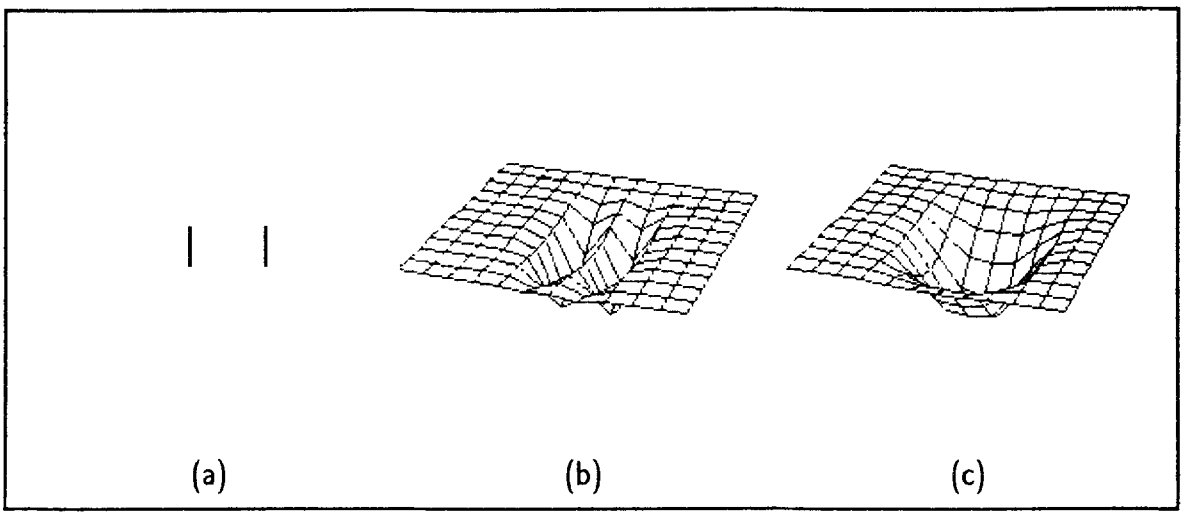

Fig. 5. (a) The tangent field, consisting of two entries with $\theta_1 = \theta_2 = \pi/2$ and $y_1 = y_2$. (b) A potential distribution obtained from (a) choosing $\sigma_E = 2.75$, $\sigma_B = \sigma_E/4 = 0.6875$, $K_E = K_B = 1$. This Gaussian separates the tangents into two different valleys. (c) A potential distribution obtained from (a) choosing $\sigma_E = 2.75$, $\sigma_B = \sigma_E/1.5 \approx 1.83$, $K_E = K_B = 1$. This Gaussian blurs the tangents into the same valley.

1989] that the vertical straight line $x = \Delta x/2$ is a $(\pi/2)$-valley of U_{TF} if and only if $\Delta x^2 < 2\,\sigma_B^2$.

To summarize, the key structural properties of the potential distribution are:

i. When $\Delta x^2 < 2\,\sigma_B^2$, the valley at $x = \Delta x/2$ is unique.
ii. When $\Delta x^2 > 2\sigma_B^2$ so that $x = \Delta x/2$ becomes a maximum, two valleys are created, one at $x = x_1^* \approx x_1$, and one at $x = x_2^* \approx x_2$. Also, the valley at $x = \Delta x/2$ is lost. The slight variation in position from x_1 and x_2 arises because of a weak interaction between nearby Gaussians; see David and Zucker [1989].
iii. When Δx becomes large, the distances $(x_1^* - x_1)$ and $(x_2 - x_2^*)$ become small.
iv. When σ_B becomes small, the distances $(x_1^* - x_1)$ and $(x_2 - x_2^*)$ become small.

The preceding analysis indicates that the valleys in the potential distribution have precisely the properties required for synthesizing global curves from the tangent field. Before proceeding to the study of dynamic curves for locating these valleys, we extend one aspect of the preceding analysis into a constraint on the parameters controlling the Gaussian generators.

3.4 Design Criterion: Blurring Parameter

Continuing the particular case of the preceding section, we know that two (vertical and parallel) tangent-field entries give rise to a unique valley when $\Delta x^2 < 2\sigma_B^2$, and to two separate valleys when $\Delta x^2 > 2\sigma_B^2$. Recall from the introduction that image curves are represented in the tangent field as *connected* sets of tangent-field entries. Thus, when two (parallel and vertical) tangent-field entries each occupy overlapping (3×3) neighborhoods, they belong to the same curve, and when they are separated by at least one empty pixel, they belong to different curves.[4] Then, any value of the blurring parameter such that

$$1^2 < 2\sigma_B^2 < 2^2 \quad \text{or} \quad \frac{1}{\sqrt{2}} < \sigma_B < \sqrt{2} \tag{5}$$

insures that the two tangent-field entries give rise to separate valleys if and only if they belong to (locally) separate curves.

More generally, it is clear that many other relations (in position and orientation), apart from the particular cases just studied, are possible, and lead to similar inequalities. For example, consider two parallel tangent-field entries with $\theta = (\pi/4)$. These entries are then located at diagonal pixels and this case is just a rotation of the preceding one. But now, the distance between two neighboring tangent-field entries is the distance across the diagonal, that is, $\sqrt{2}$. Thus, any value of σ_B such that

$$\sqrt{2}^2 < 2\sigma_B^2 < (2\sqrt{2})^2 \quad \text{or} \quad 1 < \sigma_B < 2 \tag{6}$$

insures separation of (locally) disconnected sets of tangent-field entries, and only of those sets. We submit that these two cases are typical of the general one, and summarize the analysis by saying that when

i. the analysis is restricted to nonintersecting curves;
ii. curves of the image are represented as connected sets of tangent-field entries;
iii. the choice of σ_B is made according to equations (5) and (6);

we find that

1. each connected component of the tangent field will give rise to a unique valley, built from the interaction of all the tangent-field entries of the component; thus, the valley indicates the position of the most natural curve that projects onto the connected component of tangent-field entries;
2. different connected components of the tangent field will be separated into distinct valleys. These valleys are not completely independent since there is a weak attraction between nearby valleys [David and Zucker 1989]. But this attraction is non-negligible only for very close curves (a few pixels).

Experimentally, it will be shown in section 6 that any reasonable value of σ_B relative to equations (5) and (6) (say between 0.75 and 1.5) ensures that the two above claims are verified.

4 Valley Detection

From the tangent field, we derived a potential distribution in which the valleys indicate the position of the curves in the image. We are now ready to identify these valleys to obtain a precise description of the curves. Since the valleys are, by definition, connected loci of points that are not all minima of the potential distribution, standard minimization techniques are not adequate. Rather, we use short segments of curves to capture these loci, and we define a process by which the curves migrate into the valley. Together these segments will define the global covering. Figuratively, the process works as follows. Suppose that we are given a rough approximation of where a valley lies in the potential distribution. Now, if a curve is placed at this approximate position, within a potential well and parallel to the valley, and let slide as if it were governed by a kind of gravity, then it will slide toward the bottom of the well, that is, the valley. Moreover, if the curve is not completely rigid, it will deform itself to fit the exact shape of the valley. Then, at the end of the process, when "gravitational" stability is attained, a description of the valley is given by the associated curve. This is precisely what we shall do.

Note that the above mechanism relies on having a good initial approximation, in position and orientation, of each valley. If we place a piece of curve randomly on the potential distribution and let it evolve under "gravity," the final position of the curve could be meaningless in terms of valley detection. More likely than not, the curve would span several valleys. Our initial approximation is given by the tangent field, and the wells by the Gaussian generators. It is analyzed more carefully in section 5. In this section, we develop the variational principles that govern the dynamics of each deformable curve individually; the global cover will then be computed by running all of these dynamic processes in parallel. For now, we focus on one element of the cover, and suppose it exists at some (initial) position on the potential distribution. We then develop a mathematical model to govern the movement of the curve, basically following that adopted for *snakes* [Kass et al. 1988; Terzopoulos 1987a, 1987b], by applying classical mechanics to deformable curves.

A deformable curve v is a differentiable map $v : \mathbf{R}^2 \rightarrow \mathbf{R}^2$. We write $v(s, t) = (x(s, t), y(s, t))$ with $s \in \Omega$ the space domain, and $t \in [t_0, t_1]$ the time domain. We denote $(\partial/\partial s)v(s, t)$ by $v_s(s, t)$, $(\partial^2/\partial s^2)v(s, t)$ by $v_{ss}(s, y)$ and $(\partial/\partial t)v(s, t)$ by $v_t(s, y)$.

We now define the *potential functional* $U^t(v)$ of a curve v, where the superscript indicates that the potential functional is computed at each time t. This corresponds to the potential energy of the curve. It is related to the potential distribution, but also includes internal constraints on the curve (how the curve is allowed to extend and fold). These internal constraints, corresponding to the physical concepts of tension and rigidity, must be added in order to specify completely the movement of the curve. The potential functional of a curve v at each time t is

$$U^t(v) = \frac{1}{2}\int_\Omega f(|v_s|^2) + g(|v_{ss}|^2) + U_{TF}(v)ds \qquad (7)$$

where $U_{TF}(v)$ is the value of the potential distribution at the point $(x(s, t), y(s, t))$, $f(|v_s|^2)$ controls the tension of the curve, and $g(|v_{ss}|^2)$ controls the rigidity of the curve. Among all the possible functions f and g, we simply choose the natural ones

$$f(|v_s(s, t)|^2) = \omega_1(s)|v_s(s, t)|^2$$

$$f(|v_{ss}(s, t)|^2) = \omega_2(s)|v_{ss}(s, t)|^2$$

where ω_1 and ω_2 are constants over time. A method for choosing the constants is described in section 5;

we will also see that these simple forms are sufficient to control the curve movement for our purposes.

Now that the potential functional is completely defined, we can use the methods of classical mechanics (principle of least action, Lagrange-Euler equations of motion) and obtain a system of differential equations that describes the movement of the deformable curves. Then, with known initial positions, these equations can be solved numerically. All details can be found in [Terzopoulos 1987a, 1987b] and in the Appendix to this article.

5 The Global Covering

We now proceed to put all of the previous pieces together. We infer the global curves from a tangent field by recovering a *covering* of these curves. Each element of the covering is a deformable curve that moves according to the model described in section 4, within the wells of the potential distribution described in section 3. Of course, we require good initial positions for the covering elements, so that, at "gravitational" stability, the curves will all have moved into valleys; and we require that each valley be completely covered by curves. In this section, we precisely state what we mean by a covering of the global curves, and then develop a particular covering to meet the above two requirements.

5.1 Curves and Coverings

As a mathematical object, a plane curve v is a mapping from an interval $I \subset \mathsf{R}$ to R^2, that is,

$$v(s) = (x(s), y(s)) \qquad \text{for } s \in I$$

The set $v(I) \subset \mathsf{R}^2$ is called the *trace* of the curve v.

Since our overall goal is applications in vision, we seek to recover the trace of each curve as precisely as possible. Once this trace is available, it provides a basis for inferring all the other curve properties that do not depend on the exact form of the mapping. This leads us to consider equivalence classes of curves obtained by factoring the set of curves by the equivalence relation given by the equality of trace. More formally, we propose that two curves $v : I_v \to \mathsf{R}^2$ and $u : I_u \to \mathsf{R}^2$ are said to be *equivalent* if and only if their traces coincide, that is, $v(I_v) = u(I_u)$. But an image rarely consists of a unique curve, and we want to study general classes of images. We thus say that two sets of curves U and V, where U consists of the curves $u : I_u \to \mathsf{R}^2$ and V consists of the curves $v : I_v \to \mathsf{R}^2$, are said to be *equivalent* if and only if

$$\bigcup_{u \in U} u(I_u) = \bigcup_{v \in V} v(I_v)$$

Given a set of curves, V, we denote the equivalence class formed by all sets of curves equivalent to V by $E(V)$, and any member U of $E(V)$ is called a *covering* of the set V, or is said to cover V. The elements of U (which are curves $u : I_u \to \mathsf{R}^2$) are called *covering elements*.

5.2 Construction of the Global Covering

We now develop a covering that enjoys natural representational properties. Recall that reliable local estimates of the curve positions and orientations are given in the tangent field. Now, consider each tangent-field entry as a unit-length straight curve, and let these n curves deform themselves according to the model already described. We then get the set C of these n deformed curves, where each curve $c \in C$ originated from a tangent-field entry. In this section, we will show that the set C is a covering (according to the above definition) of the global curves in an image. Or, equivalently, since the global curves are described by the valleys of the potential field we will show that each small curve c lies in a valley, and that each valley is completely covered by curves of the set C. More formally, the *global covering* of the curves in an image is the set C of the n curves

$$c_i(x, t_1) = (x_i(s, t_1), y_i(s, t_1)) \qquad i = 1, \ldots, n$$

where at initial time t_0, each $c_i(s, t_0)$ is a straight line identical to the ith entry of the tangent field, and at final time t_1, each $c_i(x, t_1)$ is a dynamic curve that has moved onto a valley of the potential distribution. Starting from now, the phrase "covering elements" will always refer to the elements of the global covering. We now complete their dynamics.

5.2 Covering Element Movement

Recall from the preceding section that for each covering element $c(s, t) = (x(s, t), y(s, t)$, with $s \in \Omega$, the potential functional is

$$U^t(c) = \int_\Omega \omega_1|c_s|^2 + \omega_2|c_{ss}|^2 + U_{TF}(c)\, ds$$
$$= \int_\Omega U_{TF}(c)\, ds + \int_\Omega \omega_1|c_s|^2 + \omega_2|c_{ss}|^2\, ds \quad (8)$$

where the first term controls the migration of the covering elements toward the valleys, and the second term controls how the covering elements are allowed to deform themselves during the migration. We call this latter term the *internal potential*, and set it so that each covering element increases its length to a prescribed length. The movement of the covering elements is thus a two-fold process, consisting of a migration and of a length increase (see figure 2). The details are as follows.

The internal potential of a covering element $c(s, t) = (x(s, t), y(s, t))$, with $s \in \Omega$ is

$$\int_\Omega \omega_1|c_s|^2 + \omega_2|c_{ss}|^2\, ds \quad (9)$$

where ω_1 and ω_2 control tension and rigidity respectively. In fact, since the minimization of (9) participates in the covering elements' movement, we observe that the covering elements tend to decrease their length for positive values of ω_1, increase their length for negative values of ω_1, become straight for positive values of ω_2, and fold and curve themselves for negative values of ω_2.

Given a prescribed mandatory length L_p, we set ω_1 to be, at each time t, $\omega_1(c) = f(L^t(c) - L_p)$, where $L^t(c)$ is the length of the covering element $c(s, t)$ at time t, that is,

$$L^t(c) = \int_\Omega |c_s(s, t)|\, ds = \int_\Omega \sqrt{x_s(s, t)^2 + y_s(s, t)^2}\, ds$$

The function f is then simply a ramp function saturating when $|L^t(c) - L_p|$ becomes big. It is easy to see that this choice achieves the desired result: a covering element c will tend to extend when $L^t(c) < L_p$, and to shrink when $L^t(c) > L_p$. It should be noted that ω_1 is independent of the parameter s, which makes the tension equally distributed over the entire covering element length.

Although the presence of a ridigity term is essential to the process, its exact value is not a key point, and it suffices to set it empirically within an order of magnitude to get good results. It was experimentally verified that changing the rigidity parameter by a reasonable amount does not change the results significantly (see [David and Zucker 1989]). But the rigidity parameter is essential to the algorithm, in order to prevent a covering element from folding on itself while finding its optimal position, since the potential distribution and the tension parameter do not prevent those movements. We simply set the rigidity parameter ω_2 to a constant value, the same for each covering element, at all times, and over the whole covering element length.

5.4 Properties of the Global Covering

Now that our model is completely defined, we can state and justify the claims made about the global covering. The global covering C, formed by the n curves c_i, which originated from the n tangent-field entries, and which evolve according to equation (8), is such that

1. Each curve lies in a valley of the potential distribution.
2. Each valley of the potential distribution is completely covered by curves of the set C.

Claim 1 is met by construction, since the potential distribution was built from the tangent field, that is, from the curves' initial positions. Then, each curve is, at initial time, within the well surrounding its associated valley, and the dynamics drive the curve onto the valley.

The truth of claim 2 is based upon three facts: global curves project into the tangent field as connected sets of tangent-field entries, the length of each covering element increases during the dynamics, and there is no movement of a covering element outside the well surrounding its valley. By the last statement, we assume that all valleys have a similar depth along their length, or equivalently, that the density of the tangent-field entries is about the same along the entire connected component. This is guaranteed by our restriction to smooth nonintersecting curves.

Note that the diagonal length across a pixel (which is the maximal distance between two neighboring elements of a connected set) is $\sqrt{2}$ units; therefore, to guarantee that neighboring covering elements overlap to give the global curves, we just have to let each covering element extend in length to some value greater than $\sqrt{2}$, since each covering element stays around its initial position. Typically, a prescribed length of 3 is used, in order to guarantee the overlap.

Since the valleys of the potential field describe the global curves of the image, we also find that the global covering is a covering of the curves in an image, that is,

$$\bigcup_{c \in C} c(\Omega, t_1) = \bigcup_{v \in V} v(I_v)$$

where V is the set of the global curves of the image.

6 Results

We now illustrate the computation of the global covering, beginning with artificial images to check the validity of the algorithm against known examples, and finally with natural images to demonstrate its robustness.

The first test shows that simple analytic curves are retrieved quite exactly from the tangent-field information, that is, that the "natural" curves indicated by the potential distribution correspond to the expected ones. The image consists of two nearby circles, with different radii, projected over a discrete sampling grid (figure 6a).

We simulate the first stage of the process (see section 2.1), which extracts the trace and tangent of these curves in the following manner. Each square of the sampling grid traversed by one of the circles belongs to the trace of this circle, and the tangent at each trace point is computed assuming that the curve crosses the center of the square; afterwards, all the computed tangents are quantized into eight classes (figure 6b). We then run our algorithm on the simulated tangent field. This tangent field is the only information that the second stage of the process uses to retrieve the global curves, the initial circles being lost by the projection onto the sampling grid. The result is shown in figures 6c and d. The curves retrieved from the potential distribution correspond very accurately to the initial circles. Also, the value of σ_B, chosen according to the design criterion, allows a clear separation between the two nearby circles, separated by only two pixels on the sampling grid.

Finally, we run our algorithm on three natural images, a fingerprint image, a radiograph of blood vessels in the brain, and a satellite image of logging roads. Since we are interested in precise extraction of the curves, without blurring of very close curves, we choose to experiment with small images so that details can be examined up close. Figure 7 presents the original images, and the resulting global curves. An exploded display of the first two tests is then presented in figures 8 and 9.

7 Summary and Conclusions

In this article we developed an algorithm for finding a representation of the global curves in an image. The novel feature of this representation is that it is formulated in mathematical terms as a covering, so that, rather than computing global curves directly, short segments of curve are computed independently from one another. The short segments thus become the elements of the cover, and global curves are given as their union. This enables us to define the algorithm in a parallel fashion, and to avoid the a priori specification of global parameters (e.g., the length of the curve).

Since the global representation is built up from local representations, the algorithm for computing global curves via coverings requires reliable local information about the curve. Within our context, which is a large project on curve detection incorporating both mathematical and biological constraints, this local information is provided by a tangent field, or a list of (quantized trace) points through which the curve passes, together with coarse estimates of its tangent and curvature at those points. The key point to emphasize here is that these local properties need not be specified to high precision; the interpolation properties implicit within the covering algorithm can deal successfully with coarse initial information. Thus the algorithm fits into a very natural information-processing hierarchy, with both degree-of-precision and local-to-global axes spanned.

The local (tangent-field) information is used in two different ways. First, a global description, in the form of a potential distribution, is synthesized from it. By definition, valleys in this potential distribution correspond precisely to curve locations. These valleys are detected by a dynamical process in which the covering elements are driven by the potential to both migrate in position and elongate in length. The second use of the local (tangent-field) information is in specifying the initial positions for each covering element. Convergence is thereby guaranteed by construction. Equivalence between the curves in the image and the covering elements is demonstrated both theoretically and by example.

Since the global curves are represented by a covering that evolves according to a dynamic process, we refer to them as *dynamic global coverings*. Although we concentrated on smooth, nonintersecting curves in this article, additional research is in progress to guarantee the validity of the results in the general case. It works as follows. Since the tangent field indicates curve intersections and discontinuities by multiple tangent field entries at the same grid coordinate, the potential distribution is split into several "layers," one per tangent-field entry, in the neighborhood of each of those coordinates. Take, for example, the case of two curves intersecting at a single point. The tangent-field entries of those curves will give rise to two independent layers in the potential distribution at this point, and the process described here can then be applied at each layer. (The advantage of

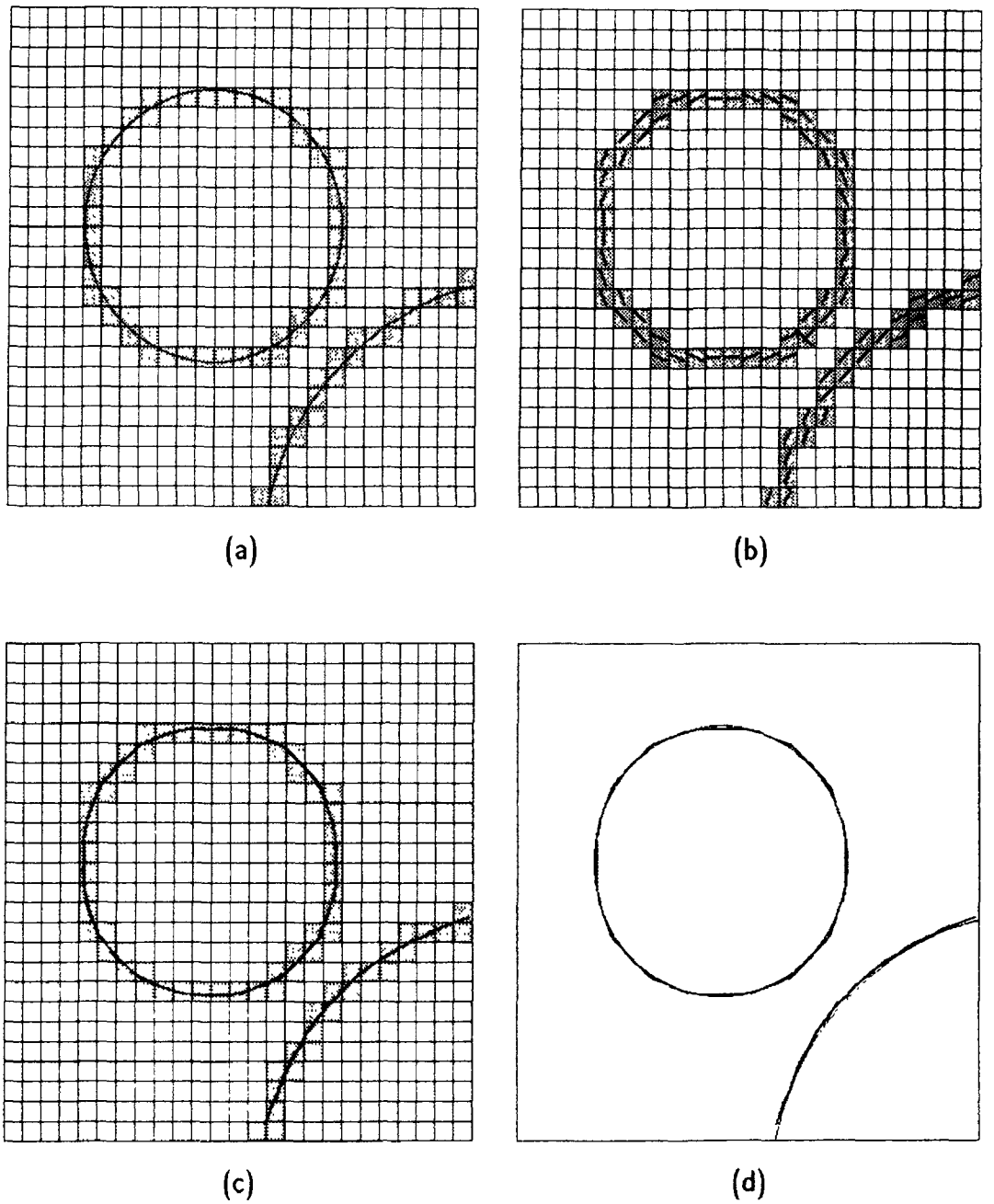

Fig. 6. (a) Two nearby circles of different radii lie over a discrete sampling grid (25 × 25). Their projection onto the grid is given by those pixels traversed by one of the two circles. (b) The tangent field for the two circles, consisting of the discretized tangent at each trace point. Note that the tangents' positions also correspond to the covering elements initial positions. (c) From their initial positons shown in (b), the covering elements have migrated into the valleys of the potential distribution, where they attain stability. The parameters of this experiment are: $\sigma_E = 2.75$, $\sigma_B = \sigma_E/3 = 0.85$, $L_p = 3.0$; ω_1 is determined by $R = 1.0$, $K = 1.0$, and $\omega_2 = 1.0$; $\mu = 2.0$, $\gamma = 2.0$. (d) The final covering elements' positions, superimposed over the initial curves.

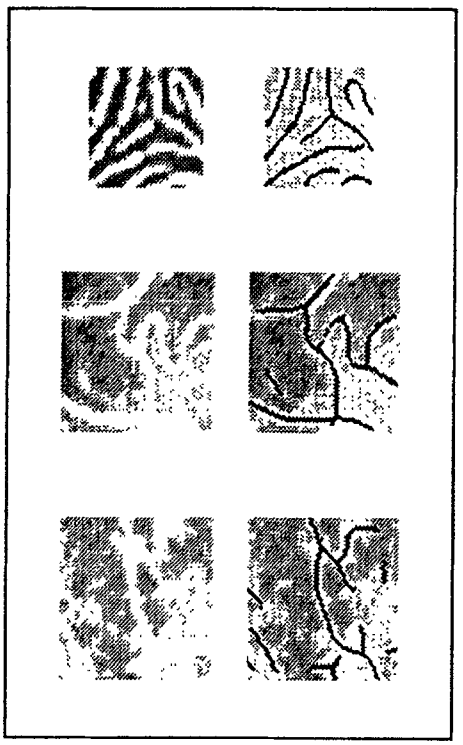

Fig. 7. Global curves of natural images. Top: A fingerprint image. Middle: A radiograph of blood vessels in the brain. Bottom: A satellite image of logging roads.

separating the tangent-field entries is, of course, to eliminate inappropriate interactions between the separate curve segments.)

The remaining step from a covering to a global representation is a specification of which covering elements belong to the same global curve. (It is usually the case, for example, that there are multiple curves within a single image.) We accomplish this by a straightforward exploitation of topological connectivity: graphically, the initial covering elements are all born with a different "color"; as two covering elements overlap dynamically, they become the same "color." In the end, "color" propagates along connected components, and the elements belonging to the covering of the same curve are all the same color, while the covers corresponding to different curves are different colors.

In conclusion, we believe that dynamic coverings have application to a variety of problems that must face the transition from local to global representations in a parallel, efficient way. For example, an extension to surface coverings could work as follows. We know that it is possible to extract, from 3D data, an intermediate structure that is the 3D equivalent of the tangent field (see Sander and Zucker [1990]). This structure mainly consists of an estimate of the tangent plane at each (3D) trace point. Starting from these planes, a generalization

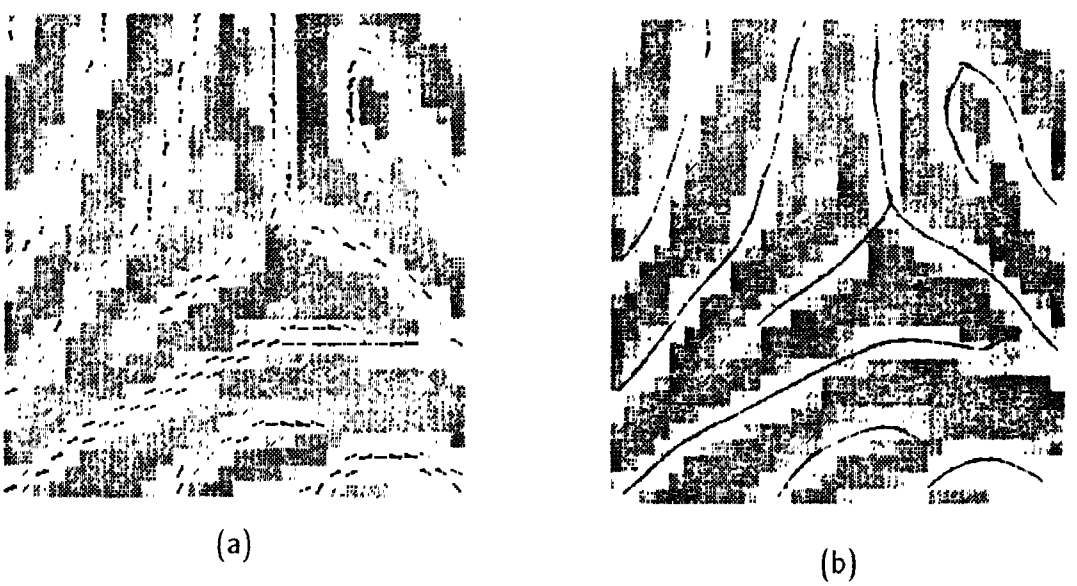

Fig. 8. Illustration of the first experiment on a fingerprint image. (a) The initial covering elements, at tangent field entries' positions. (b) Final positions of the covering elements; clearly the covering elements have migrated into the valleys, and have overlapped to form a covering. Also note that the bifurcation in the center of the image have been recovered correctly, without special treatment.

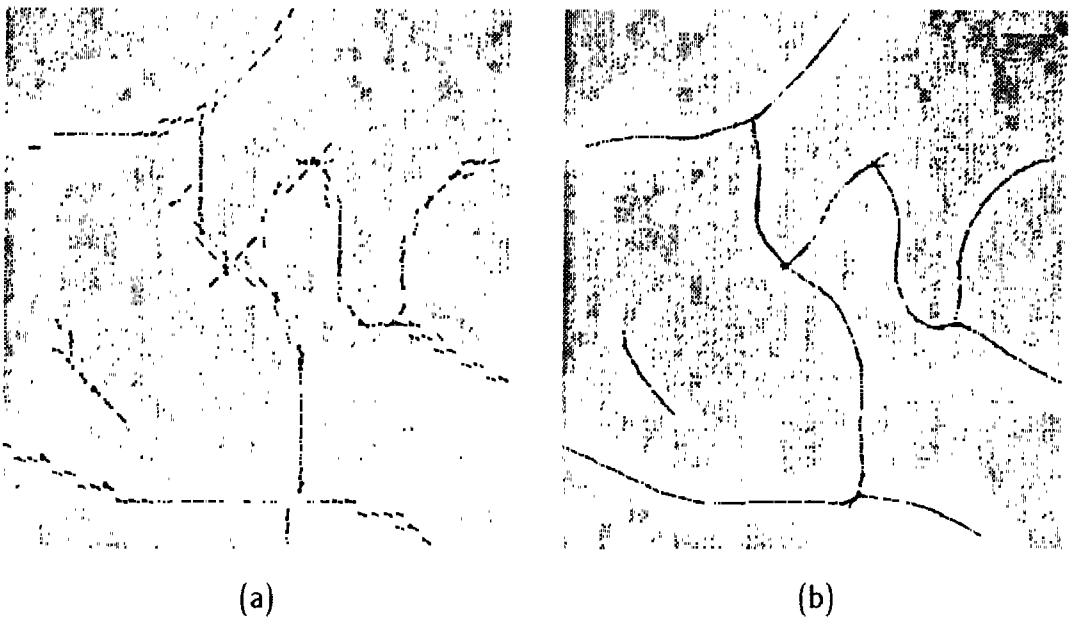

Fig. 9. The second test is performed on a biomedical image, a cerebral angiogram. (a) The initial positions of the covering elements, corresponding to the tangent-field entries. This image is very noisy, and low contrast. The tangent field was computed under the assumption that all vessels were of comparable contrast. (b) The final positions of the covering elements. The region where nearby covering elements do not overlap completely corresponds to a region where the tangent field was not completely connected.

of the model presented in this paper would evolve those local estimates to fit together into a (smooth) covering of the surface. The covering elements would thus be overlapping surface patches.

Notes

*Research supported by the Natural Sciences and Engineering Research Council of Canada and the Air Force Office of Scientific Research.

†Fellow, Canadian Institute for Advanced Research.

[1]Further computational experiments in which curved (G') generators were used led to final results virtually indistinguishable from those shown in this paper.

[2]Even if no specific treatment is performed at intersecting points and tangent discontinuities, the present algorithm does not necessarily behave badly at these locations. For example, the T-shape in the middle of the fingerprint image is correctly retrieved (see figure 8). But further analysis would have to be done in order to insure the validity of the results in the general case.

[3]Note that this also means that the presence of the $(\pi/2)$-valleys does not depend on the y coordinate.

[4]Or, they belong to different parts of the same global curve, i.e., there is a chain somewhere between the two tangents. But *locally*, they must be viewed as separated.

References

Ballard, B., and Brown, C. 1982. *Computer Vision*. Prentice Hall, Englewood Cliffs, N.J.

Benson, A., and Evans, D.J. 1977. A normalized algorithm for the solution of positive definite symmetric quintadiagnoal systems of linear equations. *ACM Trans. Mathematical Software* 3: 96–103.

Blake, A., and Zisserman, A. 1987. *Visual Reconstruction*. MIT Press: Cambridge, MA.

David, C., and Zucker, S.W. 1989. Potentials, valleys, and dynamic global coverings. Tech. Rept. TR-CIM 89-1, McGill University, Montréal.

Dobbins, A., Zucker, S.W., and Cynader, M.S. 1987. Endstopping in the visual cortex as a substrate for calculating curvature. *Nature* 329: 438–441.

Doob, J.L. 1984. *Classical Potential Theory and Its Probabilistic Counterpart*. Springer Verlag: New York.

Draper, B., Collins, R., Brolio, J., Hanson, A., and Riseman, E. 1989. The schema system. *Intern. J. Comput. Vision* 2 (3): 209–250.

Hummel, R., and Zucker, S.W. 1983. On the foundations of relaxation labelling processes. *IEEE Trans. PAMI* 5: 267–287.

Iverson, L., and Zucker, S.W. 1987. From orientation selection to optical flow: A computational perspective. In *Proc. IEEE Workshop in Computer Vision*, Miami, pp. 184–189.

Iverson, L.A. 1988. The description of image curves: Discrete forms of continuity. M.E. Thesis, McGill University, Montréal.

Kaas, W., and Witkin, A. 1987. Analyzing oriented patterns. *Comput. Vision, Graphics, and Image Process.* 37: 362–385.

Kaas, M., Witkin, A., and Terzopoulos, D. 1988. SNAKES: Active contour models. *Intern. J. Comput. Vision* 1 (4). 321–332.
Kaplan, W. 1952. *Advanced Calculus*. Addison-Wesley: Reading, MA.
Levine, M. 1985. *Vision in Man and Machine*. McGraw Hill: New York.
Lowe, D.G. 1988. Organization of smooth image curves at multiple scales. In *Proc. Intern. Conf. Comput. Vision* 2: 558–567.
Martelli, A. 1976. An application of heuristic search methods to edge and contour detection. *Comm. ACM* 19: 73.
Montanari, U. 1971. On the optimal detection of curves in noisy pictures. *Comm. ACM* 14: 335–345.
Parent, P., and Zucker, S.W. 1989. Trace inference, curvature consistency, and curve detection. *IEEE Trans. PAMI* 11: 823–839.
Rosenfeld, A., and Kak, A. 1982. *Digital Image Processing*. Academic Press: New York.
Sander, P., and Zucker, S.W. 1990. Inferring differential structure from 3-D images: Smooth cross section of fibre bundles. *IEEE Trans PAMI*, 12: 833–854.
Terzopoulos, D. 1986. Image analysis using multigrid relaxation methods. *IEEE Trans. PAMI* 8: 129–139.
Terzopoulos, D. 1987a. On matching deformable models to images. *Topical Meeting on Machine Vision*. Technical Digest Series 12: 160–163.
Terzopoulos, D. 1987b. On matching deformable models to images: Direct and iterative solutions. *Topical Meeting on Machine Vision*. Technical Digest Series 12: 164–167.
Tsotsos, J. 1987. Image understanding. In *Encyclopedia of Artificial Intelligence*. S. Shapiro (ed.). John Wiley: New York, pp. 389–409.
Zucker, S.W. 1985. Early orientation selection: Tangent fields and the dimensionality of their support, *Comput. Vision, Graphics, and Image Process.* 32: 74–103.
Zucker, S.W. 1987. The emerging paradigm of computational vision. *Annu. Rev. Comput. Sci.* 2: 69–89.
Zucker, S.W., Dobbins, A., and Iverson, L. 1989. Two stages of curve detection suggest two styles of visual computation. *Neural Computation*. 1. 68–81.
Zucker, S.W., David, C., Dobbins, A., and Iverson, L. 1988. The organization of curve detection: Coarse tangent fields and fine spline coverings. In *Proc. 2nd Intern. Conf. Comput. Vision*, Tarpon Springs, FL.

Appendix

A.1 Dynamics of the Deformable Curve

Given the potential functional $U^t(v)$, we can define the kinetic energy of the curve. Following classical mechanics, the kinetic energy functional at each time t is

$$T^t(v) = \frac{1}{2}\int_\omega \mu |v_t|^2 \, ds$$

where μ is the (constant) mass desnity.

According to the *principle of least action*, the motion of the deformable curve during the time interval $[t_0, t_1]$ is described by the functions $x(s, t)$, $y(s, t)$ for which the integral

$$\int_{t_0}^{t_1} L^t(v)\, dt$$

of the Lagrangian $L^t(v) = T^t(v) - U^t(v)$ is a minimum. It is also known that any extrema of

$$\int_{t_0}^{t_1} L^t(v)\, dt$$

must satisfy the Euler-Lagrange equations. Setting

$$f(v) = \frac{1}{2}\left(\mu|v_t|^2 - \omega_1|v_s|^2 - \omega_2|v_{ss}|^2 - U_{TF}(v)\right)$$

the Euler-Lagrange equations are

$$f_x - \frac{\partial}{\partial s}(f_{x_s}) - \frac{\partial}{\partial t}(f_{x_t}) + \frac{\partial^2}{\partial s^2}(f_{x_{ss}}) + \frac{\partial^2}{\partial t^2}(f_{x_{tt}}) + \frac{\partial^2}{\partial s \partial t}(f_{x_{st}}) = 0$$

$$f_y - \frac{\partial}{\partial s}(f_{y_s}) - \frac{\partial}{\partial t}(f_{y_t}) + \frac{\partial^2}{\partial s^2}(f_{y_{ss}}) + \frac{\partial^2}{\partial t^2}(f_{y_{tt}}) + \frac{\partial^2}{\partial s\, \partial t}(f_{y_{st}}) = 0 \qquad \text{(A1)}$$

where, for example, $f_{x_{st}}$ denotes $(\partial/\partial x_{st})f$, the derivative with respect to the variable x_{st}.

Equations (A1) hold for a conservative system. In order to dissipate the kinetic energy produced during the motion, an energy dissipation functional can be introduced, such that the deformable curves reach stable equilibrium positions. Using the Raleigh dissipation functional $D^t(v) = \int_\Omega \gamma|v_t|^2\, ds$, where γ is the (constant) damping density, (A1) becomes

$$f_x - \frac{\partial}{\partial s}(f_{x_s}) - \frac{\partial}{\partial t}(f_{x_t}) + \frac{\partial^2}{\partial s^2}(f_{x_{ss}}) + \frac{\partial^2}{\partial t^2}(f_{x_{tt}}) + \frac{\partial^2}{\partial s\, \partial t}(f_{x_{st}}) + \frac{\partial}{\partial x_t}(\gamma|v_t|^2) = 0$$

$$f_y - \frac{\partial}{\partial s}(f_{y_s}) - \frac{\partial}{\partial t}(f_{y_t}) + \frac{\partial^2}{\partial s^2}(f_{y_{ss}}) + \frac{\partial^2}{\partial t^2}(f_{y_{tt}}) + \frac{\partial^2}{\partial s\, \partial t}(f_{y_{st}}) + \frac{\partial}{\partial y_t}(\gamma|v_t|^2) = 0$$

Letting $\Gamma_x(v)$ denote $(\partial/\partial x)U_{TF}(v)$, we calculate

$$f_x = -\frac{1}{2}\Gamma_x(v)$$

$$\frac{\partial}{\partial s}(f_{x_s}) = -\frac{\partial}{\partial s}(\omega_1 x_s)$$

$$\frac{\partial^2}{\partial s^2}(f_{x_{ss}}) = -\frac{\partial^2}{\partial s^2}(\omega_1 x_{ss})$$

$$\frac{\partial}{\partial t}(f_{x_t}) = \mu x_{ttt}$$

$$\frac{\partial^2}{\partial t^2}(f_{x_{tt}}) = \frac{\partial^2}{\partial s\,\partial t}(f_{x_{st}}) = 0$$

Using identical results for f_y, $(\partial/\partial s)(f_{y_s})$, $(\partial^2/\partial s^2)(f_{y_{ss}})$, $(\partial/\partial t)(f_{y_t})$, $(\partial^2/\partial t^2)(f_{y_{tt}})$, and $(\partial^2/\partial s\,\partial t)(f_{y_{st}})$, the Lagrange equations of motion of the deformable curve are

$$\mu x_{tt} + \gamma x_t - \frac{\partial}{\partial s}(\omega_1 x_s) + \frac{\partial^2}{\partial s^2}(\omega_2 x_{ss}) = -\frac{1}{2}\Gamma_x(v)$$

$$\mu y_{tt} + \gamma y_t - \frac{\partial}{\partial s}(\omega_1 y_s) + \frac{\partial^2}{\partial s^2}(\omega_2 y_{ss}) = -\frac{1}{2}\Gamma_y(v) \tag{A2}$$

A.2 Numerical Solution

The solution of system (A2) gives the positions, in space and in time, of the curve v during the motion. We now address the numerical approximation to system (A2), following Terzopoulos [1987b].

The space domain Ω is tessellated into $N + 1$ nodes $\{0, h, 2h, \ldots, Nh\}$, $h = 1/N$. Then, the solutions to the system are the vectors $\mathbf{x}^t = (x(ih, t))_{i=0}^N = (x_i^t)_{i=0}^N$, $\mathbf{y}^t = (y(ih, t))_{i=0}^N = (y_i^t)_{i=0}^N$ at each time t. It is possible to express (A2) as the linear system

$$A\mathbf{x}^t = g_x^t \qquad A\mathbf{y}^t = g_y^t$$

Assuming temporarily that the curve is closed, two initial configurations (initial conditions) are needed. Solving for $\mathbf{x}^t$, given a time step Δt, x_t and x_{tt} are approximated by

$$x_t = \frac{x^t - x^{t-2\Delta t}}{2\Delta t}$$

$$x_{tt} = \frac{x^{t-2\Delta t} - 2x^{t-\Delta t} + x^t}{\Delta t^2}$$

where backward differences are used, since we can only rely on curve positions at previous times. This results in the system of $N + 1$ equations

$$-\frac{\partial}{\partial s}(\omega_1 x_s) + \frac{\partial^2}{\partial s^2}(\omega_2 x_{ss}) + \left(\frac{\mu}{\Delta t^2} + \frac{\gamma}{2\Delta t}\right) x_i^t$$
$$= -\frac{1}{2}\Gamma_x(x_i^{t-\Delta t}, y_i^{t-\Delta t}) + 2\frac{\mu}{\Delta t^2} x_i^{t-\Delta t} - \left(\frac{\mu}{\Delta t^2} - \frac{\gamma}{2\Delta t}\right) x_i^{t-2\Delta t}$$
$$i = 0, \ldots, N \tag{A3}$$

where $(\partial/\partial x)U_{TF}(v)$ is evaluated at time $t - \Delta t$.

The right-hand sides of equations (A3) depend only on prior configurations, and can be evaluated at each time step t, as long as two initial configurations ($\mathbf{x}^{t_0-\Delta t}$ and $\mathbf{x}^{t_0-2\Delta t}$) are given at the initial time t_0. Setting

$$g_x^t = \left[-\frac{1}{2}\Gamma_x(x_i^{t-\Delta t}, y_i^{t-\Delta t}) + 2\frac{\mu}{\Delta t^2} x_i^{t-\Delta t} - \left(\frac{\mu}{\Delta t^2} - \frac{\gamma}{2\Delta t}\right) x_i^{t-2\Delta t}\right]_{i=0}^N$$

and writing

$$A = \left(\frac{\mu}{\Delta t^2} + \frac{\gamma}{2\Delta t}\right) I + K = \alpha I + K$$

(A3) reduces to

$$(\alpha I + K)\mathbf{x}^t = g_x^t$$

where the only unknown is the *stiffness matrix* K determined by the relations between nodes at each time t. Again using finite differences to approximate the spatial derivatives $(\partial/\partial s)(\omega_1 x_s)$ and $(\partial^2/\partial s^2)(\omega_2 x_{ss})$ (see section A.4), we get

$$-\frac{\partial}{\partial s}(\omega_1 x_s) + \frac{\partial^2}{\partial s^2}(\omega_2 x_{ss}) \approx x_{i-2h}\left[\frac{\omega_2(i-h)}{h^4}\right]$$
$$+ x_{i-h}\left[\frac{-2\omega_2(i)}{h^4} - \frac{2\omega_2(i-h)}{h^4} - \frac{\omega_1(i-h)}{h^2}\right]$$
$$+ x_i\left[\frac{\omega_2(i+h)}{h^4} + \frac{4\omega_2(i)}{h^4} + \frac{\omega_2(i-h)}{h^4} + \frac{\omega_1(i)}{2h^2} - \frac{\omega_1(i-h)}{h^2}\right]$$
$$+ x_{i+h}\left[\frac{-2\omega_2(i+h)}{h^4} - \frac{2\omega_2(i)}{h^4} - \frac{\omega_1(i)}{h^2}\right]$$
$$+ x_{i+2h}\left[\frac{\omega_2(i+h)}{h^4}\right]$$

in the case of close curves (i.e., $x_{-1}^t = x_N^t$, $x_{-2}^t = x_{N-1}^t$, $x_{N+1}^t = x_0^t$, $x_{N+2}^t = x_1^t$).

Finally, setting

$$a_i = \frac{\omega_2(i+h)}{h^4}$$

$$b_i = -\frac{2\omega_2(i)}{h^4} - \frac{2\omega_2(i+h)}{h^4} - \frac{\omega_1(i)}{h^2}$$

$$c_i = \frac{\omega_2(i-h)}{h^4} + \frac{4\omega_2(i)}{h^4} + \frac{\omega_2(i+h)}{h^4} + \frac{\omega_1(i-h)}{h^2} + \frac{\omega_1(i)}{h^2}$$

for $i = 1, \ldots, N$, the stiffness matrix K is the pentadiagonal symmetric matrix

$$\begin{bmatrix}
c_0 & b_0 & a_0 & & & & a_{N-1} & b_N \\
b_0 & c_1 & b_1 & a_1 & & & & a_N \\
a_0 & b_1 & c_2 & b_2 & a_2 & & & \\
 & a_1 & b_2 & c_3 & b_3 & a_3 & & \\
\cdots & & & & & & & \\
 & & & a_{N-4} & b_{N-3} & c_{N-2} & b_{N-2} & a_{N-2} \\
a_{N-1} & & & & a_{N-3} & b_{N-2} & c_{N-1} & d_{N-1} \\
b_N & a_N & & & & a_{N-2} & b_{N-1} & c_N
\end{bmatrix}$$

Then, $A = \alpha I + K$ is also pentadiagonal symmetric and can be solved very efficiently in $O(N)$ time and space [Benson and Evans 1977], by factoring A into triangular and diagonal matrixes.

A.3 Inserting Position Discontinuities

Up to now, the deformable curve was assumed to be closed. We now relax this assumption by inserting a discontinuity between node $N + 1$ and node 0.

To insert a position discontinuity at $s = (i - 1/2)h$ between nodes $i - 1$ and i, the stiffness matrix K has to be changed to eliminate relations between nodes $i - 1$ and i. The equations that relate nodes $i - 1$ and i are given by the matrix rows $i - 2$, $i - 1$, i and $i + 1$, and are the following

$$\begin{aligned}
&a_{i-4}x_{i-4} + b_{i-3}x_{i-3} + c_{i-2}x_{i-2} \\
&\quad + \mathbf{b}_{i-2}\mathbf{x}_{i-1} + \mathbf{a}_{i-2}\mathbf{x}_i = g_{i-2} \\
&a_{i-3}x_{i-3} + b_{i-2}x_{i-2} + \mathbf{c}_{i-1}\mathbf{x}_{i-1} \\
&\quad + \mathbf{b}_{i-1}\mathbf{x}_i + a_{i-1}x_{i+1} = g_{i-1} \\
&a_{i-2}x_{i-2} + \mathbf{b}_{i-1}\mathbf{x}_{i-1} + \mathbf{c}_i\mathbf{x}_i \\
&\quad + b_ix_{i+1} + a_ix_{i+2} = g_i \\
&\mathbf{a}_{i-1}\mathbf{x}_{i-1} + \mathbf{b}_i\mathbf{x}_i + c_{i+1}x_{i+1} \\
&\quad + b_{i+1}x_{i+2} + a_{i+1}x_{i+3} = g_{i+1} \qquad \text{(A4)}
\end{aligned}$$

From (A4), we immediately get $b_{i-1} = a_{i-2} = a_{i-1} = 0$, which implies $\omega_2(i - 1) = \omega_2(i) = \omega_1(i - 1) = 0$. This dictates the 9 entries to change in the marix, which are

1. $a_{i-2} = 0$
2. $a_{i-1} = 0$
3. $b_{i-2} = -\dfrac{2\omega_2(i - 2h)}{h^4} - \dfrac{\omega_1(i - 2h)}{h^2}$
4. $b_{i-1} = 0$
5. $b_i = -\dfrac{2\omega_2(i + h)}{h^4} - \dfrac{\omega_1(i)}{h^2}$
6. $c_{i-2} = \dfrac{\omega_2(i - 3h)}{h^4} + \dfrac{\omega_2(i - 2h)}{h^4} + \dfrac{\omega_1(i - 3h)}{h^2} + \dfrac{\omega_1(i - 2h)}{h^2}$
7. $c_{i-1} = \dfrac{\omega_2(i - 2h)}{h^4} + \dfrac{\omega_1(i - 2h)}{h^2}$
8. $c_i = \dfrac{\omega_2(i + h)}{h^4} + \dfrac{\omega_1(i)}{h^2}$
9. $c_{i+1} = \dfrac{4\omega_2(i + h)}{h^4} + \dfrac{\omega_2(i + 2h)}{h^4} + \dfrac{\omega_1(i)}{h^2} + \dfrac{\omega_1(i + h)}{h^2}$

In our implementation, this is used only once, to relax the condition that the curve must be a closed one.

A.4 Development of the Spatial Derivatives

We indicate here how the numerical approximation to spatial derivatives are obtained. Centered, backward and forward differences are mixed in order to obtain a symmetric matrix.

$$\begin{aligned}
\frac{\partial}{\partial s}(\omega_1 x_s) &\approx \frac{\omega_1(i)x_s(i) - \omega_1(i - h)x_s(i - h)}{h} \\
&\approx \frac{\omega_1(i)}{h}\left[\frac{x_{i+h} - x_i}{h}\right] \\
&\quad - \frac{\omega_1(i - h)}{h}\left[\frac{x_i - x_{i-h}}{h}\right] \\
&= x_{i-h}\left[\frac{\omega_1(i - h)}{h^2}\right] \\
&\quad + x_i\left[\frac{-\omega_1(i - h) - \omega_1(i)}{h^2}\right] \\
&\quad + x_{i+h}\left[\frac{\omega_1(i)}{h^2}\right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial s^2}(\omega_2 x_{ss}) &\approx [\omega_2(i + h)x_{ss}(i + h) - 2\omega_2(i)x_{ss}(i) \\
&\quad + \omega_2(i - h)x_{ss}(i - h)] \div h^2 \\
&\approx \frac{\omega_2(i + h)}{h^2}\left[\frac{x_{i+2h} - 2x_{i+h} + x_i}{h^2}\right] \\
&\quad - \frac{2\omega_2(i)}{h^2}\quad\frac{x_{i+h} - 2x_i + x_{i-h}}{h^2} \\
&\quad + \frac{\omega_2(i - h)}{h^2}\left[\frac{x_i - 2x_{i-h} + x_{i-2h}}{h^2}\right]
\end{aligned}$$

$$= x_{i-2h}\left[\frac{\omega_2(i-h)}{h^4}\right] + x_{i-h}\left[\frac{-2\omega_2(i)-2\omega_2(i-h)}{h^4}\right] + x_i\left[\frac{\omega_2(i+h)+4\omega_2(i)+\omega_2(i-h)}{h^4}\right] + x_{i+h}\left[\frac{-2\omega_2(i+h)-2\omega_2(i)}{h^4}\right] + x_{i+2h}\left[\frac{\omega_2(i+h)}{h^4}\right]$$

Thus,

$$-\frac{\partial}{\partial s}(\omega_1 x_s) + \frac{\partial^2}{\partial s^2}(\omega_2 x_{ss})$$

is numerically approximated by

$$x_{i-2h}\left[\frac{\omega_2(i-h)}{h^4}\right] + x_{i-h}\left[\frac{-2\omega_2(i)}{h^4} - \frac{2\omega_2(i-h)}{h^4} - \frac{\omega_1(i-h)}{h^2}\right] + x_i\left[\frac{\omega_2(i+h)}{h^4} + \frac{4\omega_2(i)}{h^4} + \frac{\omega_2(i-h)}{h^4} + \frac{\omega_1(i)}{2h^2} + \frac{\omega_1(i-h)}{h^2}\right] + x_{i+h}\left[\frac{-2\omega_2(i+h)}{h^4} - \frac{2\omega_2(i)}{h^4} - \frac{\omega_1(i-h)}{h^2}\right] + x_{i+2h}\left[\frac{\omega_2(i+h)}{h^4}\right]$$