Differential Morphology and Image Processing

Petros Maragos, *Fellow, IEEE*

Abstract—**Image processing via mathematical morphology has traditionally used geometry to intuitively understand morphological signal operators and set or lattice algebra to analyze them in the space domain. In this paper, we provide a unified view and analytic tools for a recently growing part of morphological image processing that is based on ideas from differential calculus and dynamical systems. This part includes both recent and some earlier ideas on using partial differential or difference equations (PDEs) to model distance propagation or nonlinear multiscale processes in images. We briefly review some nonlinear difference equations that implement discrete distance transforms and relate them to numerical solutions of the eikonal equation of optics. We also review some nonlinear PDEs that model the evolution of multiscale morphological operators and use morphological derivatives. Among the new ideas presented, we develop some general *2-D max/min-sum difference equations* that model the space dynamics of 2-D morphological systems (including the distance computations) and some nonlinear signal transforms, called *slope transforms*, that can analyze these systems in a transform domain in ways conceptually similar to the application of Fourier transforms to linear systems. Thus, *distance transforms* are shown to be bandpass slope filters. We view the analysis of the multiscale morphological PDEs and of the *eikonal* PDE solved via weighted distance tranforms as a unified area in nonlinear image processing, which we call *differential morphology*, and briefly discuss its potential applications to image processing and computer vision.**

I. Introduction

MORPHOLOGIAL image processing has been based traditionally on modeling images as sets or as points in a complete lattice of functions and viewing morphological image transformations as set or lattice operators. Thus, so far, the two classic approaches to analyzing or designing the deterministic systems of mathematical morphology have been (i) *geometry*, by viewing them as image set transformations in Euclidean spaces; and (ii) *algebra*, to analyze their properties using set or lattice theory. Geometry was used mainly for intuitive understanding, and algebra was restricted to the space domain. Despite their limitations, these approaches have produced a powerful and self-contained broad collection of nonlinear image analysis concepts, operators, and algorithms. In parallel to these directions, there is a recently growing part of morphological image processing that is based on ideas from differential calculus and dynamical systems. It combines some early ideas on using simple morphological operations to obtain signal gradients and some recent ideas on using differential equations to model nonlinear multiscale processes or distance propagation in images. In this paper, we present a unified view of the various interrelated ideas in this area, and develop some systems analysis tools in both the space and a (slope) transform domain.

The main tools of morphological image processing are a broad class of nonlinear signal operators, also called morphological systems. These are (simple or complex) parallel and/or serial interconnections of the basic morphological *dilation* $\oplus$ or morphological *erosion* $\ominus$ operations

$$(f \oplus g)(x) \equiv \bigvee_{y \in \mathbf{E}} f(y) + g(x-y) \tag{1}$$

$$(f \ominus g)(x) \equiv \bigwedge_{y \in \mathbf{E}} f(y) - g(y-x) \tag{2}$$

where $\bigvee$ and $\bigwedge$ denote supremum and infimum, and $\mathbf{E} = \mathbf{R}^d$ or $\mathbf{Z}^d$. Compositions of erosions and dilations yield two useful smoothing filters: the *opening* $f \mapsto (f \ominus g) \oplus g$ and *closing* $f \mapsto (f \oplus g) \ominus g$. The above morphological signal operations and their combinations have found a broad range of applications in image processing and computer vision; examples include problems in noise suppression, feature extraction, skeletonization, multiscale analysis, size distributions, segmentation, and shape recognition; see [15], [25], and [43] for surveys and more references.

The abundant use of set and lattice theory as the main analytic tools for morphological operations has been partly due to the relative intractability of these operations using standard ideas from calculus. Among the very few early connections between morphology and calculus were the morphological gradients. Specifically, given a function f: $\mathbf{R}^2 \to \mathbf{R}$, we define its isotropic *morphological sup-derivative* at a point x by

$$\mathcal{M}(f)(x) \equiv \lim_{r \downarrow 0} \frac{(f \oplus rB)(x) - f(x)}{r} \tag{3}$$

where $rB = \{rb\colon b \in B\}$ is a disk B scaled to radius r, and $f \oplus B$ and $f \ominus B$ are *flat* dilations and erosions of f by a planar set B, i.e., special cases of (1) and (2) where g is zero over its support set B as follows:

$$(f \oplus B)(x) \equiv \bigvee_{y \in B} f(x-y)$$

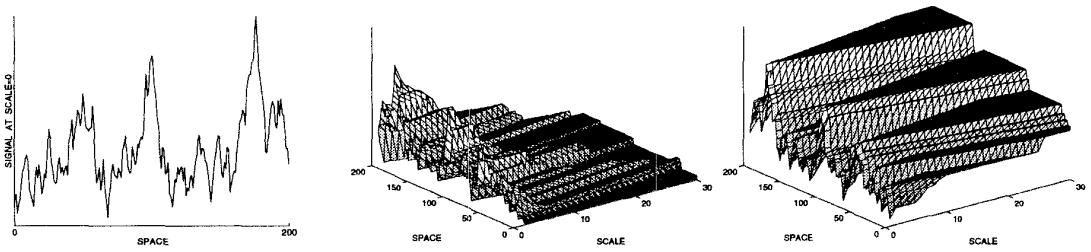
$$(f \ominus B)(x) \equiv \bigwedge_{y \in B} f(x+y).$$

The derivative $\mathcal{M}$ has been used in image analysis for edge detection. For example, $[\mathcal{M}(f) + \mathcal{M}(-f)]/2$ is Beucher's morphological derivative [4], which becomes equal to $||\nabla f||$ when f is differentiable.

Manuscript received January 1, 1995; revised October 18, 1995. This work was supported by the National Science Foundation under Grant MIP-9396301 and Grant MIP-9421677.

The author is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: maragos@gatech.edu).

Publisher Item Identifier S 1057-7149(96)04172-3.



(a) (b) (c)

Fig. 1. (a) Original 1-D signal $f(x)$ at scale $t = 0$. (b) Multiscale erosion. (c) Multiscale dilation of $f(x)$ by a set $B = [-1, 1]$ for scales $t \in [0, 30]$.

A more recent applications area where calculus-based ideas have been used in morphology is that of multiscale analysis. Multiscale signal analysis is a useful and often required framework for many tasks such as feature/object detection, motion detection, and multiband frequency analysis. While the majority of such approaches used so far have been linear, a general understanding arises for the limitations or inability of linear systems to successfully model several image processing problems, and the need grows for developing nonlinear approaches. Recently, some nonlinear partial differential equations (PDEs) have been developed in [8] and [9] to model multiscale morphological operators as dynamical systems evolving in scale-space. To illustrate the basic idea via a 1-D example, let us define the multiscale flat dilation and erosion of a 1-D signal $f(x)$ by the set $B = [-1, 1]$ at scales $t \geq 0$ as the space-scale functions

$$\delta(x, t) = \bigvee_{|y| \leq t} f(x - y);$$
$$\varepsilon(x, t) = \bigwedge_{|y| \leq t} f(x + y).$$

An example of δ and ε is shown in Fig. 1. Then the PDE generating these multiscale flat dilations or erosions is [9],

$$\frac{\partial \delta}{\partial t} = \pm \left| \frac{\partial \delta}{\partial x} \right|$$

where + is for dilation and − yields erosion. To deal with possible discontinuities in these partial derivatives, we can replace them with morphological derivatives (explained later). Recent works related to the above PDE and its 2-D extensions can be found in [1], [26], [46], and [47]. Further, the morphological PDEs are related to broader classes of nonlinear PDEs that have been developed recently to model nonlinear dynamics in image processing. For references, see [2] and the papers in [18]. This rapidly growing interest in using PDEs for image processing is a natural continuation of earlier efforts in computer vision to model several image analysis tasks related to physical phenomena. Examples include shape from shading and optical flow [17] and modeling multiscale analysis via the heat diffusion equation [20].

Multiscale dilations and erosions of binary images can be also obtained via distance transforms. Discrete distance transforms can be implemented fast via 2-D min-sum difference equations [6], [35]. For example, the city-block distance transform of a discrete binary image $f[m, n]$ can be obtained by running the following recursive equation:

$$u[m, n] = \min\left(u[m-1, n] + 1,\ u[m, n-1] + 1,\ f[m, n]\right)$$

in a forward and backward image scan. Distance transforms of continuous-domain binary images can be obtained by propagating wavefronts [5], [22], [41] using ideas similar to optic wave propagation according to Huygen's principle. Finally, resulting from Fermat's principle of least time, the wavefront propagation in geometrical optics is governed by the *eikonal equation*, a nonlinear PDE. This eikonal PDE has found applications in gridless halftoning [31], [42], [48] and shape from shading [17], [48]. Among other methods, the eikonal equation can also be solved using weighted distance transforms [19], [48], which can be implemented via min-sum difference equations with space-varying coefficients.

We have outlined above some close relationships between the morphological derivatives, the PDEs for multiscale morphology, the eikonal PDE of optics, and the difference equations used to implement distance transforms. The unifying theme is a collection of nonlinear differential/difference equations modeling the scale or space dynamics of morphological systems. We call this area *differential morphology*. Whereas classical morphological image processing is based on set and lattice theory, differential morphology offers calculus-based tools and some exciting connections to the physics of wave propagation.

The use of derivatives and differential equations in image processing has been limited so far mostly to *linear* schemes; e.g., the Laplacian used in edge detection and the isotropic diffusion PDE used in multiscale vision. To analyze them there are existing tools both in the space domain (using linear systems theory) as well as in the frequency domain (using Fourier transforms). In this paper, we develop analysis tools for the nonlinear systems used in differential morphology, which have many similarities to the tools used to analyze linear differential schemes. Specifically, in Section II we present analytic methods to determine the output and properties of these nonlinear systems in the spatial domain based on their *impulse response* or on *2-D max-sum or min-sum difference equations* that describe the discrete space dynamics of these systems. Further, in Section III, to understand their behavior in a transform domain—the slope domain—we develop

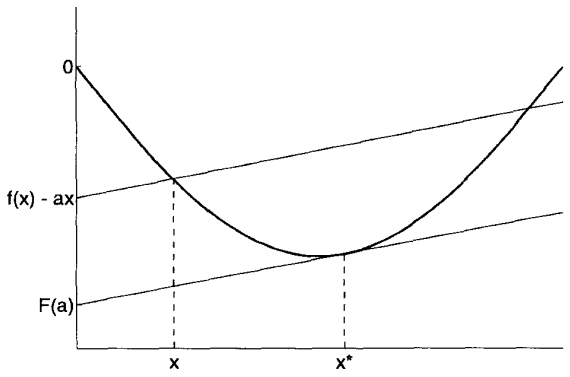


Fig. 2. Convex signal f, its tangent with slope $= \alpha$, and a line parallel to the tangent.

signal transforms, called *slope transforms*, whose properties and application to morphological systems has some striking conceptual similarities with Fourier transforms and their application to linear systems. Our discussion on morphological systems and slope transforms focuses on 2-D signal and systems; a similar work for 1-D signals and systems can be found in [23] and [24]. Despite the wide applicability of morphological systems to image processing and computer vision, so far their analysis has lacked a transform domain. Thus, our work on slope transforms offers a way to fill this gap. Some other works connecting ideas between the slope domain and morphology can be found in [12], [14], and [26]. Section IV outlines the interpretation of distance transforms as bandpass slope filters. Min-sum difference equations, their application to computing discrete distance transforms, and the analysis of both in the slope domain is described in Section V. The eikonal equation, its solution via continuous weighted distance transforms, and an outline of its applications to image processing and vision are discussed in Section VI. In Section VII, we briefly review from [9] some PDEs for multiscale dilations and erosions, and then refine them using morphological derivatives, which also yields a slope domain interpretation. Finally, we conclude with some comments and ideas for further research.

II. 2-D Morphological Systems

Assume signals $f(x)$ defined on a 2-D continuous ($\mathbf{E} = \mathbf{R}^2$) or discrete domain ($\mathbf{E} = \mathbf{Z}^2$) and taking values in $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty, \infty\}$. Their argument may occasionally be written as a two-component vector. In convex analysis [34] and optimization [3] the nonlinear operation $\oplus$ is called *supremal convolution* and an operation closely related to $\ominus$ is the *infimal convolution*

$$(f \,\square\, g)(x) \equiv \bigwedge_y f(y) + g(x-y). \tag{4}$$

Note that $\square$ is closely related to $\ominus$ because $f(x) \,\square\, g(x) = f(x) \ominus [-g(-x)]$. Henceforth, we shall refer to $\oplus$ and $\square$ as the *max-sum* and *min-sum convolution* to distinguish them from the more abstract concept of a dilation and erosion system defined below.

A signal operator or system Ψ: $f \mapsto \Psi(f)$ is called

- **dilation** [44], [15] if $\Psi[\bigvee_i f_i(x)] = \bigvee_i \Psi[f_i(x)]$,
- **erosion** [44], [15] if $\Psi[\bigwedge_i f_i(x)] = \bigwedge_i \Psi[f_i(x)]$,
- **shift-invariant** if $\Psi[f(x-y)] = \Psi(f)(x-y)$,
- **translation-invariant** if $\Psi[c + f(x-y)] = c + \Psi(f)(x-y)$,

for any signals f_i, f, horizontal shifts y and vertical shifts c.

A. Translation-Invariant Systems

Of particular interest in this paper are operators $\mathcal{E}$: $f \mapsto \mathcal{E}(f)$ that are *erosion- and translation-invariant (ETI)* systems. Such systems are shift-invariant and obey an infimum-of-sums superposition.

$$\mathcal{E}\left[\bigwedge_i c_i + f_i(x)\right] = \bigwedge_i c_i + \mathcal{E}[f_i(x)]. \tag{5}$$

Similarly, *dilation- and translation-invariant (DTI)* systems are shift-invariant and obey a supremum-of-sums superposition as in (5) but with $\bigwedge$ replaced by $\bigvee$. Two elementary signals useful for analyzing such systems are the *zero impulse*

$$\mu(x) \equiv \begin{cases} 0, & x = 0 \\ -\infty, & x \neq 0 \end{cases}$$

and the *zero step*

$$\lambda(x) \equiv \begin{cases} 0, & x \geq 0 \\ -\infty, & x < 0 \end{cases}$$

where, if $x = (x_1, x_2) \in \mathbf{R}^2$, then $x \geq 0$ means that both $x_1, x_2 \geq 0$. Occasionally, we shall refer to μ as an "upper" impulse, and to its negated version $-\mu$ as a "lower" impulse. A signal can be represented as a sup or inf of weighted upper or lower impulses; i.e., if f assumes all its values in $\mathbf{R} \cup \{+\infty\}$, then

$$f(x) = \bigwedge_y f(y) - \mu(x-y). \tag{6}$$

If we consider the lower *impulse response* of the ETI system, defined as the system's output when the input is the zero lower impulse, we find [23] that the system's action is equivalent to a min-sum convolution of the input with its lower impulse response,

$$\mathcal{E} \text{ is ETI} \iff \mathcal{E}(f) = f \,\square\, g_\wedge, \quad g_\wedge \equiv \mathcal{E}(-\mu).$$

Similarly, a system $\mathcal{D}$ is DTI iff $\mathcal{D}(f) = f \oplus g_\vee$, where $g_\vee \equiv \mathcal{D}(\mu)$ is the system's upper impulse response. Thus, DTI and ETI systems are uniquely determined in the spatial domain by their impulse responses, which also control their causality and stability [23].

To create a *transform domain* for morphological systems, we first note that the planes $f(x) = \alpha \cdot x + b$ are *eigenfunctions* of any DTI system $\mathcal{D}$ or ETI system $\mathcal{E}$ because

$$\mathcal{D}[\alpha \cdot x + b] = \alpha \cdot x + b + G_\vee(\alpha)$$
$$\mathcal{E}[\alpha \cdot x + b] = \alpha \cdot x + b + G_\wedge(\alpha)$$

where $\alpha \cdot x \equiv \alpha_1 x_1 + \alpha_2 x_2$ for $\alpha = (\alpha_1, \alpha_2)$ and $x = (x_1, x_2)$ in $\mathbf{R}^2$, and

$$G_\vee(\alpha) \equiv \bigvee_x g_\vee(x) - \alpha \cdot x,$$
$$G_\wedge(\alpha) \equiv \bigwedge_x g_\wedge(x) - \alpha \cdot x$$

are the corresponding eigenvalues, called, respectively, the upper and lower *slope response* of the DTI and ETI system. They measure the amount of shift in the intercept of the input hyperplanes with slope vector α. They are also conceptually similar to the frequency response of linear systems.

B. Shift-Varying Systems

For shift-varying dilation and erosion systems, it is still possible to define some type of impulse response and slope response, which are, however, shift-varying. Specifically, consider an erosion system $\mathcal{E}$ that commutes with vertical signal shifts but is horizontally shift-varying, and let $g(x, y) \equiv \mathcal{E}[-\mu(x - y)]$ be its response due to an input lower impulse located at $x = y$. Then, by (6)

$$\mathcal{E}[f(x)] = \bigwedge_y f(y) + g(x, y).$$

If the input is a plane $f(x) = \alpha \cdot x$, the output is

$$\mathcal{E}[\alpha \cdot x] = \alpha \cdot x + G_\wedge(\alpha, x). \tag{7}$$

where the function

$$G_\wedge(\alpha, x) = \bigwedge_y g(x, x - y) - \alpha \cdot y \tag{8}$$

can play the role of a slope response for the shift-varying erosion system. Note that, if $\mathcal{E}$ is *shift-invariant*, then $g(x, y) = g(x - y, 0)$, which we write simply as $g(x - y)$; further, we call $g(x) = \mathcal{E}[-\mu(x)]$ the system's impulse response, and this brings us back to our previous discussion on ETI systems.

III. 2-D Slope Transforms

Viewing the slope response as a signal transform with variable the slope vector α, we define for any 2-D signal $f(x)$ its *upper slope transform* as the 2-D function $F_\vee$: $\mathbf{R}^2 \to \overline{\mathbf{R}}$ defined by

$$F_\vee(\alpha) \equiv \bigvee_{x \in \mathbf{E}} f(x) - \alpha \cdot x \tag{9}$$

and as its *lower slope transform*[1] the function

$$F_\wedge(\alpha) \equiv \bigwedge_{x \in \mathbf{E}} f(x) - \alpha \cdot x. \tag{10}$$

As shown in Fig. 2 for a 1-D signal, $f(x) - \alpha x$ is the intercept of a line with slope α passing from the point $(x, f(x))$ on the signal's graph. Hence, for each α, the lower slope transform of f is the minimum value of this intercept, which occurs when the above line becomes a tangent. For 2-D signals, the tangent line becomes a tangent plane. Examples of lower slope transforms are shown in Fig. 3.

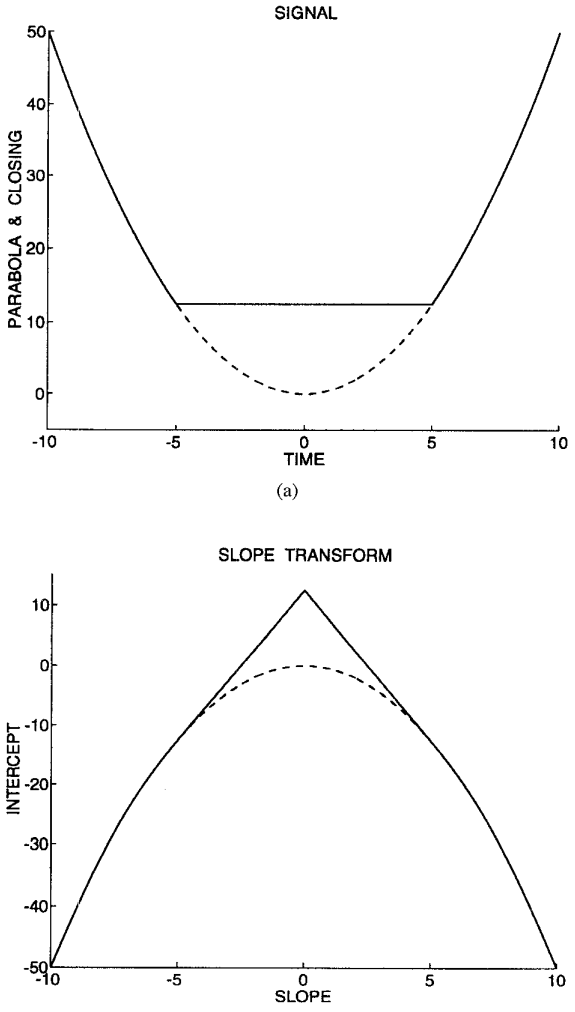


Fig. 3. (a) Convex parabola signal $f(x) = x^2/2$ (in dashed line) and its morphological closing (in solid line) by a flat structuring element $[-5, 5]$. (b) Lower slope transform $F_\wedge(\alpha) = -\alpha^2/2$ of the parabola (in dashed line) and of its closing (in solid line).

In general, a 2-D signal $f(x)$ is covered from above by all the planes $F_\vee(\alpha) + \alpha \cdot x$ whose infimum creates an *upper envelope*

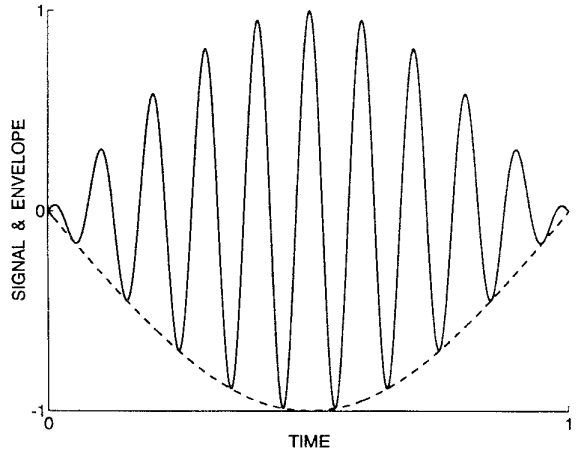
$$\hat{f}(x) \equiv \bigwedge_{\alpha \in \mathbf{R}^2} F_\vee(\alpha) + \alpha \cdot x \tag{11}$$

and $f(x)$ is covered from below by planes $F_\wedge(\alpha) + \alpha \cdot x$ whose supremum creates the *lower envelope*

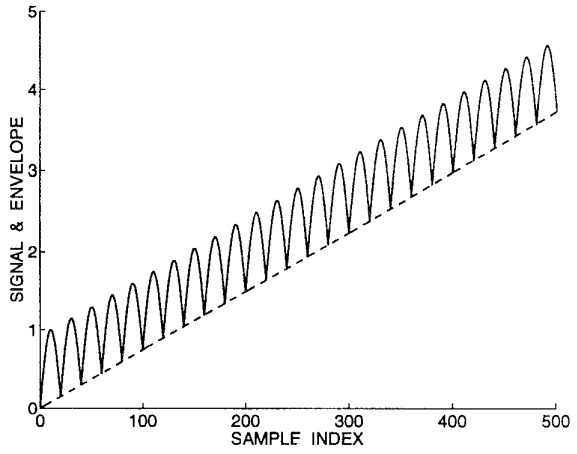
$$\check{f}(x) \equiv \bigvee_{\alpha \in \mathbf{R}^2} F_\wedge(\alpha) + \alpha \cdot x. \tag{12}$$

We view the signal envelopes $\hat{f}(x)$ and $\check{f}(x)$ as the "inverse" upper and lower slope transform of $f(x)$, respectively. Examples are shown in Fig. 4. The next theorem (whose proof

[1] In convex analysis [34], given a convex function h there uniquely corresponds another convex function $h^*(\alpha) = \bigvee_x \alpha \cdot x - h(x)$ called the *Fenchel conjugate* of h. The lower slope transform of h and its conjugate function are closely related since $h^*(\alpha) = -H_\wedge(\alpha)$.



(a)



(b)

Fig. 4. Signals f (in solid lines) and their lower envelopes $\check{f}$ (in dashed lines) obtained via the composition of the lower slope transform and its inverse. The signals are (a) a cosine whose amplitude has been modulated by a slower cosine pulse and (b) the output f of a recursive discrete ETI system generated by the min-sum difference equation $f[n] = \min(-\mu[n], \bigwedge_{1\le k\le 20} f[n-k] + a_k)$, where $a_k = \sin(\pi k/21)$ for $k = 1, \cdots, 20$.

is similar to the proof of the corresponding theorem for 1-D signals given in [24]) shows that slope transforms create convex or concave functions.

Theorem 1: For any signal f: $\mathbf{R}^2, \to \overline{\mathbf{R}}$:

1) $F_\vee$ and $\check{f}$ are convex, whereas $F_\wedge$ and $\hat{f}$ are concave;
2) for all x, $\check{f}(x) \le f(x) \le \hat{f}(x)$;
3) $\hat{f} = f$ if f is concave and upper semi-continuous. Likewise, $\check{f} = f$ if f is convex and lower semicontinuous;
4) $\hat{f}$ is the smallest concave upper envelope of f, and $\check{f}$ is the greatest convex lower envelope of f.

Tables I and II list several properties and examples of the 2-D upper slope transform. The most striking is that (dilation) max-sum convolution in the time/space domain corresponds to addition in the slope domain. Note the analogy with linear systems where linearly convolving two signals in space corresponds to multiplying their Fourier transforms. Very

TABLE I
PROPERTIES OF 2-D SLOPE TRANSFORM

Signal: $f(x)$ *	Transform: $F_\vee(\alpha)$ *
$\bigvee_i c_i + f_i(x)$	$\bigvee_i c_i + F_i(\alpha)$
$f(x - x_0)$	$F(\alpha) - \alpha \cdot x_0$
$f(x) + \alpha_0 \cdot x$	$F(\alpha - \alpha_0)$
$f(rx),\ r \in \mathbf{R}$	$F(\alpha/r)$
$rf(x)\ ,\ r > 0$	$rF(\alpha/r)$
$f(x) \oplus g(x)$	$F(\alpha) + G(\alpha)$
$\bigvee_y f(x) + g(x + y)$	$F(-\alpha) + G(\alpha)$
$f(x) \le g(x)\ \forall x$	$F(\alpha) \le G(\alpha)\ \forall \alpha$
$g(x) = \begin{cases} f(x), & \|x\|_p \le r \\ -\infty, & \|x\|_p > r \end{cases}$	$G(\alpha) = F(\alpha) \square r\|\alpha\|_q$ $1/p + 1/q = 1$
* $x = (x_1, x_2) \in \mathbf{R}^2$	$\alpha = (\alpha_1, \alpha_2) \in \mathbf{R}^2$

TABLE II
EXAMPLES OF 2-D SLOPE TRANSFORMS

Signal: $f(x)$	Transform: $F_\vee(\alpha)$
$\alpha_0 \cdot x$	$-\mu(\alpha - \alpha_0)$
$\alpha_0 \cdot x + \lambda(x)$	$-\lambda(\alpha - \alpha_0)$
$\mu(x - x_0)$	$-\alpha \cdot x_0$
$\lambda(x - x_0)$	$-\alpha \cdot x_0 - \lambda(\alpha)$
$\begin{cases} 0, & \|x\|_p \le r \\ -\infty, & \|x\|_p > r \end{cases}, p \ge 1$	$r\|\alpha\|_q,\ \ \frac{1}{p} + \frac{1}{q} = 1$
$-\alpha_0 \|x\|_p,\ \ \alpha_0 > 0$	$\begin{cases} 0, & \|\alpha\|_q \le \alpha_0 \\ +\infty, & \|\alpha\|_q > \alpha_0 \end{cases}$
$\sqrt{1 - x_1^2 - x_2^2},\ \ x_1^2 + x_2^2 \le 1$	$\sqrt{1 + \alpha_1^2 + \alpha_2^2}$
$-(x_1^2 + x_2^2)/2$	$(\alpha_1^2 + \alpha_2^2)/2$
$-(\|x_1\|^p + \|x_2\|^p)/p\ ,\ p > 1$	$(\|\alpha_1\|^q + \|\alpha_2\|^q)/q$
$\exp(x_1 + x_2)$	$\alpha_1\alpha_2 - \sum_{i=1}^{2} \alpha_i \log \alpha_i$

similar properties also hold for the 2-D lower slope transform, the only differences being the interchange of suprema with infima, concave with convex, and the supremal $\oplus$ with the infimal convolution $\square$.

For *differentiable* signals, the maximization or minimization of the intercept $f(x) - \alpha \cdot x$ involved in both slope transforms can also be done, for a fixed α, by finding its value at the stationary point x^* such that $\nabla f(x^*) = \alpha$. At the point $(x^*, f(x^*))$ the plane becomes tangent to the graph. See Fig. 2 for a 1-D example. This extreme value of the intercept (as a function of the slope α) is the *Legendre transform* of the signal f,

$$F_L(\alpha) \equiv f\left[(\nabla f)^{-1}(\alpha)\right] - \alpha \cdot (\nabla f)^{-1}(\alpha).$$

It is extensively used in mathematical physics [11]. If the signal $f(x)$ is concave or convex and has an invertible gradient, its Legendre transform is single-valued and equal (over the slope regions it is defined) to the upper or lower transform; e.g., see the last four examples in Table II. If a differentiable signal is neither convex nor concave, or if it does not have an invertible gradient, the Legendre transform is multivalued; i.e., $F_L(\alpha)$ is a set of real numbers for each α. This multivalued Legendre transform is defined in [12] as a "slope transform" and is expressed via stationary points; i.e., $F_L(\alpha) = \{f(x^*) - \alpha \cdot x^* \colon \nabla f(x^*) = \alpha\}$. Its properties in [12] are similar to the properties of the upper/lower slope transform,

but there are also some important differences as explained in [23] and [24].

Given a *discrete-domain* 2-D signal $f[m, n]$, we define its lower slope transform by

$$F_\wedge(\alpha_1, \alpha_2) = \bigwedge_{m=-\infty}^{\infty} \bigwedge_{n=-\infty}^{\infty} f[m, n] - (m\alpha_1 + n\alpha_2) \quad (13)$$

and likewise for its upper slope transform using $\bigvee$. The properties of these slope transforms for signals defined on the discrete plane are almost identical to the ones for signals defined on $\mathbf{R}^2$. See [23] and [24] for details.

IV. Distance Transforms and Slope Filters

The distance between two planar points depends on the choice of the norm $||x||$ for a 2-D vector $x = (x_1, x_2)$. Typical norms are

$$||(x_1, x_2)||_p \equiv \begin{cases} (|x_1|^p + |x_2|^p)^{1/p}, & p = 1, 2, 3, \cdots \\ \max(|x_1|, |x_2|), & p = \infty \end{cases}.$$

Their associated 2-D unit balls are

$$B_p \equiv \{x \in \mathbf{R}^2 : ||x||_p \leq 1\}.$$

Thus, given a set $S \subseteq \mathbf{R}^2$, its *distance transform* (also known as its *distance function*) with respect to the $||\cdot||_p$ norm is defined as

$$D_p(S)(x) \equiv \bigwedge_{y \in S^c} ||x - y||_p \quad (14)$$

where $S^c = \mathbf{R}^2 \setminus S$ is the complement of S. It has many applications in image analysis and computer vision. For example, its thresholds at levels $r > 0$ yield the multiscale erosions of S by the balls rB_p of radius r. Further (for $p = 2$), its local maxima provide the points of the skeleton (medial) axis of S [5]. Then, if we consider the $(0, +\infty)$ indicator function of S

$$I_S(x) \equiv \begin{cases} 0, & x \in S \\ +\infty, & x \notin S \end{cases} \quad (15)$$

and the convex conical structuring function

$$g(x) = ||x||_p \quad (16)$$

it follows that

$$D_p(S)(x) = \bigwedge_{y \in \mathbf{R}^2} ||x - y||_p + I_{S^c}(y) = (I_{S^c} \,\square\, g)(x). \quad (17)$$

Hence, the distance transform can be obtained from the min-sum convolution of the indicator function of the set with the conical norm function. This min-sum convolution is equivalent to passing the input signal, i.e., the set's indicator function, through an ETI system with slope response

$$G_\wedge(\alpha) = \bigwedge_{x \in \mathbf{R}^2} ||x||_p - \alpha \cdot x. \quad (18)$$

By using Hölder's inequality

$$|\alpha \cdot x| \leq ||x||_p ||\alpha||_q, \quad \frac{1}{p} + \frac{1}{q} = 1 \quad (19)$$

where q is the conjugate exponent of p, we find that

$$G_\wedge(\alpha) \geq \bigwedge_{x \in \mathbf{R}^2} ||x||_p (1 - ||a||_q).$$

Therefore, $G_\wedge$ is equal to

$$G_\wedge(\alpha) = \begin{cases} 0, & ||\alpha||_q \leq 1 \\ -\infty, & ||\alpha||_q > 1 \end{cases}. \quad (20)$$

That is, the distance transform is the output of an ideal-cutoff slope-selective filter that rejects all input planes whose slope vector falls outside the unit ball with respect to the $||\cdot||_q$ norm, and passes all the rest unchanged.

V. 2-D Min-Plus Difference Equations and Discrete Distance Transforms

A. Min-Sum Difference Equations and Discrete Slope Filters

The space dynamics of a large class of 2-D discrete ETI systems can be described by the following general 2-D *min-sum difference equation*

$$u[m, n] = \left(\bigwedge_{(i,j) \in M_o} a_{ij} + u[m - i, n - j] \right) \wedge \left(\bigwedge_{(k,\ell) \in M_i} b_{k\ell} + f[m - k, n - \ell] \right) \quad (21)$$

which we view as a 2-D discrete nonlinear system, mapping the input signal f to the output u. The *masks* M_o, M_i are pixel coordinate sets that determine which output and input samples will be added with constant weights to form the current output sample. Similarly, the dynamics of DTI systems can be described by max-sum difference equations as in (21) but with $\bigwedge$ replaced by $\bigvee$. For erosion (resp. dilation) systems the useful information in a signal f exists only at points x where $f(x) < +\infty$ [resp. $f(x) > -\infty$]. The vast majority of discrete max/min-sum convolutions $\oplus / \square$ used in applications employs a finite-support structuring element, and they can be modeled by the above max/min-sum difference equations by ignoring the recursive part (i.e., if all $a_{ij} = \pm\infty$). The only exception has been recursive erosions with simple masks that have been used for fast generation of discrete distances or related operations on binary images [6], [21], [35]. We shall show that the min-sum equations with a recursive part correspond to min-sum convolution of the input with an infinite-support structuring element.

As explained in [13] for 2-D linear difference equations, the recursive computability of (21) depends on i) the shape of the output mask $M_o = \{(i, j) : a_{ij} < +\infty\}$ determining which past output samples are involved in the recursion; ii) the boundary conditions, i.e., the locations and values of the output samples $u[n, m]$ that are prespecified as initial conditions; and iii) the scanning order in which the output samples should be computed. We assume boundary conditions of value $+\infty$ and of a shape (dependent on M_o and the scanning order) appropriate so that the difference equation is an ETI system recursively computable. Obviously, $(0, 0) \notin M_o$. The

(a) (b) (c)

Fig. 5. Regions of support of binary slope responses of discrete ETI systems representing (a) the forward pass of the cityblock distance transform; (b) the forward pass of the chamfer (3, 4) distance transform; (c) the chamfer (3, 4) distance transform.

nonrecursive part of (21) represents a min-sum convolution of the input array $f[m, n]$ with the 2-D finite-support structuring function $b[m, n] = b_{mn}$. This is well understood from the existing theory and geometric intuition about erosions by finite-support structuring functions. Thus, we henceforth focus only on the recursive version of (21) by setting $b_{k\ell} = +\infty$ except from $b_{00} = 0$. This yields the autoregressive 2-D min-sum difference equation

$$u[m, n] = \left(\bigwedge_{(i,j)\in M_o} a_{ij} + u[m-i, n-j] \right) \wedge f[m, n]. \tag{22}$$

If $g[m, n] = \mathcal{E}(-\mu[m, n])$ is the impulse response of the corresponding ETI system $\mathcal{E}$: $f \mapsto u$, then

$$u = f \square g.$$

Finding a closed-formula expression for g is generally not possible. However, we can first find the slope response G and then, via inverse lower slope transform, find the impulse response g or its envelope $\check{g}$. (For notational simplicity, we dropped here the subscript $\wedge$ from g and G.) By considering the 2-D finite-support signal of the mask coefficients

$$a[m, n] \equiv \begin{cases} a_{mn}, & (m, n) \in M_o \\ +\infty, & \text{else} \end{cases} \tag{23}$$

we can rewrite (22) as

$$u[m, n] = (u[m, n] \square a[m, n]) \wedge f[m, n]. \tag{24}$$

Applying lower-slope transforms to both sides of (24) yields

$$U_\wedge(\alpha) = [U_\wedge(\alpha) + A_\wedge(\alpha)] \wedge F_\wedge(\alpha) \tag{25}$$

where, for $\alpha = (\alpha_1, \alpha_2)$

$$A_\wedge(\alpha_1, \alpha_2) = \bigwedge_{(i,j)\in M_o} a_{ij} - i\alpha_1 - j\alpha_2. \tag{26}$$

Then, since $U_\wedge(\alpha) = G(\alpha) + F_\wedge(\alpha)$ and assuming that $F_\wedge(\alpha)$ is finite, we have

$$G(\alpha) = [G(\alpha) + A_\wedge(\alpha)] \wedge (0). \tag{27}$$

Thus $G(\alpha) \leq 0$. To find a nontrivial (i.e., different than $-\infty$) solution G, first note that (27) implies

$$G(\alpha) = \begin{cases} 0 \text{ or } -\infty, & \text{if} \quad A_\wedge(\alpha) > 0 \\ \leq 0, & \text{if} \quad A_\wedge(\alpha) = 0. \\ -\infty, & \text{if} \quad A_\wedge(\alpha) < 0 \end{cases}$$

Hence, the largest nontrivial solution of (27) is

$$G(\alpha) = -I_C(\alpha) = \begin{cases} 0, & \alpha \in C \\ -\infty, & \alpha \notin C \end{cases} \tag{28}$$

where C is the convex planar region

$$C \equiv \{(\alpha_1, \alpha_2): i\alpha_1 + j\alpha_2 \leq a_{ij} \; \forall \; (i, j) \in M_o\}. \tag{29}$$

Thus, the system acts as an *ideal-cutoff spatial slope filter* passing all input lower-slope vectors α in the planar region C unchanged and rejecting the rest. The inverse slope transform on G yields the lower envelope $\check{g}$ of the impulse response g. Over short-scale periods g has the shape induced by the sequence $\{a_{ij}\}$. But over scales much longer than the size of the output coefficient mask M_o, g behaves like its lower envelope $\check{g}$. See the 1-D example of Fig. 4(b). Together G and $\check{g}$ can describe the long-scale dynamics of the system. In addition, if g is a plane over its support, then the above analysis is also exact for the short-scale behavior.

Example 1: Let $M_o = \{(0, 1), (0, 1)\}$, and

$$u[m, n] = \min(u[m-1, n] + a_{10}, u[m, n-1] + a_{01}, f[m, n]). \tag{30}$$

Assuming boundary conditions $u[m, n] = +\infty$ if $n < 0$ or $m < 0$ and a bottom-left to top-right scanning order, the impulse response (found by induction) and slope response (shown in Fig. 5(a) for $a_{10} = a_{01} = 1$) are

$$\begin{aligned} g[m, n] &= a_{10}n + a_{01}m - \lambda[m, n] \\ G(\alpha_1, \alpha_2) &= \lambda(a_{10} - \alpha_1, a_{01} - \alpha_2). \end{aligned} \tag{31}$$

Thus, this system acts as a 2-D lowpass spatial slope filter passing all input lower slopes $\alpha_1 \leq a_{10}$ and $\alpha_2 \leq a_{01}$, and rejecting the rest. In this case, $g = \check{g}$ is convex. This example demonstrates that the impulse response of ETI systems described by min-sum difference equations with a recursive part has an infinite support.

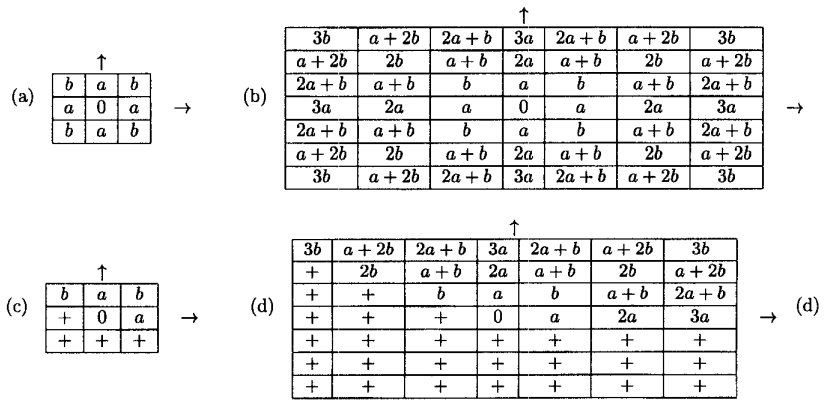


Fig. 6. Coefficient masks and impulse responses of ETI systems associated with computing the (a, b) chamfer distance transform. (a) Local distances within the 3 × 3-pixel unit "disk". (b) Distances from origin by propagating three times the local distances in (a) (also equal to a 7×7-pixel central portion of the infinite impulse response of the overall system associated with the distance transform). (c) Coefficient mask for the min-sum difference equation computing the forward pass for the chamfer distance. (d) 7×7-pixel portion of the infinite impulse response of the system corresponding to the min-sum difference equation computing the forward pass ($+ = +\infty$).

B. Discrete Distance Transforms

Discrete distance transforms, especially of binary images, have been useful for a variety of image analysis tasks including multiscale erosions/dilations, skeletonization, segmentation, smoothing, and path finding [6], [21], [35], [36], [49], [50]. Their early theory was developed in [35] and [36] based on either sequential or parallel operations. Later generalizations including improved approximations to the Euclidean distance were developed by Borgefors [6] based on "chamfer" metrics. The general chamfer distance transform is obtained by *propagating local distances* within a small, e.g., a 3 × 3-pixel neighborhood. The fixed local distances shown in Fig. 6(a) have the constraints $0 < a < b < 2a$. Given a subset X of the discrete plane $\mathbf{Z}^2$, the local distance propagation will yield upon completion its (a, b) chamfer distance transform

$$D_{a,b}(X)[m, n] \equiv \bigwedge_{(i,j)\in X^c} ||(m-i, n-j)||_{a,b} \tag{32}$$

where $||\cdot||_{a,b}$ is the (a, b) *chamfer norm*

$$||(m, n)||_{a,b} \equiv \max(|m|, |n|)a + \min(|m|, |n|)(b-a).$$

This function is very similar to the distance transform of sets in the real plane, but with the difference that the $||\cdot||_p$ norms are replaced by the chamfer norms.

The local distance propagation was implemented in [6] and [35] sequentially via simple recursive min-sum difference equations. We shall show that these equations correspond to ETI systems with infinite impulse responses and binary slope responses. The distance propagation can also be implemented in parallel via nonrecursive min-sum equations, which correspond to ETI systems with finite impulse responses, as explained next.

Example 2: Consider a 2-D min-sum autoregressive difference equation with output mask $M_o = \{(0, 1), (0, 1), (1, 1), (-1, 1)\}$ and coefficients as in Fig. 6(c) as follows:

$$u[m, n] = \min(u[m-1, n] + a, u[m, n-1] + a, u[m-1, n-1] + b, u[m+1, n-1] + b, f[m, n]). \tag{33}$$

Let X be a set representing a discrete binary image and consider the $(0, +\infty)$ binary image array

$$f = I_{X^c}$$

with 0 marking background/source pixels of X and $+\infty$ marking foreground/object pixels. Then consider the following distance transformation of f obtained in two passes: During the *forward pass* (33) is recursively run over $f[m, n]$ in a bottom-to-top left-to-right scanning order. The forward pass mapping $f \mapsto u$ is an ETI system with an infinite impulse response (found via induction)

$$g_1[m, n] = \begin{cases} ||(m, n)||_{a,b}, & m+n \geq 0,\ n \geq 0 \\ +\infty, & \text{else} \end{cases} \tag{34}$$

A truncated version of g_1 is shown in Fig. 6(d). The slope response $G_1(\alpha)$ of this ETI system is equal to the indicator function of the region shown (for $a = 3$, $b = 4$) in Fig. 5(b). During the *backward pass* a similar recursion as in (33) but in the opposite scanning order and using as output mask the reflected version of Fig. 6(c) is run over the previous result $u[m, n]$ to yield a signal $d[m, n]$ which is the final distance transform of $f[m, n]$. The backward pass mapping $u \mapsto d$ is an ETI system with an infinite impulse response $g_2[m, n] = g_1[-m, -n]$ and with a slope response $G_2(\alpha) = G_1(-\alpha)$.

Since min-sum convolution is an associative operation, the distance transform mapping $f \mapsto d$ is an ETI system equivalent to a min-sum convolution with an infinite impulse response $g = g_1 \square g_2$ as follows:

$$\begin{aligned} d &= (f \square g_1) \square g_2 \\ &= f \square (g_1 \square g_2) \\ &= f \square g. \end{aligned} \tag{35}$$

The overall slope response

$$\begin{aligned} G(\alpha) &= G_1(\alpha) + G_2(\alpha) \\ &= G_1(\alpha) + G_1(-\alpha). \end{aligned} \tag{36}$$

of this distance transform ETI system is the indicator function of a bounded convex region shown in Fig. 5(c) for $a = 3$, $b =$

4. Further, by using induction on (m, n) and symmetry, we find that $g = g_1 \square g_2$ is equal to

$$g[m, n] = ||(m, n)||_{a,b}. \tag{37}$$

A truncated version of g is shown in Fig. 6(b). Note that $g[m-i, n-j]$ is the (a, b) chamfer distance between two pixels (m, n) and (i, j). Further

$$\begin{aligned}(f \square g)[m, n] &= \bigwedge_{(i,j)\in \mathbf{Z}^2} I_{X^c}(i, j) + g[m-i, n-j] \\ &= \bigwedge_{(i,j)\in X^c} g[m-i, n-j].\end{aligned}$$

Thus, our analysis above has provided an alternative proof (using ETI systems theory) that the two-pass computation via recursive min-sum difference equations whose coefficients are the local chamfer distances yields the (a, b) chamfer distance transform of [6] as follows:

$$d = I_{X^c} \square g = D_{a,b}(X). \tag{38}$$

Two special cases are the two well-known cityblock and chessboard discrete distances [35]. The *cityblock* distance transform is obtained using $a = 1$ and $b = +\infty$, or equivalently $b = 2$, i.e., using the five-pixel diamond as the unit "disk." It is an ETI system with impulse response $g[m, n] = |m| + |n|$ and slope response the indicator function of the unit square $\{\alpha: ||\alpha||_\infty = 1\}$. The difference equation and impulse response for the forward pass are given by (30) and (31) for $a_{10} = a_{01} = 1$. Similarly, the *chessboard* distance transform is obtained using $a = b = 1$. It is an ETI system with impulse response $g[m, n] = \max(|m|, |n|)$ and slope response the indicator function of the unit diamond $\{\alpha: ||\alpha||_1 = 1\}$.

Example 3: The (a, b) distance transform can be implemented alternatively using parallel operations. Namely, let

$$g_0[m, n] \equiv \begin{cases} g[m, n], & |m|, |n| \le 1 \\ +\infty, & \text{else} \end{cases} \tag{39}$$

be the 3×3-pixel central portion of g in (37). It can be shown (via induction) that, the kth-fold min-sum convolution of g_0 with itself yields g in the limit

$$g = \lim_{k\to\infty} \underbrace{(g_0 \square g_0) \cdots \square g_0}_{k \text{ times}}. \tag{40}$$

Figure 6(b) shows the intermediate result for $k = 3$ iterations. Similar finite decompositions of discrete conical functions into min-sum convolutions of smaller kernels have been studied in [45]. Consider now the nonautoregressive min-sum difference equation

$$d^{(k)}[m, n] = \bigwedge_{i=-1}^{1} \bigwedge_{j=-1}^{1} g_0[i, j] + d^{(k-1)}[m-i, n-j] \tag{41}$$

run iteratively for $k = 1, 2, \cdots$ starting from $d^{(0)} = f$. Each iteration is equivalent to a min-sum convolution of the previous result with a finite impulse response equal to g_0. By iterating these local distances to the limit, the final distance transform is obtained as follows:

$$d = \lim_{k\to\infty} d^{(k)}. \tag{42}$$

In practice, when the input image f has finite support, the number of required iterations is finite and bounded by the image diameter.

Thus, in general, the 2-D discrete distance transforms are ETI systems whose slope responses are indicator functions of symmetric polygonal approximations to the disk in the slope plane and whose impulse responses are approximations to space cones. In the space domain they are min-sum convolutions with infinite conical functions, whereas in the slope domain they are 2-D ideal-cutoff bandpass slope filters.

C. Gray-Weighted Distance Transforms

Given a gray-level image $f[m, n]$ and a set of reference points (the "sources") S, the *gray-weighted distance transform* [38] finds at each pixel $p = [m, n]$ the smallest sum of values of f over all possible paths connecting p to the sources S. It can also be viewed as a procedure of finding paths of minimal "cost" among nodes of a weighted graph. It can be computed by running a min-sum difference equation like the one implementing the (a, b) chamfer distance transform but with spatially-varying coefficients proportional to the local input image values, as follows:

$$\begin{aligned}u_{i+1}[m, n] = \min(&u_i[m-1, n] + af[m, n], \\ &u_i[m, n-1] + af[m, n], \\ &u_i[m-1, n-1] + bf[m, n], \\ &u_i[m+1, n-1] + bf[m, n], u_i[m, n]).\end{aligned} \tag{43}$$

Starting from $u_0 = I_S$, a sequence of functions u_i is iteratively computed by running (43) over the image domain in a forward scan for even i, whereas for odd i an equation as in (43) but with a reflected output mask is run in a backward scan. In the limit $i \to \infty$ the final gray-weighted distance transform u_∞ is obtained. In practice, this limit is reached after a finite number of passes. The final transform depends on both the sources and the gray values. Fig. 7 shows examples of weighted distance transforms computed by iterating (43) until convergence, which was reached at 45 and 17 iterations when f was the negated image of Fig. 7(a) and (d), respectively, and sources were at the four-line image boundary.

If the image f is binary and the sources are placed over all points of its closed boundary, then the gray-weighted distance transform reduces to the previous unweighted distance transform.

VI. Eikonal PDE and Distance Propagation

A. Eikonal Equation and Geometrical Optics

The main postulate of geometrical optics [7], [39] is Fermat's *principle of least time*. For notational simplicity, let us assume a 2-D, i.e., planar medium with (possibly space-varying) refractive index $\eta(x, y) = c_0/c(x, y)$ defined as the

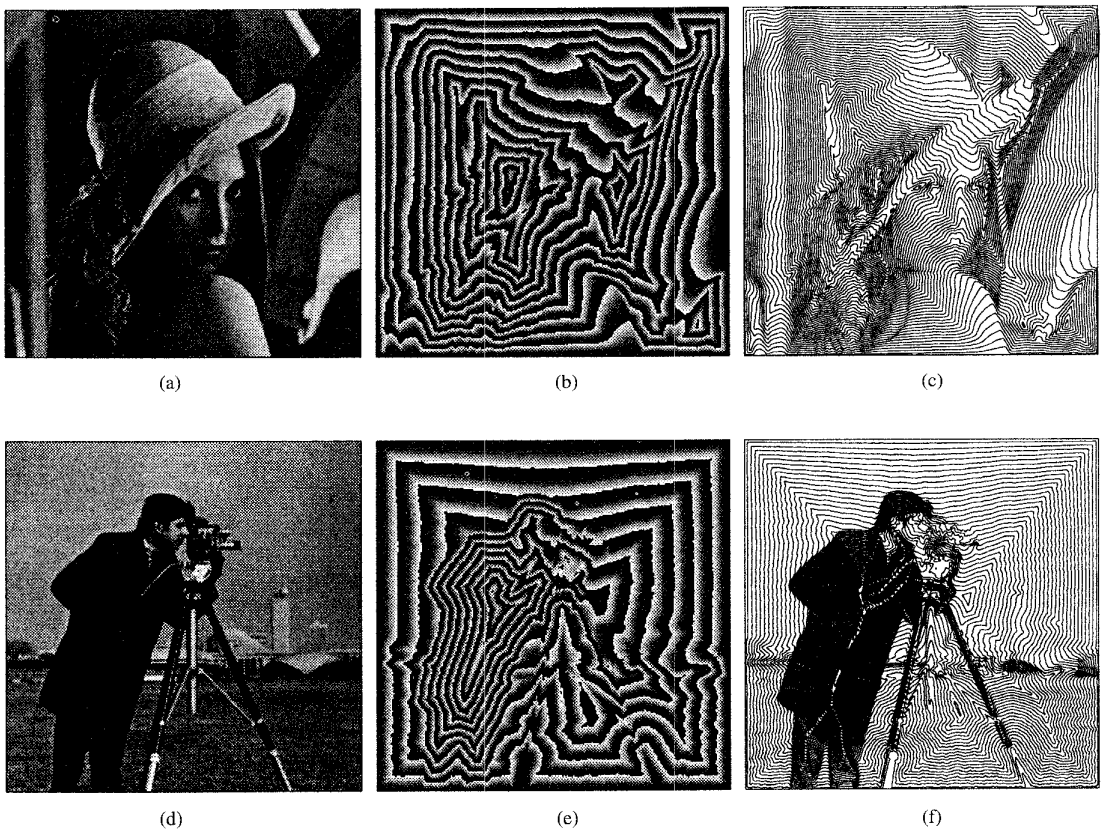


Fig. 7. (a) Original gray image I (Lena) with intensity values in [1, 256]. (b) Gray-weighted distance transform of $257 - I$ w.r.t. the (5, 7) chamfer metric using as sources the four-image boundary lines and displayed modulo 1440. (c) 50 isolevel contour lines of the distance transform of $257 - I$. (d) Original gray image (Cameraman). (e) and (f) show same experiments as in (b) and (c) but when I is the image of (d).

ratio of the speed c_0 of light in free space divided by its speed $c(x, y)$ in the medium. Given two points A and B in such a medium, the optical path length along a ray trajectory Γ_{AB} (parameterized by ℓ) between points A and B is

$$\text{optical path length} = \int_{\Gamma_{AB}} \eta\left(\Gamma_{AB}(\ell)\right) d\ell = c_0 T(\Gamma_{AB}) \tag{44}$$

where $d\ell$ is the differential length element along this trajectory, and $T(\Gamma_{AB})$ is the time required for the light to travel this path. Then, Fermat's principle states that light will choose a path between A and B that minimizes the optical path length. While in homogeneous media, where η is constant, Fermat's principle reduces to simply choosing the shortest distance and, hence, the straight line connecting A and B, in heterogeneous media the path of least time is generally not straight. For example, Fig. 8(a) shows a medium with piecewise-constant index and two points A and B in two regions each with a different but constant index; there, the path of least time is piecewise linear.

An alternative viewpoint of geometrical optics is to consider the scalar function $\Phi(x, y)$, called the *eikonal*, whose isolevel contours are normal to the rays. Thus, the eikonal's gradient $||\nabla\Phi||$ is parallel to the rays. It can be shown [7] using calculus of variations that Fermat's principle is equivalent to the following PDE:

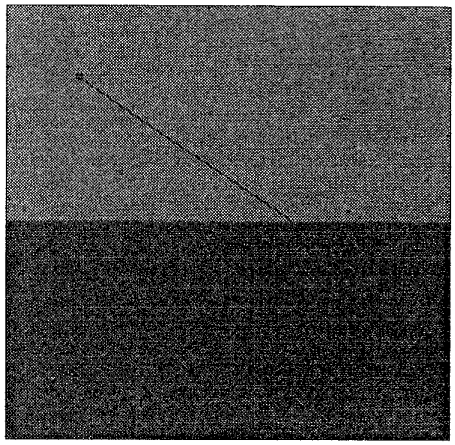
$$||\nabla\Phi(x, y)|| = \sqrt{\left(\frac{\partial\Phi}{\partial x}\right)^2 + \left(\frac{\partial\Phi}{\partial y}\right)^2} = \eta(x, y) \tag{45}$$

called the *eikonal equation*. Thus, the minimal optical path length between two points located at A and B is

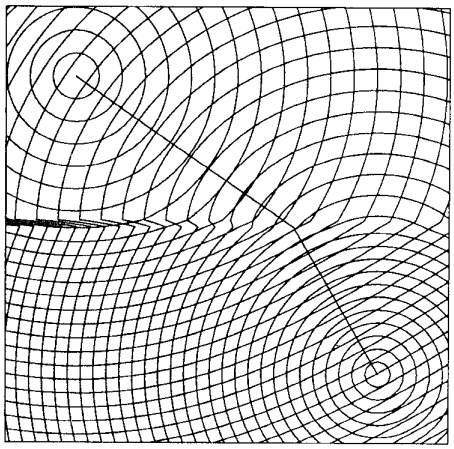
$$\Phi(B) - \Phi(A) = \inf_{\Gamma_{AB}} \int_{\Gamma_{AB}} \eta(\Gamma_{AB}(\ell))\, d\ell. \tag{46}$$

B. Wavefront and Distance Propagation

Assume an optical wave propagating in a 2-D medium of index $\eta(x, y)$ at wavelengths much smaller than the image objects, so that ray optics can approximate wave optics. Then, the eikonal Φ of ray optics is proportional to the phase of the wavefunction. Hence, the isolevel contour lines of Φ, i.e., $\Phi(x, y) = \text{constant}$, are the wavefronts. Assuming that at time $t = 0$ there is an initial wavefront at a set of source points S_i, we can trace the wavefront propagation using Huygen's *envelope construction*: Namely, if we dilate the points $P = (x, y)$ of the wavefront curve at a certain time t with circles of infinitesimal radius $c(x, y)\, dt$, the envelope of these circles yields the wavefront at time $t + dt$. If $T(P)$ is the time required for the wavefront to arrive at P from the



(a)



(b)

Fig. 8. (a) Two point light sources in two media with different refractive indexes, and the path of least propagation time (the lighter region corresponds to $\eta = 1$, and the darker to $\eta = 1.7$). (b) Wavefronts originating from the point sources and propagating with velocities $c = c_0/\eta$. (Adapted from Fig. 5 of [19].)

sources, then by (44) and (46)

$$\Phi(P) = c_0 T(P) = \inf_i \left\{ \inf_{\Gamma_{S_i P}} \int_{\Gamma_{S_i P}} \eta(\Gamma_{S_i P}(\ell))\, d\ell + \Phi(S_i) \right\}. \quad (47)$$

Thus we can equate the eikonal $\Phi(x, y)$ to the *weighted distance function* between a point (x, y) and the sources along a path of minimal optical length and also see Φ as proportional to the wavefront arrival time $T(x, y)$. See Fig. 8(b) for an example of wavefront propagation through a heterogeneous medium. This weighted distance transform uses the Euclidean norm for horizontal path distances and the index values $\eta(x, y)$ as vertical weights. Viewing the solution to the eikonal PDE as a weighted distance function dates back as early as [22], was discussed in [48], and was proven in [37].

In a *homogeneous* medium of index $\eta \equiv 1$ having a planar support X and sources located all over its boundary, the wavefronts propagate at constant velocity c_0 and can be obtained from the multiscale dilations of the initial front by disks whose radii are proportional to the time of propagation [5]. The eikonal function Φ is then simply equal to the unweighted distance transform of X, since the paths of minimal optical length will be straight lines. However, in a *heterogeneous* medium of planar support X, the wavefronts result from successive dilations by small disks of spacially varying radii, and the eikonal Φ becomes a weighted distance function whose values depend on the geometry of X, the geometry of the sources, and the index field η.

C. Applications

Three important problems in image processing and computer vision are shape from shading [17], gridless image halftoning [31], and watershed image segmentation [4]. Recently, all these problems have been approached from the viewpoint of solving the eikonal equation driven by different fields of refractive indices.

The goal of *shape from shading* is to find a 3-D surface $z(x, y)$ which, when illuminated, yields a 2-D image matching a given intensity image $I(x, y)$. Assuming a Lambertian image surface, the unknown function z satisfies an eikonal equation [17], [48] as follows:

$$||\nabla z(x, y)|| = \sqrt{\frac{1}{I^2(x, y)} - 1}. \quad (48)$$

Numerical solutions to the eikonal, or equivalently to finding weighted distance transforms, solving shape from shading problems have been given in [48] using discrete distance algorithms implemented via queues and in [19] using curve evolution implemented via the numerical algorithms of [30].

An efficient morphological approach to *image segmentation* is the *watershed* [4], which transforms an image f to the crest lines separating adjacent catchment basins that surround regional minima or other "marker" sets of feature points. From the work in [27]–[29] and [33], it has been established that computing the watershed is equivalent to finding a skeleton by influence zones with respect to a weighted distance function that uses the regional minima as sources and $||\nabla f||$ as the field of indices. (If other markers different than the minima are to be used as sources, then the homotopy of the function must be modified to impose these markers as minima.) Hence, it has been proposed in [27]–[29] to use existing efficient watershed algorithms for finding the solution to the eikonal.

Another interesting application is *gridless image halftoning*. Inspired by the use of the eikonal function's contour lines [42] for visually perceiving an intensity image $I(x, y)$, the work in [48] and especially in [31], [32] attempts to solve the PDE

$$||\nabla \Phi(x, y)|| = \text{const} - I(x, y) \quad (49)$$

and create a binary *gridless* halftone version of $I(x, y)$ as the union of the isolevel curves of the eikonal function $\Phi(x, y)$. The larger the intensity value $I(x, y)$, the smaller the local density of these contour lines in the vicinity of (x, y). To illustrate the power of this approach, shown in Figs. 7 and 9, we have computed numerical solutions of the eikonal PDE in (49) using discrete weighted distance

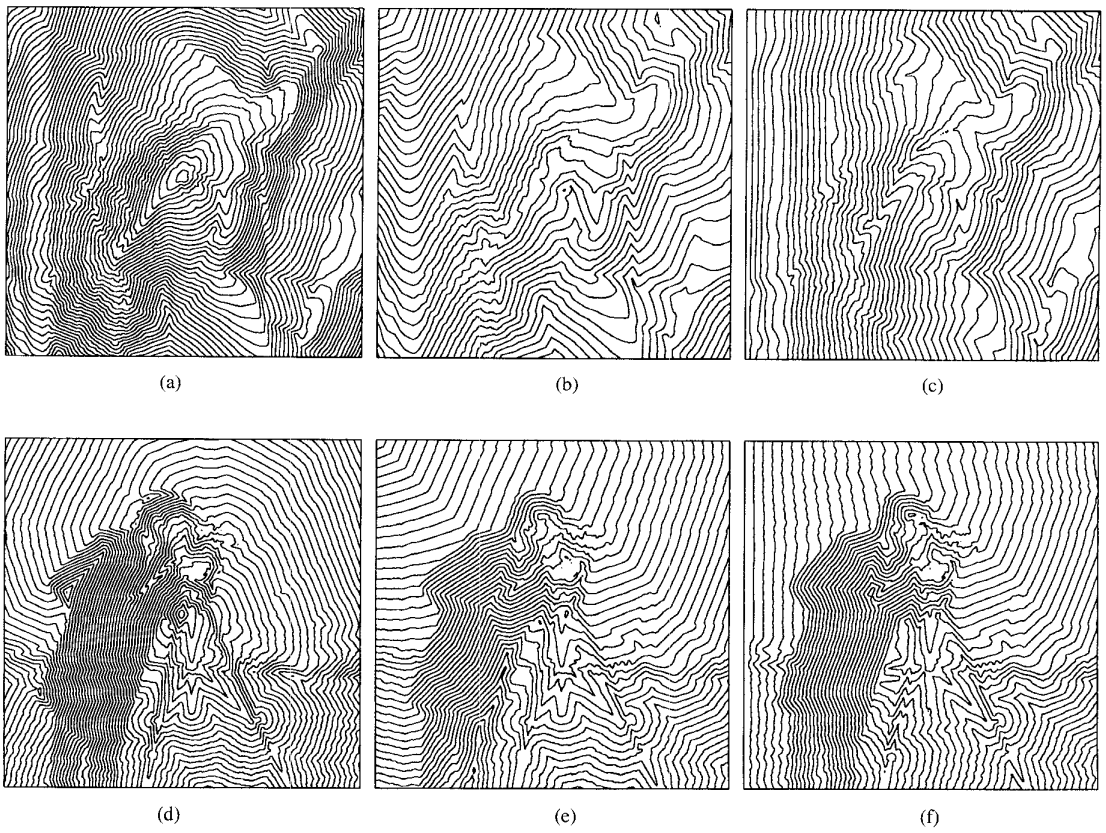


Fig. 9. Gridless halftoning of gray-level images I with intensities values in [1, 256] via plotting 50 isolevel contour lines of the gray-weighted distance transform of $257 - I$ for various locations of the light sources. (a), (b), and (c) show results when I is the Lena image of Fig. 7(a), and the source is: (a) the center point of the image; (b) the top left image corner point; and (c) the left image boundary. (d), (e), and (f) show same experiments as in (a), (b), and (c) but when I is the Cameraman image of Fig. 7(d).

transforms, implemented[2] by iterating the min-sum difference (43) with $a = 5$, $b = 7$. Extending the discussion in [48], the rationale for such a solution is that, away from the sources, this difference equation mapping $f \mapsto u$ corresponds to

$$\bigvee_{(i,j)\in B} \frac{u[m,n] - u[m-i, n-j]}{a_{ij}} = f[m, n] \qquad (50)$$

where a_{ij} are the chamfer weights inside B equal to the union of the output mask and its reflection. The left side of (50) is the (weighted) discrete morphological derivative $\mathcal{M}(-u)$ equal to $u - u \ominus B$, where horizontal distances are weighted by a_{ij}. Thus, since in the continuous case $\mathcal{M}(-u) = ||\nabla u||$, (50) is an approximation of the eikonal. In fact, as established in [27], it is possible to recover a digital image u from its half morphological gradient $u - u \ominus B$ using weighted distance transforms if one places the source points in each regional minimum of u. The experimental results shown in Fig. 9 for various light source locations verify the conclusion of [32] that this eikonal-based approach to gridless halftoning is indeed very promising and can simulate various artistic effects.

[2]Alternative implementations of weighted distance transforms include queue-based algorithms using special data structures such as "buckets" with priorities [49, ch. 1] or hierarchical queues [27].

VII. Multiscale Morphological PDE's

Most of the work in multiscale image analysis involves obtaining the multiscale linear convolutions

$$\gamma(x, y, t) \equiv \int\int_{\mathbf{R}^2} \frac{f(x - v, y - w)}{\sqrt{4\pi t}} \cdot \exp\left(-\frac{v^2 + w^2}{4t}\right) dv\, dw$$

of the original image $f(x, y)$ with a 2-D Gaussian function whose variance $(2t)$ is proportional to scale t. The popularity of this approach is due to its linearity and the fact that the multiscale function γ can be generated from the isotropic heat *diffusion equation*

$$\frac{\partial \gamma}{\partial t} = \frac{\partial^2 \gamma}{\partial x^2} + \frac{\partial^2 \gamma}{\partial y^2} \qquad (51)$$

with initial condition $\gamma(x, y, 0) = f(x, y)$. The big disadvantage of the Gaussian multiscale approach is the fact that linear smoothers blur and shift important image features, e.g., edges. In contrast, morphological smoothing filters, such as openings and closings, can smooth while preserving important image features and correspond to simple filtering operations. Examples are shown in Fig. 10. So far the vast majority of implementations of multiscale morphological filtering had been

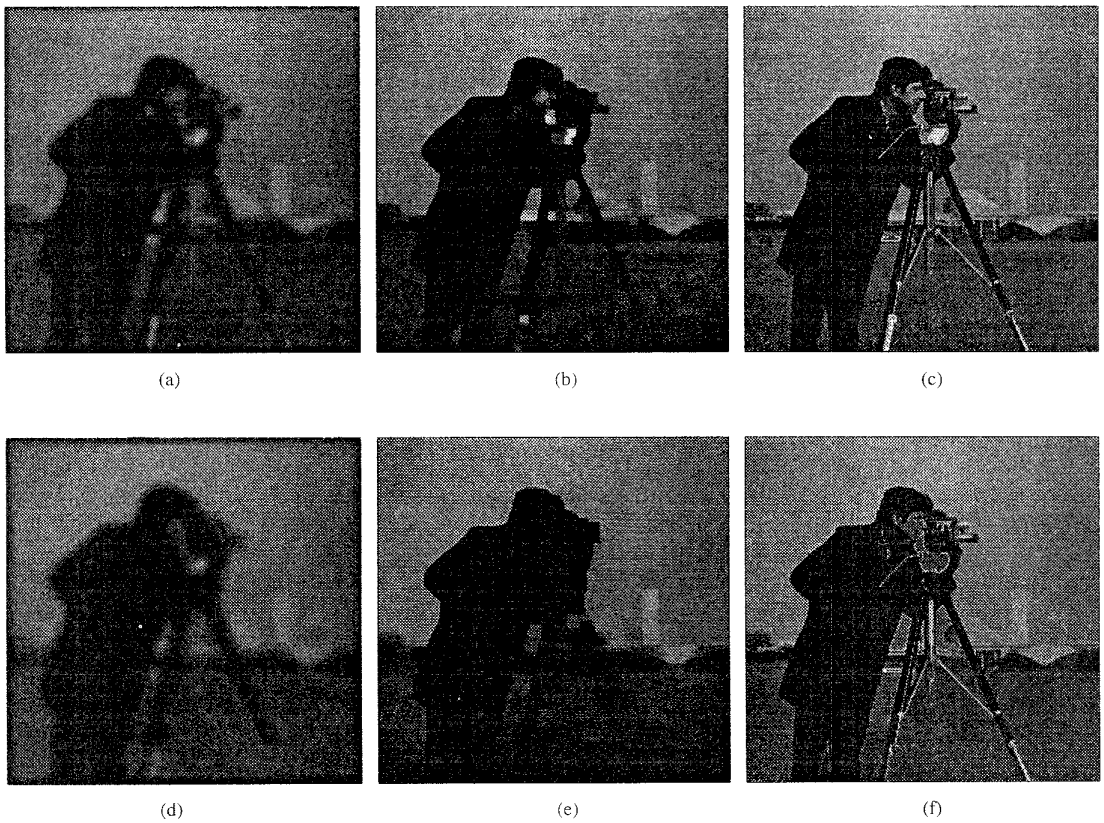


Fig. 10. Multiscale smoothings of the original Cameraman Fig. 7(d) at scales $t = 4$ for (a), (b), and (c) and $t = 8$ for (d), (e),and (f). The smoothers are: (a), (d) Linear convolutions with Gaussians of standard deviation $\sqrt{2t}$; (b), (e) morphological clos-opening by a square of $(t + 1) \times (t + 1)$ pixels; (c), (f) clos-opening by reconstruction [40] by a square of $(t + 1) \times (t + 1)$ pixels.

discrete. In [8] and [9] nonlinear PDEs were developed for generating continuous-scale morphological erosions/dilations and openings/closings. Herein we shall limit our discussion to PDEs for erosions and dilations by *flat* structuring functions whose support sets B_p are the unit balls corresponding to the norms $||\cdot||_p$.

The *multiscale dilation* and *erosion* of f: $\mathbf{R}^2 \rightarrow \mathbf{R}$ by a compact convex set B_p at scale t are defined as the space-scale functions

$$\delta(x, y, t) \equiv (f \oplus tB_p)(x, y)$$
$$\varepsilon(x, y, t) \equiv (f \ominus tB_p)(x, y)$$

where $\delta(x, y, 0) = \varepsilon(x, y, 0) = f(x, y)$. The multiscale morphological PDEs describe the evolution of the multiscale dilation and erosion functions in scale-space by relating the infinitesimal rate of change along the scale direction t to the infinitesimal rates of change along the spatial directions x, y. Thus for the multiscale dilation the goal is to find a PDE for the following evolution equation:

$$\frac{\partial \delta}{\partial t}(x, y, t) = \lim_{r \downarrow 0} \frac{\delta(x, y, t + r) - \delta(x, y, t)}{r}. \tag{52}$$

Given the semigroup property of multiscale dilations we can write the above as

$$\frac{\partial \delta}{\partial t}(x, y, t) = \lim_{r \downarrow 0} \frac{\delta(x, y, t) \oplus rB_p - \delta(x, y, t)}{r}. \tag{53}$$

Consider now the general *morphological sup-derivative*[3] $\mathcal{M}(f)$ of f defined as in (3) by replacing the disk B with the general ball B_p. Then, it follows that the infinitesimal rate of change of δ along the scale direction t is equal to its partial morphological sup-derivative in the spatial domain, as follows:

$$\frac{\partial \delta}{\partial t} = \mathcal{M}_{xy}(\delta) \tag{54}$$

where the partial sup-derivative is defined as

$$\mathcal{M}_{xy}(\delta)(x, y, t) \equiv \lim_{r \downarrow 0} \frac{\bigvee_{||(a,b)||_p \leq r} \delta(x + a, y + b, t) - \delta(x, y, t)}{r}. \tag{55}$$

The general PDE of (54) can deal with possible discontinuities in the signal derivatives.

If δ is *differentiable* at (x, y, t), then

$$\delta(x + a, y + b, t) - \delta(x, y, t) = a\frac{\partial \delta}{\partial x} + b\frac{\partial \delta}{\partial y} + o(||(a, b)||)||(a, b)|| \tag{56}$$

[3]The limit in the definition of the morphological derivative may not exist. For example, if $f(x) = |x|^p$ and $B = [-1, 1]$, then $\mathcal{M}(f)(0)$ is equal to 1 if $p = 1$, 0 if $p = 2$, and $+\infty$ if $p = 1/2$. If the limit does not exist, the "lim" could be replaced by "limsup."

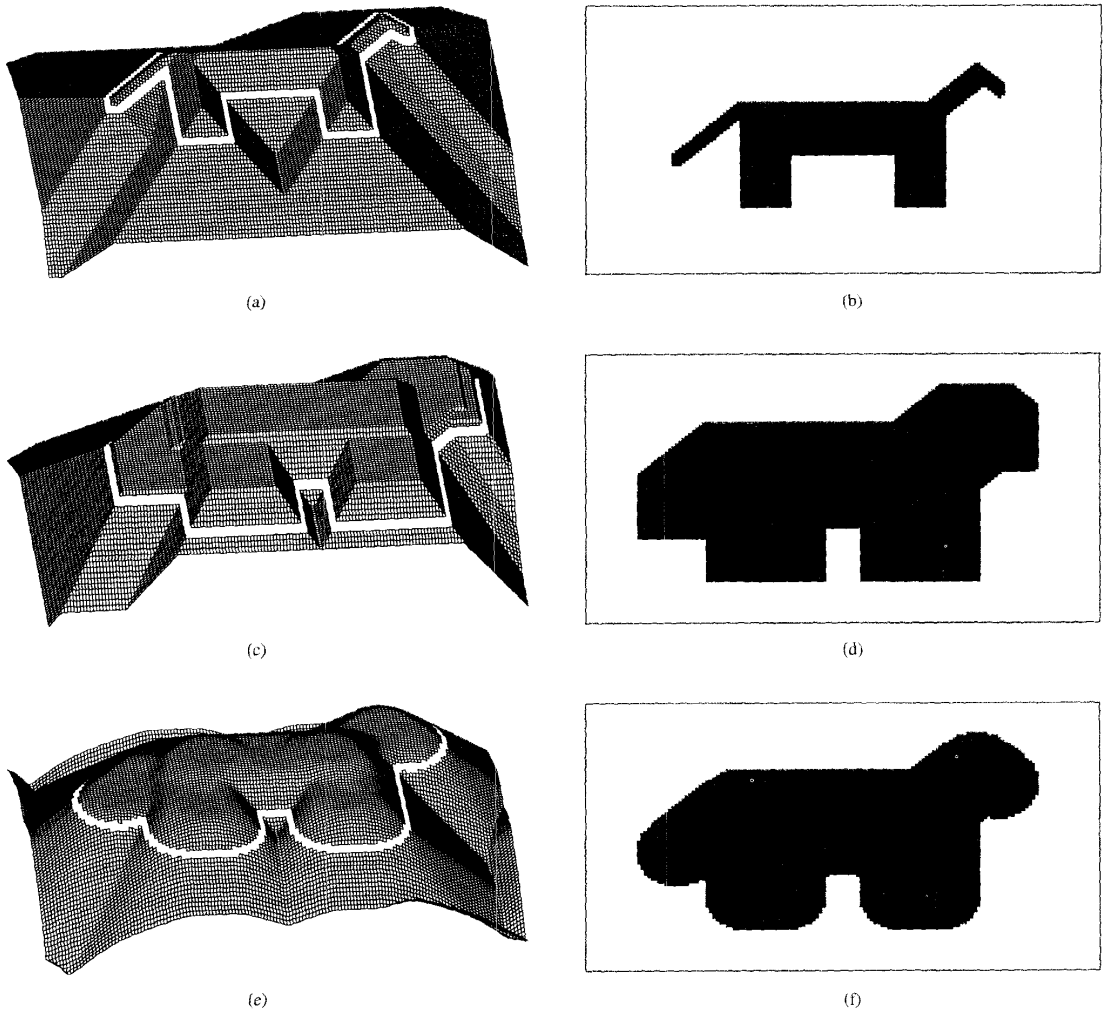


Fig. 11. (a) Original digital gray-level image f and its contour at level=0. (b) Binary image X from thresholding f at level= 0. (c) Flat dilation $f \oplus B$ of f by a discrete disk, i.e., a square of $(2t+1) \times (2t+1)$ pixels with $t = 5$. (d) Binary image $X \oplus B$ from thresholding $f \oplus B$ at level $= 0$. (e) Dilation of f by running the PDE $\partial\delta/\partial t = ||\nabla\delta||_2$ via curve evolution for scales $t \in [0, 5]$ with initial condition $\delta(x, 0) = f(x)$. (f) binary image from thresholding the image in (e) at level $= 0$.

and in the limit $r \downarrow 0$, using Hölder's inequality (19), (55) becomes $\mathcal{M}_{xy}(\delta) = ||\nabla\delta||_q$, which simplifies (54) to

$$\frac{\partial\delta}{\partial t} = ||\nabla\delta||_q \tag{57}$$

with $q = p/(p-1)$. This PDE states that the velocity of the multiscale dilation by a ball B_p along the scale direction t is equal to the q-norm of its spatial gradient over the x, y plane. The multiscale erosions ε satisfy equations identical to the above dilation PDEs except that one must multiply the gradient norms by -1; i.e., $\partial\varepsilon/\partial t = -||\nabla\varepsilon||$. These simple but nonlinear PDEs are satisfied at points where the data are smooth, i.e., the partial derivatives exist. Starting from a continuous f, the multiscale functions δ and ε remain continuous at all x, y, t. However, even if f is differentiable, as the scale t increases the multiscale erosions/dilations can create discontinuities in their derivatives; then these derivatives and the generator PDEs have to be interpreted correctly at such points according to the specific case. This problem is solved if we assume that left and right derivatives exist in each direction and replace the conventional derivatives with morphological derivatives.

In [41] the dilation PDE $\partial\delta/\partial t = ||\nabla\delta||_2$ was numerically solved using robust algorithms from [30] for curve evolution. Thus multiscale dilations by disks can be modeled in a continuous framework, i.e., via PDEs, and then implemented via algorithms that can approximate arbitrarily well the shape of the Euclidean disks on the discrete plane. Fig. 11 shows the results of a simulation to compare the traditional dilation of digital images via discrete max-sum convolution of the image by digital approximations to disks, e.g., squares, versus

a dilation, which is the solution $\delta(x, y, t)$ of the dilation PDE implemented as in [41] using the curve evolution algorithm of [30]. Comparing both the gray-level images and their binary versions (from thresholding at level $= 0$) it is evident that the PDE approach to multiscale dilations can give much better approximations to Euclidean disks and, hence, avoid the abrupt shape discretization inherent in modeling digital multiscale dilations using discrete disks.

The eikonal PDE $||\nabla u(x, y)|| = \eta(x, y)$ relates the magnitude of the eikonal's gradient to the index field. Hence, it is related to the multiscale morphological PDEs where the velocity along the scale direction is equal to the norm of the spatial gradient. Actually, in [41] continuous-scale dilations of binary images have been implemented by embedding the binary image as the zero-level threshold set of a function ϕ and solving the PDE $\partial\phi/\partial t = ||\nabla\phi||$ using the numerical algorithms of [30] that are suitable for PDEs of the eikonal type. Also, in [19] the curve evolution approach to distance transforms is related to solutions of eikonal-based PDEs of the type $u_t = ||\nabla u(x, y)|| - \eta(x, y)$ at steady state ($u_t = 0$).

Finally note that, given that $||\alpha||_q$ is the upper slope transform of the indicator function of the ball B_p (see Table II), it follows that

$$\frac{\partial\delta}{\partial t} = \mathcal{M}_{xy}(\delta) = \mathcal{S}^{\vee}(-I_{B_p})(\nabla\delta) \tag{58}$$

where $\mathcal{S}^{\vee}(f)$ denotes the upper slope transform of a signal f. Namely, the velocity of δ along the scale direction t (and its spatial morphological derivative) is equal to the slope transform of the indicator function of the ball B_p evaluated at a slope equal to the spatial gradient of δ. Details of this slope transform interpretation of morphological derivatives and dilation PDEs can be found in [16]. Also, in [26] the PDEs describing the scale-space evolution of multiscale dilations of umbras (i.e., hypographs and epigraphs) of signals were expressed in terms of Fenchel conjugates (closely related to the slope transforms) using ideas from convex analysis; however, note that set dilation of signal umbras does not always correspond to max-sum convolution $\oplus$ of signals.

VIII. Conclusions

We have developed 2-D max/min-sum difference equations that model the space dynamics of 2-D morphological systems and slope transforms that can analyze these systems in a transform domain. We have also introduced the concepts of impulse response and slope response that are useful in understanding the behavior of distance transforms as infinite-extent min-sum convolutions in the space domain and as bandpass spatial slope filters in the slope domain. Some of these ideas were also used to analyze the eikonal PDE solved via distance transforms and the evolution PDEs modeling multiscale dilations and erosions. We view the analysis of these nonlinear PDEs and the max/min-sum difference equations as a unified area in nonlinear image processing, which we call *differential morphology*.

The eikonal PDE has many potential applications, including solving problems in distance-path finding, segmentation, gridless halftoning, and shape from shading. However, its computation via weighted distance transforms has considerable complexity. Since we have shown that distance transforms are slope filters, it remains to be seen whether a faster distance computation can be found based on slope transforms; the appealing fact here is the correspondence of min-sum signal convolutions in space to transform additions in the slope domain. However, since weighted distance transforms require difference equations with spatially varying coefficients, a study is also needed of the shift-varying slope response of adaptive min-sum difference equations.

The dilation PDEs suggest new ways to view and implement morphological multiscale filtering that avoid the shape discretization effects inherent in all discrete implementations as demonstrated in [41] by implementing continuous-scale morphological dilations of binary images via curve evolution. Given that good approximations to Euclidean distance transforms and, hence, to multiscale dilations of binary images by disks, can be efficiently computed using chamfer distance transforms, a comparison remains to be done to study the relative advantages of the curve evolution versus the chamfer distance approach in terms of goodness of approximation and computational complexity. A preliminary work in [10] indicates that the PDE approach, implemented via curve evolution, can achieve a better approximation to Euclidean distances and multiscale analysis than the chamfer distance transforms, but at a higher computational complexity.

Acknowledgment

The author expresses appreciation to A. Butt for preparing Figs. 8, 10, and 11, and for insightful discussions; and F. Meyer and other anonymous reviewers for many helpful comments that improved this paper.

References

[1] L. Alvarez, F. Guichard, P. L. Lions, and J. M. Morel, "Axiomatization et nouveaux operateurs de la morphologie mathematique," *C. R. Acad. Sci. Paris*, t.315, serie I, pp. 265–268, 1992.

[2] L. Alvarez and J. M. Morel, "Formalization and computational aspects of image analysis," *Acta Numer.*, pp. 1–59, 1994.

[3] R. Bellman and W. Karush, "On the maximum transform," *J. Math. Anal. Appl.*, vol. 6, pp. 67–74, 1963.

[4] S. Beucher and C. Lantuejoul, "Use of watersheds in contour detection," in *Proc. Int. Workshop Image Processing: Real-time Edge & Motion Detect./Estimat.*, Rennes, France, 1979.

[5] H. Blum, "Biological shape and visual science (part I)," *J. Theoret. Biol.*, vol. 38, pp. 205–287, 1973.

[6] G. Borgefors, "Distance transformations in digital images," *Comput. Vision, Graphics, Image Processing*, vol. 34, pp. 344–371, 1986.

[7] M. Born and E. Wolf, *Principles of Optics*, 1987 ed. Oxford, UK: Pergamon, 1959.

[8] R. W. Brockett and P. Maragos, "Evolution equations for continuous-scale morphology," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, San Francisco, CA, Mar. 1992.

[9] ———, "Evolution equations for continuous-scale morphological filtering," *IEEE Trans. Signal Processing*, vol. 42, pp. 3377–3386, Dec. 1994.

[10] M. A. Butt and P. Maragos, "Comparison of multiscale morphology approaches: PDEs implemented via curve evolution versus Chamfer distance transforms," in *Proc. Int. Symp. on Mathematical Morphology and Its Application to Image and Signal Processing*, Atlanta, GA, May 1996.

[11] R. Courant and D. Hilbert, *Methods of Mathematical Physics*. New York: Wiley, 1992.

[12] L. Dorst and R. van den Boomgaard, "Morphological signal processing and the slope transform," *Signal Processing,* vol. 38, pp. 79–98, July 1994.
[13] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1984.
[14] P. K. Ghosh, "An algebra of polygons through the notion of negative shapes," *CVGIP: Image Understanding,* vol. 54, no. 1, pp. 119–144, 1991.
[15] H. J. A. M. Heijmans, *Morphological Image Operators.* Boston, MA: Academic, 1994.
[16] H. J. A. M. Heijmans and P. Maragos, *Lattice Calculus of the Morphological Slope Transform,* CWI Rep. BS-R9531, Dec. 1995.
[17] B. K. P. Horn, *Robot Vision.* Cambridge, MA: MIT Press, 1986.
[18] Special session on "Nonlinear dynamics for image processing," in *Proc. Int. Conf. Image Processing,* Austin, TX, Nov. 1994.
[19] R. Kimmel, N. Kiryati, and A. Bruckstein, "Sub-pixel distance maps and weighted distance transforms," *J. Math. Imag. Vision,* to be published.
[20] J. J. Koenderink, "The structure of images," *Biol. Cybern.,* vol. 50, pp. 363–370, 1984.
[21] B. Lay, "Recursive algorithms in mathematical morphology," *Acta Stereol.,* vol. 6, no. III, pp. 691–696, 1987.
[22] G. Levi and U. Montanari, "A grey-weighted skeleton," *Inform. Contr.,* vol. 17, pp. 62–91, 1970.
[23] P. Maragos, "Morphological systems: slope transforms and max–min difference and differential equations," *Signal Processing,* vol. 38, pp. 57–77, July 1994.
[24] ——, "Slope transforms: Theory and application to nonlinear signal processing," *IEEE Trans. Signal Processing,* vol. 43, pp. 864–877, Apr. 1995.
[25] P. Maragos and R. W. Schafer, "Morphological systems for multidimensional signal processing," *Proc. IEEE,* vol. 78, pp. 690–710, Apr. 1990.
[26] J. Mattioli, "Differential relations of morphological operators," in *Proc. 1st Int. Workshop Math. Morph. Applic. Signal Processing,* Barcelona, Spain, May 1993.
[27] F. Meyer, "Integrals and gradients of images," *Proc. SPIE vol. 1769: Image Algebra Morph. Image Processing III,* pp. 200–211, 1992.
[28] ——, "Topographic distance and watershed lines," *Signal Processing,* vol. 38, pp. 113–125, July 1994.
[29] L. Najman and M. Schmitt, "Watershed of a continuous function," *Signal Processing,* vol. 38, pp. 99–112, July 1994.
[30] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton–Jacobi formulations," *J. Comput. Phys.,* vol. 79, pp. 12–49, 1988.
[31] Y. Pnueli and A. M. Bruckstein, "Gridless halftoning: A reincarnation of the old method," Technion, Israel, CIS Rep. #9323, Oct. 1993.
[32] ——, "Digi$_D$ürer—a digital engraving system," *Visual Comput.,* vol. 10, pp. 277–292, 1994.
[33] F. Preteux, "On a distance function approach for gray-level mathematical morphology," in *Mathematical Morphology in Image Processing,* E. R. Dougherty, Ed. New York: Marcel Dekker, 1993.
[34] R. T. Rockafellar, *Convex Analysis.* Princeton, NJ: Princeton Univ. Press, 1972.
[35] A. Rosenfeld and J. L. Pfaltz, "Sequential operations in digital picture processing," *J. Assoc. Comput. Mach* vol. 13, pp. 471–494, Oct. 1966.
[36] ——, "Distance functions on digital pictures," *Pattern Recog.,* vol. 1, pp. 33–61, 1968.
[37] E. Rouy and A. Tourin, "A viscocity solutions approach to shape from shading," *SIAM J. Numer. Anal.,* vol. 29, no. 3, pp. 867–884, June 1992.
[38] D. Rutovitz, "Data structures for operations on digital images," in *Pictorial Pattern Recog.,* G. C. Cheng *et al.,* Eds. Washington, DC: Thompson, 1968, pp. 105–133.
[39] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics.* New York: Wiley, 1991.
[40] P. Salembier and J. Serra, "Morphological multiscale image segmentation," in *Proc. SPIE vol. 1818: Visual Commun. Image Proc. '92,* 1992, pp. 620–631.
[41] G. Sapiro, R. Kimmel, D. Shaked, B. Kimia, and A. Bruckstein, "Implementing continuous-scale morphology via curve evolution," *Pattern Recogn.,* vol. 26, no. 9, pp. 1363–1372, 1993.
[42] M. Schröder, "The eikonal equation," *Math. Intelligencer,* vol. 1, pp. 36–37, 1983.
[43] J. Serra, *Image Analysis and Mathematical Morphology.* New York: Academic, 1982.
[44] J. Serra, Ed., *Image Analysis and Mathematical Morphology, Vol. 2: Theoretical Advances.* New York: Academic, 1988.
[45] F. Y.-C. Shih and O. R. Mitchell, "Decomposition of gray-scale morphological structuring elements," *Pattern Recog.,* vol. 24, pp. 195–203, 1991.
[46] R. van der Boomgaard, "Mathematical morphology: Extensions toward computer vision," Ph.D. dissertation, Univ. of Amsterdam, 1992.
[47] R. van der Boomgaard and A. Smeulders, "The morphological structure of images: The differential equations of morphological scale-space," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 16, pp. 1101–1113, Nov. 1994.
[48] P. W. Verbeek and B. J. H. Verwer, "Shading from shape, the eikonal equation solved by grey-weighted distance transform," *Pattern Recogn. Lett.,* vol. 11, pp. 618–690, 1990.
[49] B. J. H. Verwer, "Distance transforms: metrics, algorithms, and applications," Ph.D. dissertation, Tech. Univ. of Delft, The Netherlands, 1991.
[50] L. Vincent, "Morphological algorithms," in *Mathematical Morphology in Image Processing,* E. R. Dougherty, Ed. New York: Marcel Dekker, 1993.

Petros Maragos (S'81–M'85–SM'91–F'95), for photograph and biography, please see this issue, p. 807.