

# Gridless Halftoning: A Reincarnation of the Old Method

YACHIN PNUELI AND ALFRED M. BRUCKSTEIN

Computer Science Department, Technion, Israel Institute of Technology, Haifa, Israel

Received October 4, 1993; revised August 31, 1995; accepted September 8, 1995

Despite continuing research and steady progress in the field of digital halftones, there is a feeling that a wide gap in quality still remains between the best available and the best achievable results. A glance at man-made halftones readily confirms this feeling. To bridge this gap, we propose the use of computer-generated, yet *gridless* halftones. This involves solving the halftoning problem on the continuous 2D plane rather than on the usual discrete grid of pixels. In this article we outline this new approach, describe its expected advantages over existing techniques and demonstrate some of them via a prototype system, *DigiDürer*, developed for this purpose. © 1996 Academic Press, Inc.

## 1. HALFTONING—DO WE NEED A NEW APPROACH?

A *halftone* is a binary picture,  $h(x, y)$ , each point being either completely black or completely white, that gives the impression of an image containing shades of gray. *Halftoning* is the process of transforming a gray-level image,  $f(x, y)$ , into a halftone, with the aim that when  $h(x, y)$  is displayed on a bilevel medium, the *human observer* will see in it as good an approximation as possible of the original  $f(x, y)$ .

Both applications and algorithms for digital halftones appeared almost as soon as computers became capable of storing and generating images. Most of us are probably familiar with early attempts at digital halftoning via the use of a line printer (for example see [21] and Fig. 1).

As display devices became more sophisticated and computer power increased, the methods commonly used today were invented. These include *clustered ordered dither* (Fig. 2), which is a computerized version of the analog halftone-screen process used in the print industry, *dispersed ordered dither* (Fig. 3) due to Bayer [5], which is more suitable to CRT-terminal-type devices, and *error diffusion* methods, of which the Floyd Steinberg [11] and Jarvis *et al.* [16] algorithms (Fig. 4), published in 1975 and 1976, respectively, were the first and still are the most commonly used to date.

On the theoretical side, an important step was taken by Allebach in 1981 [3] when he suggested formulating

halftoning as an *optimization process* based on modeling the response of the human visual system. Research since has led to a series of increasingly sophisticated models and algorithms for halftoning. For examples see Ulichney's blue noise [34] simulated annealing techniques [9], space-filling curve techniques [30, 35, 37, 38] halftones via pyramids [23], neural network approaches [1, 4, 18] (Fig. 5) and the recently published unified framework for simulated annealing and neural network methods [43]. See [3] and [34] for a bibliography on the field up to the early 1980s and Chapter 2 of [25] for an overview of work done since.

One may ask how much of an improvement do we get with each of these increasingly sophisticated methods and whether we are approaching the "optimum" we aim for? If a reasonable and accepted model of the human visual system was available, we could have answered this question objectively (say by grading, for a given input image, the outputs generated by the different methods, according to a cost function measuring the difference between the input image and the different outputs as perceived via the model). However, even without an objective measure, one cannot deny that, whereas much progress has been achieved and increasingly complex methods do tend to produce increasingly better-looking outputs (compare Figs. 2–5), there is also a "leveling off" effect in quality—in the sense that the first two or three publications in a specific context produce notable jumps in quality, the improvement in later methods being often marginal.

If this was the whole story, one would be tempted to conclude that we are apparently approaching the optimum. However going back to times before the dawn of the computer age (and even before the invention, for newspaper printing, of the analog halftone screen) one can easily find etchings, engravings, and woodcuts, that are examples of halftones of a quality dramatically superior to those produced by any of our existing computerized methods (Fig. 6).<sup>1</sup> Although there is much that is art in such man-made

<sup>1</sup> Provided they are displayed in the same resolution. Although no woodcut compares in quality of rendition to a halftone at 2500 dpi, existing methods could not compete with woodcuts if limited to the resolution used in them.





FIG. 2. Halftoning with clustered ordered dither.

*provided or some sort of automatic image understanding.*

In the sequel we define gridless halftone methods, and discuss the advantages we expect to gain from the new approach. We also remark on the difficulty of fairly comparing them to the usual grid-based methods. In Section 3 we develop one such method for halftoning with line elements (the first halftone-by-lines method proposed to our knowledge) and describe the system based on it, which we call **DigiDürer**. We also show that with no understanding whatsoever of the image, this method provided good results only on a restricted class of inputs. In Section 4 we show how, given the type of information considered available via existing (so-called low-level) computer vision algorithms, we can modify the basic **DigiDürer** to produce high-quality results on arbitrary images. In Section 5 we show that additional understanding of the scene can be used to generate halftones of even better quality, more

closely resembling man-made engravings. Section 6 shows that our method generalizes to halftoning with (almost) arbitrary shape elements. We conclude in a lighter tone and show how the machinery available in **DigiDürer** can be used to generate halftones that give the impression of graphic art (of course at the expense of being less true, in the strict sense, to the original image).

Most of the material in this paper appears in Chapter 4 of [25] (in Hebrew) and some of the topics of Sections 3, 4, and 6 are discussed, from a different view point, in [24].

## 2. WHAT ARE GRIDLESS HALFTONES?

### 2.1. Defining Physically Realizable Gridless Halftones

It is customary in the literature to view *gray-level images* as the family of functions from  $\mathbb{R}^2$  (the real 2D plane) to the segment  $[0, 1]$ , *discrete images* or *images on the grid* are then viewed as the family of functions from  $\mathbb{Z}^2$  (the



FIG. 3. Halftoning with dispersed ordered dither.

2D integer lattice) to  $[0, 1]$ . In this framework halftones are defined as functions from  $\mathbb{Z}^2$  to the set  $\{0, 1\}$ .

Analogously, we define gridless halftones as the family of functions from  $\mathbb{R}^2$  to the set  $\{0, 1\}$ . However, while this definition removes the restrictions of the grid, it is clearly too liberal, as the following example demonstrates: for a given image  $f(x, y)$  let  $h^*(x, y)$  be a function such that for any (even infinitesimally small) region  $A$  the ratio of black to white points in  $h^*(x, y)$  is equal to  $\int_A f(x, y) dA$ .

It is easy to see that  $h^*(x, y)$  (which can be defined as a fractal) is theoretically the ideal solution to the halftone optimization problem, yet is useless for all practical purpose as no bilevel physical device can ever implement it. To avoid such pitfalls we must impose some limitations on gridless halftones so that they become *physically realizable gridless halftones*.

The least restrictive and most natural limitation is a *covering* condition, stating that there is some finite area

shape  $S$  (or set of shapes  $S_1, \dots, S_n$ ) such that  $h(x, y)$  can be *covered* by  $S$  (or  $S_1, \dots, S_n$ ). Formally, it should be possible to place a finite number of copies of  $S$  in the plane such that for each point  $(x, y)$ :  $h(x, y) = 1$  if and only if  $(x, y)$  belongs to one of these copies.

For specific physical devices one can impose additional constraints as necessary. For example:

- Orientation. Limiting copies of the covering shape to being placed only at specific orientations.
- Disjointness. Limiting copies of the covering shape to being disjoint.
- Discrete localization. Limiting copies of the covering shape to being placed only such that their centers of gravity are located on a specific grid. (Note that we return to the usual grid-based definition only in the specific case when the discretization of the size of the shape equals the discretization of its localization.)



FIG. 4. Halftoning with Jarvis *et al.*, [16] error diffusion.

Any halftone that obeys the covering condition will be considered physically realizable (since it is realizable by at least one physical device—the human hand holding an appropriately shaped tool). From now on by *gridless halftones* we refer only to those that are physically realizable.

## 2.2. Expected Advantages

What are the advantages of gridless halftone methods and what are the reasons to expect them?

1. Improved quality. Both in general and specifically in the range of resolutions where the graphic elements are visible. As already mentioned, only for gridless halftones do we have a “proof” that better quality is achievable. Moreover there is much literature aimed at teaching the would-be artist how to achieve such quality. Although we do not expect this literature to provide us with ready-made

computer algorithms, it is definitely relevant in our search for them.

2. Suitability to a variety of output devices. Current methods are almost exclusively tailored to raster-type devices. Removing the grid restriction allows for methods more directly suited to any type of existing or imaginable device (by selecting appropriate conditions of covering, orientation etc.). For example, a pen plotter requires halftones with covering shapes that are lines of a given width. While it is true that today raster devices dominate the image reproduction market, we believe that part of the reason is precisely the fact that they are the ones best-supported by current halftone methods.

3. Robustness to printing errors. It is well known (see, e.g., [2, 34] and others) that the print medium is especially prone to reproduction errors. Effectively this sets a limit on the resolution, since, as the pattern we wish to display



FIG. 5. Neural network optimized halftoning [18].

becomes finer, the error term becomes more dominating. However, the errors in the reproduction of an ink dot are of two types, in the size of the dot and in its placement. Since the different types of errors stem from different physical limitations, they are not necessarily related and hence impose different resolution limitations. In grid-based methods these limitations must be combined into a single-resolution parameter—the pixel size. Gridless methods on the other hand can easily enforce the two types of limitations separately, via the size of the graphic elements used and the constraints on where the elements can be placed.

4. Utilizing information about the image. It is obvious that an algorithm that understands the contents of an image has a better chance of producing a higher quality halftone. Yet in the literature there is hardly any mention of such algorithms. The reason for this is that in practice it is very difficult to meaningfully integrate information

about the contents of a picture into any of the proposed halftone cost functions. We believe that part of this difficulty stems from the different “language” and symbols people use when thinking of real world scenes and those they use in matrix operations and cost functions.<sup>3</sup> Gridless methods transcend this “language barrier” since their graphic elements and constraints are more familiar to us as human artists and observers and hence we can draw on existing experience outside the scope of digital halftoning. Of course *a posteriori* it is possible that once a “gridless image content-using algorithm” is found, the knowledge gained can be imported to grid-based methods as well.

<sup>3</sup> The only exceptions are some low-level features of vision such as edges which are meaningful in both terminologies. Indeed one occasionally finds halftone methods that “sharpen” or “enhance” edges (see [17, 26] and the remarks on them in [3, 16] and the remarks in [33]).

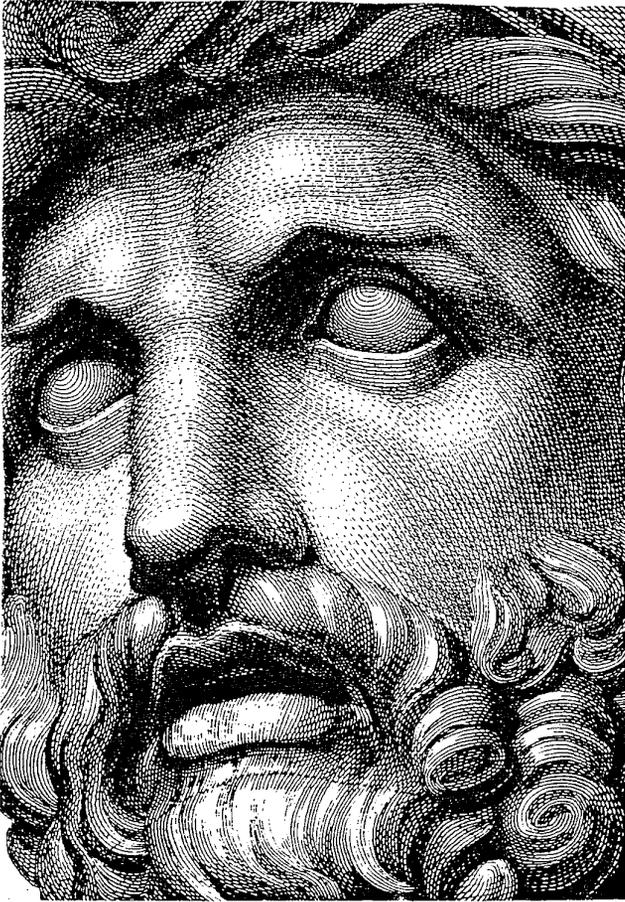


FIG. 6. A man-made halftone artwork [14].

5. Freedom of style. In gridless methods one is not constrained to pixel elements—the variety of graphic elements thus available is limited only by one’s imagination and the capabilities of his output device. This is true even when the underlying device is pixel based, since by labeling different combinations of pixels as *shapes* we can regain this freedom (albeit at the cost of poorer resolution). While it is true that for applications which require strict photo-realism this tradeoff is never advantageous, there are other applications where it can be quite beneficial.

6. User intervention. Advantages 4 and 5, above, give rise to yet another benefit—that of allowing user intervention. While existing methods also allow user intervention, they are typically so constrained that there is very little one can do. With gridless methods, on the other hand, the large room for variations makes interactive user intervention both meaningful and desired. Also, we have the aforementioned advantage that the graphic elements used are familiar to the user.

We note that the lack of a well-defined and agreed-upon model of the human visual system prevents us

from analytically proving the advantages we enumerated above. We therefore resort (as is common in digital halftones) to demonstrating their feasibility via the development of a prototypical gridless halftone method which we named **DigiDürer**, in honor of the great Renaissance etcher and engraver Albrecht Dürer.

### 2.3. Comparing Grid-Based Halftones to Gridless Halftones

Whenever the merits of different halftone techniques are discussed, an important issue is to compare all results at the same *resolution*. When one is limited to a given family of halftones, say grid-based halftones with square pixels, or gridless halftones coverable by circular shapes, defining the resolution is easy (for example, the area of the pixel or the covering circle). However, when one compares methods subject to different sets of constraints, they cannot be effectively calibrated to the same resolution.

There are two ways to somehow make up for this deficiency. The first is to select one of the features of the resolution as dominant and use it as the measure we seek. The most common such feature is the area of the graphic element (used by Ulichney [34] and others to compare halftones on square and hexagonal grids). Although we admit that this measure is always unfair to the more-restrictive method in a comparison, we adopt it in most of the sequel because it is commonly used and easy to calibrate, and also because many times it is indeed the area of the graphic element which imposes the physical limitation on making the resolution any finer. Hence, unless otherwise stated, the areas of the shapes (or the width of the lines) in figures ahead equals the area (or the width) of one pixel from Figs. 2–5.

The other comparison possibility is to use the output device itself as the measure of resolution. This means that given two methods A and B and an input image  $f(x, y)$ , the comparison is made between the best-displayable resolution of A-as-applied-to-  $f(x, y)$  and the best-displayable resolution of B-as-applied-to-  $f(x, y)$ . Since the final purpose of all halftones is the display on a physical device, this type of comparison is in a sense “the fairest of them all.” However it has the big disadvantage that now the calibration of a halftone method to a specific device becomes an integral part of the method, while typically we would like to separate the more general issue of selecting a halftone technique from the more technical issue of properly calibrating it to a specific device. Since this calibration for any halftone method to a specific device requires a technical expertise of the device which we did not have, we did not use this method of comparison.

width = 1 pixel, evolution speed = regular



FIG. 7. Halftoning with the eikonal equation (note shock-waves!).

### 3. DigiDürer HALFTONES VIA CURVE EVOLUTION

#### 3.1. The Eikonal Equation

Our aim is to develop a gridless halftone method. We begin by limiting ourselves to a single family of graphic elements—that of curved lines. There are a number of reasons for this choice:

- In graphic artwork line elements dominate, and since our motivation is partially to mimic such artwork, line elements are indeed most appropriate.
- To date there is no halftone method which uses lines as its basic graphic element, whereas there are devices and contexts where these elements are preferable.
- We found that for our algorithmic purposes, every graphic element can be regarded as either a *line* or a *shape* according to whether it has one or two bounded dimensions. The fact that one of the dimensions of a *line* is

unbounded allows us to generate *line*-halftones by placing the origins of these elements on a planar curve. For *shapes*, on the other hand, we have to specify both coordinates. Thus in some sense *line*-halftoning is a one-dimensional problem whereas *shape*-halftoning is two-dimensional. (More on this in Section 6 when we develop a *shape*-halftoning method).

Next we define the “rules” by which we want our lines to abide. Observing “lines” in man-made halftones we notice that “lines-in-art” are of two kinds, *contour-lines* and *tone-lines*. While from the artistic or graphic point of view contour-lines are very important, from the point of view of halftones they play a minor role and hence we shall ignore them (until Section 7). Tone-lines on the other hand are exactly those we wish to capture. We notice that the behavior of tone-lines in artwork, although globally complex, is locally quite simple: adjacent lines are locally similar to each other

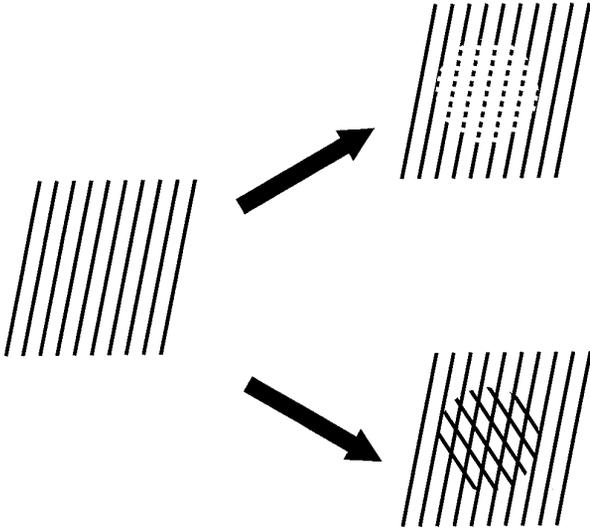


FIG. 8. Tone correction employing further degrees of freedom (line styles and overlays).

and their local density is proportional to the grayness we wish the observer to see. These rules are consistent with the common paradigm of all halftone algorithms:

*Halftone Paradigm.* The *local* density of black elements should be proportional to the *local* grayness of the original image.

We take this paradigm to be the basic driver of our method. One way to mathematically translate this paradigm into a line generation algorithm is via the *eikonal* equation:

$$\|\nabla H(x, y)\| = \sqrt{\frac{\partial H^2}{\partial y} + \frac{\partial H^2}{\partial x}} = f(x, y). \quad (1)$$

It can be easily seen that the equiheight lines of the solution  $H(x, y)$  of this equation obey the halftone paradigm, see the Appendix.

The idea of using this equation to generate lines which make up a picture was originally suggested by Schröder [27]. His motivation was a gestalt-type illusion where locally the picture is meaningless yet globally an image can be seen.

For an image  $f(x, y)$ , a halftone with line elements of width  $w$  is therefore the set of equiheight contours of the “potential field”  $H(x, y)$  solving Eq. (1), sampled at a rate proportional to  $w$ . To completely define the field  $H(x, y)$  one also has to specify, as an initial condition, at least one equiheight contour.

Since we are not interested in the field  $H(x, y)$  as such, but only its equiheight contours, it makes sense to compute them directly via curve evolution methods. We note that the curve evolution approach is a powerful tool for transforming an image into a set of lines and has been applied successfully in a variety of fields such as shape

from shading, curve offsets, image morphology, finding geodesics between two points on a surface, and more (see [6–8, 10, 19, 22, 28, 29, 31] for a partial list of algorithms and applications).

The details of our algorithms already appear in [24] and hence are not repeated here. However, the important points to observe are the following:

- Curve evolution can be computed either in the “Lagrangian” formalism where a line is computed directly from its predecessor or in the “Eulerian” formalism where an implicit function of a higher dimension is used. In this last approach the curves are the zero level-sets of a bivariate function at different moments in time. The advantage here is that the “topology takes care of itself” that is saddle points and extrema of the potential field are automatically dealt with correctly and singularities are always resolved according to the Huygens principle of propagation. Because of this, the Eulerian formalism is currently the preferred one for most curve evolution applications. Its disadvantage is that because the curves do not evolve directly but rather are a by-product of the higher dimensional function it is more difficult to intervene in the evolution process when there are considerations outside the scope of the “strict” rules of the evolution. Because this intervention plays a cardinal part in later stages of **DigiDürer** we implemented the Lagrangian formalism.

- Initial conditions. When viewing lines in artwork one observes that selecting an optimal initial condition is a delicate task requiring much understanding of the contents of the image rendered. Happily, although this observation is true, even arbitrary initial conditions (the only viable option when no information on image content is available) yield reasonable output quality. It is only when one is heavily using content information from the image to improve the halftone generation process that one must also be so careful in selecting the initial condition.

Figure 7 gives a feeling of what can be obtained with a crude implementation of this method. We note that since we only took local effects into consideration, as expected, the output image is of high quality locally yet globally suffers from quite noticeable defects (specifically the discontinuities in the line slopes, or *shocks* ones sees in many regions of the picture).

### 3.2. Improving the Method

It is clear from Fig. 7 that to achieve acceptable quality halftones, we must impose some additional “global” constraints. These can be formulated via the following smoothness paradigm:

*Smoothness Paradigm.* The curves that compose the halftone should be nice and smooth (i.e., everywhere differentiable).



FIG. 9. Halftoning by line-style modulation, a limit case of the image modification method.

3.2.1. *The  $\alpha$ -weighted Diffusion Method.* Can the eikonal equation satisfy this requirement? To answer this question let us rewrite the equation as an evolution equation for  $L(s, t)$ —the parameterized equipotential lines (with  $s$  tangent and  $t$  normal to the curve):

$$\frac{\partial L}{\partial t} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \frac{\partial L / \partial s}{\|\partial L / \partial s\|} f^{-1}(x, y).$$

In this form we see that our equation is hyperbolic (or a *reaction equation*) and hence our smoothness requirement can only be satisfied when either the input image is almost constant or when the output is restricted to a close proximity of the initial condition [32].

It is well known that these restrictions do not apply to equations of the parabolic (or diffusion) form [32], hence

we modify the “driving” equation by adding an  $\alpha$ -weighted “diffusion” term to it:

$$\frac{\partial L}{\partial t} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \frac{\partial L / \partial s}{\|\partial L / \partial s\|} f^{-1}(x, y) + \alpha \frac{\partial^2 L}{\partial s^2}. \quad (2)$$

It can be shown that for large enough values of  $\alpha$  the smoothness requirement will be satisfied for almost any (differentiable) input image.

However, this equation adheres to the halftone paradigm only for values of  $\alpha$  below some threshold (which is proportional to the curvature of  $L$ ). Moreover, above a very low value of  $\alpha$  the diffusion effect clearly manifests itself in a loss of sharpness in the output. By setting  $\alpha$  adaptively over the image (from a maximal value in regions of roughly constant graylevel to a zero value at edges where



FIG. 10. Halftoning with the image modification method, eikonal equation with diffusion and linestyle modulation.

retaining sharpness is most important) we can somewhat control this effect, however, never enough to fully sustain the smoothness paradigm.

What we do achieve by using a *reaction-diffusion* variation of the eikonal equation, is an increase in the distance the solution can stray from the initial condition, or, conversely, an increase in the tolerated rate of change in the input, before losing the smoothness paradigm.

3.3.3. *The Image Modification Method.* Since we still cannot handle general inputs, but only images restricted to some bounded rate of change, this result is obviously unsatisfactory. However, assume we impose this restriction only on the curve evolution part of the algorithm: for any input  $f(x, y)$ , we feed to the curve evolution procedure a modified image  $\tilde{f}(x, y)$  whose rate of change is bounded. Clearly the resulting output would conform to the smoothness paradigm; however, since it will not obey the halftone paradigm (with respect to  $f(x, y)$ ) it

would be unacceptable as a halftone. It could, however, form the basis of one if we have some additional degree of freedom:

Assume that in some region,  $\tilde{f}(x, y)$  is darker than  $f(x, y)$ . The density of lines in the output over this region would hence be too high. If we are allowed to dash some of our lines, we can, by locally controlling the amount of dashedness, restore to the region its correct tone. Conversely if  $f(x, y)$  is lighter than  $f(x, y)$  we could use the residue  $(f(x, y) - \tilde{f}(x, y))$  as a new image whose halftone when overlayed on the original halftone restores the correct tone (Fig. 8).<sup>4</sup> When the width of the lines is at our control, we can utilize it to correct the tone as well.

<sup>4</sup> Technically this is slightly more complicated since one has to ensure that the two sets of lines are of different orientations to avoid interference effects, and one has to correct for overlaps and the response of the human visual system to intersections of lines which is higher than the simple sum of the dark area.



FIG. 11. Halftoning with the image modification method, eikonal equation with diffusion, linestyle modulation and overlay.

Using these *image modification* techniques it is possible to bring any image to conform with the smoothness paradigm and at the same time halftone it correctly. However, these techniques also have their price, which is best demonstrated, by the extreme case illustrated in Fig. 9, where we narrowed the permissible rate of change over  $\tilde{f}(x, y)$  to zero. As expected, the output of the curve evolution procedure is a sequence of evenly spaced parallel lines. The halftone effect is thus generated entirely by properly dashing the lines or increasing their width. While the result is definitely a halftone (which might even be judged to be of good quality), it can clearly be seen that a grid has sneaked in. In fact it can be shown that this limiting case is analogous to a variation of clustered ordered dither known as a linear screen.

When the rate of change on an image is restricted to a range wider than a point, we no longer get the “strict” grid as above, however, in a sense we still impose some

form of a grid on the output. From our point of view this is undesirable for two reasons—the first, which is more “ideological,” is that we set out to find *gridless* methods, the second is that for raster devices this “grid” can never be in phase with the grid of the raster and hence the output is much more susceptible to Moiré patterns. Figures 10 and 11 show that results can be obtained via the  $\alpha$ -weighted diffusion technique together with image modification techniques. Figure 10 uses linestyle modulation, while Fig. 11 uses the overlay method.

#### 4. UTILIZING LOW-LEVEL VISION INFORMATION

If we refuse to accept any imposed grid which is noticeable, the *image modification method* like the  $\alpha$ -weighted diffusion becomes only of limited power. In this case we find that it is impossible for many images to abide by the smoothness paradigm without losing some other property

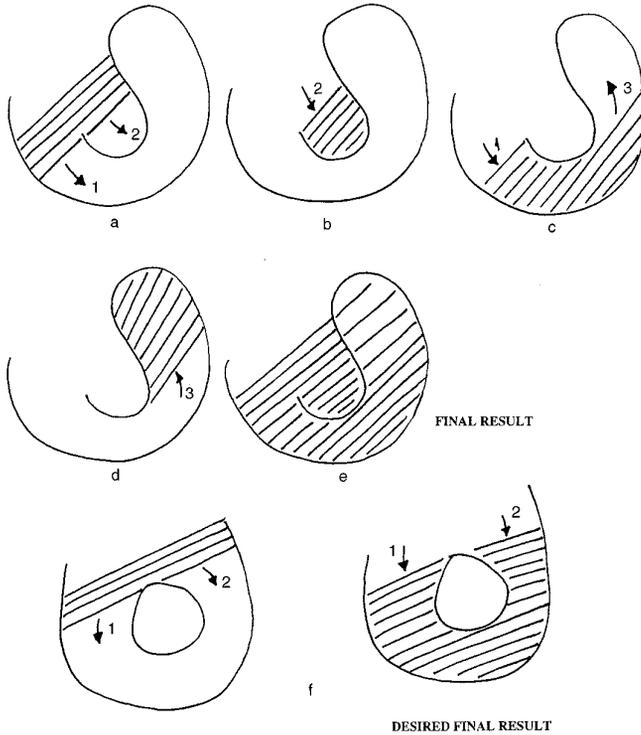


FIG. 12. The “cutting at edges” algorithm solving problems of curve propagation around edges and obstacles.

which we deem highly desirable. Happily, even for high-quality outputs it is sufficient to use a weaker version of this paradigm: notice that edges are discontinuities in the image, and hence the halftone can also be discontinuous across them. We get the following:

*Modified Smoothness Paradigm.* Curves which compose the halftone should be smooth (everywhere differentiable) except across edges where they do not even have to be continuous.

To see why this modified paradigm is so much easier to satisfy, we return to Fig. 7, where practically all noticeable shocks can be traced back to discontinuities in the original image. Analytically we can show that even locally the original smoothness paradigm could never be satisfied over a discontinuity in the input without sacrificing the halftone paradigm.<sup>5</sup> The modified paradigm, on the other hand, can always be enforced locally. Globally we still might have a problem with shocks which evolve slowly, however against these the above  $\alpha$ -weighted diffusion and image modification techniques are quite effective.

It remains to show how to evolve curves under this

<sup>5</sup> Either the output is true to the original and then not smooth or it is smooth and then not true to the original.

weaker paradigm. The first issue is detecting and locating the edges. Since this is a heavily researched area, we shall not dwell on it and just remark that robust and accurate algorithms for edge detection exist. This is especially so for our case as we do not need to locate all edges in the image, but only the most dominant and clear-cut edges (the more diffused an edge is the slower the rate of evolution of the shock it generates).

Given the edges, we are still faced with the issue of how to evolve the curves. Naively the way to do this is via the *cutting at edges algorithm* demonstrated in Fig. 12. The idea is to evolve the curves as before with the following additional rules.

- When a curve crosses an edge, it is separated into two segments, one on each side of the edge. Each segment is then evolved separately (Figs. 12a–12c).
- If a curve can be extended beyond its end-point in the tangent direction without crossing an edge, this is done. The evolution then continues in two directions, in the “forward” direction with the extended curve and in the “backward” direction with the extension only (Figs. 12d–12e).

For most practical cases this algorithm works properly,

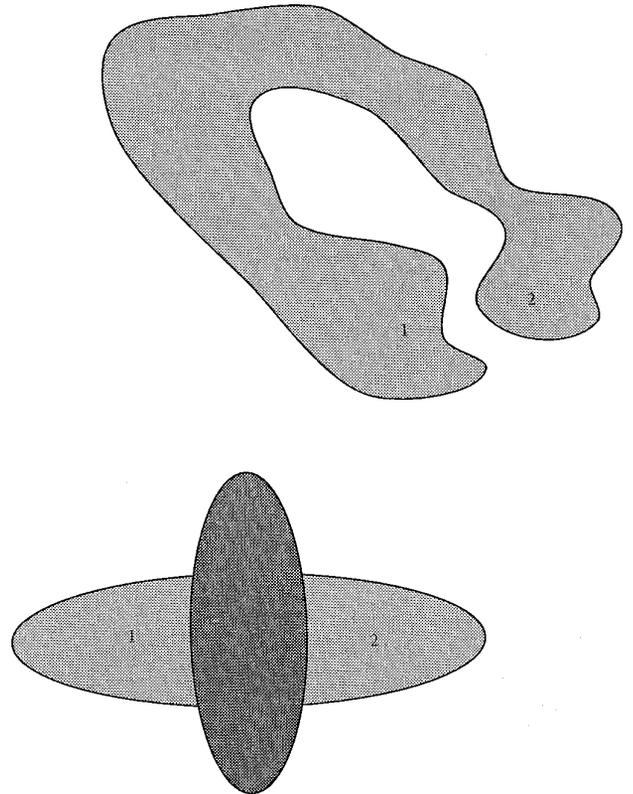


FIG. 13. Separate areas must sometimes have related halftone patterns.



FIG. 14. *DigiDürer* output utilizing edge/segmentation information.

however, there is one notable exception in surfaces with holes (Fig. 12f). When the curve meets the hole it becomes separated into two segments which evolve independently on the two sides of the hole. This leads to trouble since, once the evolution process passes beyond the hole, the two segments should merge back into one, which is unlikely to happen once they have been separated. A solution to this problem is the *segmentation and interpolation algorithm* which operates as follows:

- For each hole, the original gray-level values are ignored and new values are interpolated from the values of the area surrounding the hole.
- The resulting image (which is devoid of holes) is halftoned by the usual method.
- This halftone is output—but only for the area which was originally outside the holes.

- Each hole region is then halftoned separately using its original gray-level values.

We note that in some cases we would like to use this method even when the area we halftone does not “strictly” have holes, but nonetheless we want the halftoning pattern in different regions to be “related,” though separated by edges (Fig. 13). Typically this can happen when we have at our disposal good image segmentation information.

Examples of halftone results via the combination of the *segmentation and interpolation algorithm* and the *cutting at edges algorithm* can be seen in Figs. 14–16.

## 5. UTILIZING HIGHER LEVEL INFORMATION

In the process of developing an algorithm that satisfies both the halftoning paradigm and the modified smoothness

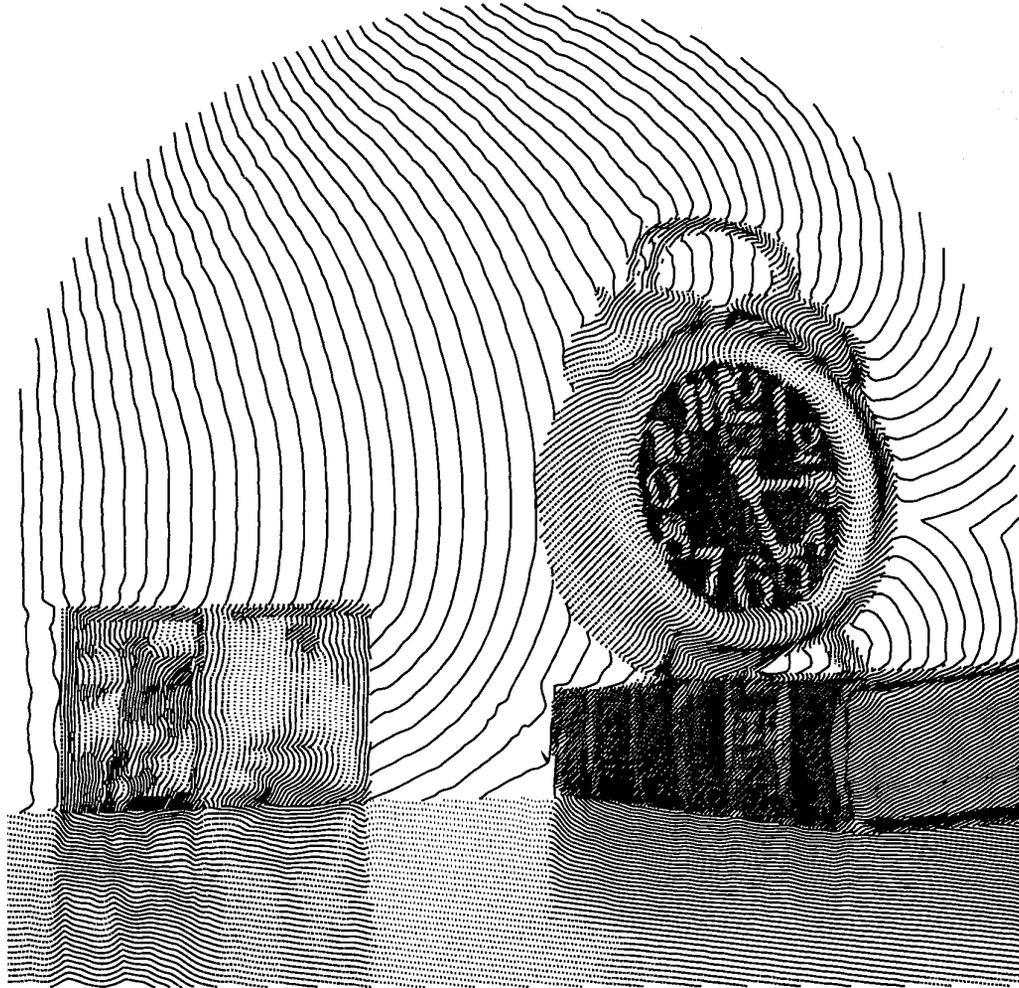


FIG. 15. *DigiDürer* output utilizing edge/segmentation information.

paradigm we have shown some ways to improve the quality of the halftone by the use of information about edges in the image and about its segmentation into *objects*.<sup>6</sup> In this section we show two examples of how the framework we have set up allows us to meaningfully integrate additional information into the halftone process (provided such information is available).

### 5.1. *Dealing with Glares and Shadows*

The first such example is glares and shadows. For our purposes, areas of glare or shadow are parts of an image that appear brighter or darker than adjacent parts belonging to the same semantic (real world) object.

Taking man-made halftones as our yard stick, areas that

belong to the same semantic object should display the same patterns of halftone elements. The problem with glares or shadows is that if we were to halftone them blindly they would display patterns differing considerably from other areas in the same object (as the perceived gray level is different). Applying the algorithms of Section 4 does not help, since, while avoiding shocks that might result from the high rate of change in the gray levels, the discontinuity between the glare or shadowed region and its surrounding is stressed. However if the fact that some region suffers from glare or shadow is known beforehand (either via detection algorithms, human input or from prior ray tracing information) we can use the *image modification* techniques to generate a correct halftone: first we “remove” the shade or glare via some interpolation method, then halftone the modified image and then we correct the tone with any of the tone-correcting techniques outlined in Section 3.2.2. An example can be seen in Fig. 17.

<sup>6</sup> Objects either in the syntactic sense, that is, connected regions of similar gray levels, or in the semantic sense, that is, regions which correspond to the same real-world object.



FIG. 16. **DigiDürer** output utilizing edge/segmentation information.

### 5.2. Enhancing the 3D Structure

A second example is that of enhancing the three-dimensional structure of elements in an image. Again from observing man-made halftones we learn that lines or other graphic elements which compose a halftone typically follow either equiheight or steepest-descent contours of the rendered 3D object.

In fact these directions are so dominant in artwork that one may wonder how do we get acceptable quality at all without taking the 3D structure into consideration. The answer to this question is that, on the one hand, the dominance of a specific solution does not imply its uniqueness and, on the other hand, we have some “luck” in the fact that to a large extent we always (even unknowingly) consider the 3D structure. This surprising fact follows from the observation that for the purpose of *shape from shading*, apart from shadows or glares, a Lambertian model of light reflectance is often appropriate (see [13]). Under this

model the “shape from shading” equation for  $H$ , the true height function of the scene, becomes

$$\sqrt{\frac{\partial H^2}{\partial x} + \frac{\partial H^2}{\partial y}} = T(f(x, y)),$$

where  $T(f)$  is a monotone of  $f$  over  $[0, 1]$ , which is very close both in form and behavior to the eikonal equation governing our curves (see [8]).

When we are given the three-dimensional structure of the scene we halftone, we can, however, do much better: instead of evolving our lines according to the grayness of the image, we can evolve them according to the true 3D structure of the scene—thus enhancing the perceived 3D structure and adhering more closely to the artist’s way of rendering. As above, to ensure a correct rendition of the tone, we may modify the width and dashedness of our lines or apply more than one ensemble of curves (See [42] for

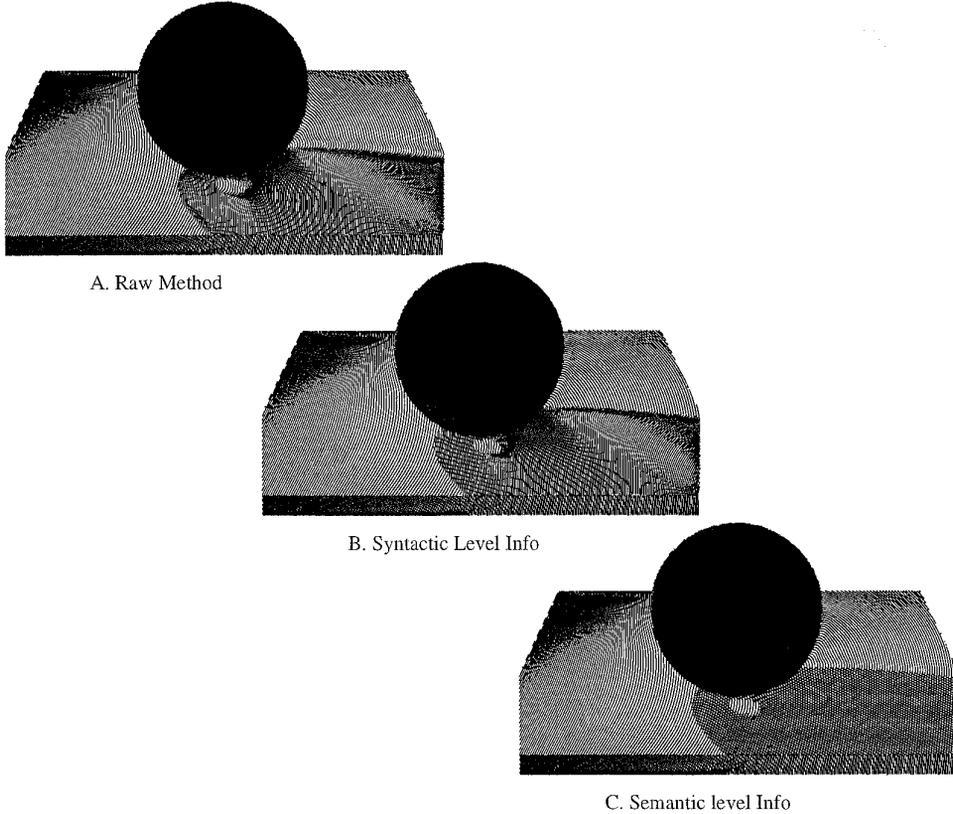


FIG. 17. Dealing with glares and shadows.

an implementation of such ideas that appeared while this paper was under review.) It should be stressed that here too we take advantage of the resemblance in behavior between the eikonal equation and the Lambertian shape from shading equation, through the fact that the corrections made are small enough for the image modification techniques to work.

## 6. GENERALIZING FROM LINES TO AREA SHAPES

We now discuss gridless halftones with graphic elements which are finite area shapes as opposed to lines. Abstractly, for some shape  $A$  and a given image  $f(x, y)$ , the problem is to find a finite set of real valued locations on the plane such that placing scaled copies of  $A$  at those locations generates a binary function  $h(x, y)$  that obeys the halftone condition and minimizes some cost function, based on some model of the human visual system. Stated thus, the problem is two-dimensional. It turns out that we can separate it into two one-dimensional problems.

Let  $H(x, y)$  be a potential field which satisfies the following equation:

$$\|\nabla H(x, y)\| = \sqrt{\frac{\partial H^2}{\partial y} + \frac{\partial H^2}{\partial x}} = \sqrt{f(x, y)}. \quad (3)$$

Let  $L(t)$  be a set of equipotential lines of  $H$  such that the potential difference between  $L(t_i)$  and  $L(t_i + 1)$  is  $\Delta s$ . Now consider a division of each line  $L(t_i) \in L(t)$  into segments such that

$$\int_a^b \sqrt{f(x, y)} ds = \Delta t, \quad (4)$$

where  $a$  and  $b$  are the end points of a segment and the integration is along the direction tangent to  $L(t_i)$ .

It is easy to see that if we select  $\Delta s$  and  $\Delta t$  such that  $\Delta s \times \Delta t = |A|$  (the area of  $A$ ) then by placing one copy of  $A$  in the middle of each segment, we get a halftone (provided the overlapping between adjacent copies is negligible).

Thus we can solve the two-dimensional shape-placing problem by applying the same one-dimensional line-placing algorithm twice: first we find equiheight contours via Eq. (3) (this is exactly our line-placing algorithm but with a modified image), and then we find the location of the shapes along each line via Eq. (4) (this is an integrate to threshold algorithm, which is no more than a one-dimensional version of our line-placing algorithm, with the direction of integration—over the image—predetermined to be the direction tangent to an equiheight contour). For dot



FIG. 18. Gridless halftoning with dot elements.

elements it makes sense to use  $\Delta t = \Delta s$  and we get results as in Fig. 18.

Notice that the only parameter of the covering shape  $A$  we use is its area. This is not much of a surprise since the halftoning paradigm considered deals only with this parameter. More surprising is the fact that indeed the exact shape of the area does not matter as long as it is round enough.<sup>7</sup> Figures 19–22 demonstrate this effect and Fig. 23 shows that it breaks down when the shape is highly anisotropic.

One may argue that for noncircular shapes we should use different values for  $\Delta s$  and  $\Delta t$  and possibly utilize the orientation freedom which we have. For example, for rectangles with an aspect ratio of 1:3,  $\Delta s = 1/2 \Delta t$  should be used and the shapes should be placed on the plane so that

the elongated side of the rectangles is in the direction normal to the lines. However, initial tests suggest that the improvement is not dramatic and that it is difficult to generate reasonable halftones with highly noncircular shapes.

An interesting side effect of the algorithm's independence of the elements' shape is that we can select for every location an arbitrary shape as long as it is round enough and has the right area. Figure 24 demonstrates this effect. It was generated by first applying the gridless shape halftone algorithm to generate the placement of the shapes and then at each location slightly different shapes were chosen according to the strength and direction of the local gradient in  $f(x, y)$ .

A recently proposed method of halftoning via “adaptive screening” is based on the same idea of using image context to control the shape of dots. See [41], a paper that appeared while this work was under review.

<sup>7</sup> Formally, the ratio between  $|A|$  and the area of the bounding circle of  $A$  is not too small.



FIG. 19. Gridless halftoning using various local shapes, round dots (low resolution).

## 7. ART-PLAY: MODIFYING THE DISTANCE FUNCTION

There was a time when critics believed that the measure of quality for drawings and paintings should be how much they resemble “real objects.” A more modern approach argues that a drawing or painting should display objects the way artists see them or even the way they feel about them. In this section we take a “digital” leap in this direction and describe variations of our method generating outputs that are halftones only inasmuch as they are black and white images anchored in some gray-level image but do not obey the halftone paradigm that guided us so far.

As in previous sections our motivation comes from observing the labors of the artist. These days, when photographs are common, it is not unusual to see “artistic” graphic reproductions of a photograph. If the medium used is bilevel (say ink on paper) such a reproduction is a halftone and if the artist was competent most people would judge it to be of high quality. Yet comparing such reproductions to what usually are regarded as halftones (say reproductions of photographs which appear in print) we readily notice vast differences (see for example the way portraits appear in *The Wall Street Journal*). The most striking is the fact that the graphic reproductions often contain mostly contour lines (that,



FIG. 20. Gridless halftoning using various local shapes, squares (low resolution).

strictly speaking, do not exist in the original at all), tone rendition not being even attempted except perhaps in the darkest regions of the image.

Our aim in this section is to point out possible ways of designing algorithms for this purpose. It should be mentioned that this idea is not entirely new, since already in 1976, Gard [12] made an initial step in this direction and suggested a halftone method based an edge detection procedure followed by a threshold operation to render the darker regions of the output in black. The method we wish to use is as follows:

Regard our gridless halftone method of previous sections as a collection of algorithms enveloping each other, each

“layer” doing some preprocessing and postprocessing services on the inputs and outputs of the inner layers. At the heart of this “onion” lies the basic curve evolution algorithm and it is there that the output is made to “resemble” the input. Our idea is to keep our method intact yet replace its heart with some other curve evolution rule more appropriate to the task at hand.

One such rule could be the following *3D-equidistance* rule: Interpret the image as a 3D surface (the gray level being the height, or  $z$  coordinate), and let the curve evolve by moving at a constant speed on this 3D surface in the direction of its normal. What we display is the projection of the trace of the curve in the  $(x, y)$ -plane



FIG. 21. Gridless halftoning using various local shapes, crosses (low resolution).

at constant time intervals. The result can be seen in Fig. 25.

The motivation for this rule comes from the CAD/CAM field, where, as suggested to us by G. Elber, applying this rule on a true 3D surface yields curves that can be used as tool paths for generating the surface with a milling machine. It is also the rule of evolution used in [19] to find shortest distances between points on the 3D surface. When applied to images, it has two features most desirable from our point of view of halftoning with an eye toward graphic artwork: the first is a strong bias towards edges, since at edges the surface is steep and more “time” is needed to

traverse the same distance in the  $(x, y)$  plane. The second is the seemingly random noise on the rest of the image due to the typically very noisy least significant bit of the gray level that dominates the local behavior of the curves.

We shall not elaborate on the usefulness of certain types of noise in many halftone applications (see Ulichney’s blue noise concept in [34] among others). Rather we wish to concentrate on the bias towards edges. Off-hand it might seem that to produce the more “graphic” appearance we aim for, all we have to do is add the edges (from some edge detection algorithm) to our existing *DigiDürer* halftone outputs. However initial tests suggest that this does not



FIG. 22. Gridless halftoning using various local shapes, nonsymmetric round shapes (low resolution).

readily work as the edges added “on top” of the halftone always appeared artificial and disjoint from the halftone patterns.

Now that we have on the one hand a halftone algorithm and on the other a curve evolution rule biased towards edges it seems promising to search for some new rules that combine them. Although this search is still far from over we already have initial results, as can be seen in Fig. 26. Here we used the 3d equidistance rule with the modification that the gray level also determines the difficulty of the terrain—so that the darker the area the slower the rate of advancement. Although this result is definitely not

yet a graphic artwork we feel that a step in this direction has been made.

## 8. CONCLUDING REMARKS

We have presented a new framework for (automatically) generating gridless halftones of gray-level images. The basic idea was trying to emulate the techniques used by graphic artists who manage to produce amazingly good halftone art by various engraving methods. While we are still far from achieving artistic outputs, our digital halftoning/engraving system called **DigiDürer** does pro-



FIG. 23. Gridless halftoning using various local shapes, elongated bars (low resolution).

duce high-quality gridless halftones based on a curve evolution solution to a suitably postulated eikonal equation incorporating additional information often available about the scene rendered. The **DigiDürer** system we created and tested can already be used to produce very good “engraving” style halftones for a variety of printing needs. We hope this work will open new avenues of research into the “no man’s land” still separating computer graphics algorithms and artistic pursuits.

#### APPENDIX: HALFTONING VIA THE EIKONAL EQUATION

Suppose we want to determine a bivariate function  $H(x, y)$ , so that the local density of its equal height contours

(level sets) corresponding to equally spaced levels, defined by  $\{H(x, y) = nh \mid n \in \mathbf{Z}\}$ , will be proportional to a given image  $f(x, y) > 0$ . Assuming that in the neighborhood of a point  $(x_0, y_0)$  the bivariate function can be approximated by a planar patch, we can write

$$H(x, y) = H(x_0, y_0) + p(x_0, y_0)(x - x_0) + q(x_0, y_0)(y - y_0),$$

where  $p$  and  $q$  are the partial derivatives of  $H(x, y)$ , w.r.t. the  $x$  and  $y$  coordinates. In the neighborhood of  $(x_0, y_0)$ , we have that the level sets corresponding to  $nh$  are equally spaced parallel lines, separated by a distance  $d$ , which is given by



FIG. 24. Gridless adaptive halftoning using different shapes at different locations.

$$d[p^2(x_0, y_0) + q^2(x_0, y_0)]^{1/2} = h.$$

The “local density” of the parallel lines is proportional to the inverse of  $d$ , hence we need to have

$$[p^2(x_0, y_0) + q^2(x_0, y_0)]^{1/2} = \frac{h}{d} = f(x_0, y_0).$$

This shows that, indeed, a solution of the partial differential equation

$$\left[ \frac{\partial H(x, y)}{\partial x} + \frac{\partial H(x, y)}{\partial y} \right]^{1/2} = f(x, y)$$

will yield a bivariate function  $H(x, y)$  obeying the desired “halftone paradigm.”

Suppose now that we want to solve the above partial

differential equation for  $H(x, y)$  and we are given as a starting point an equal height contour of  $H(x, y)$ . Following a method first described in [8], in the shape from shading context, we shall parametrize the level sets of  $H(x, y)$  by an arbitrary curve parametrization and the height of the equal-height contours as

$$L(s, nh) = \{[x(s, nh), y(s, nh)] \mid H(x(s, nh), y(s, nh)) = nh\}$$

and devise an algorithm to determine  $L(s, (n + 1)h)$  from  $L(s, nh)$  assuming a small  $h$ . We then note that, along any equal height contour, the curve tangent must be orthogonal to the local gradient of  $H$ . Hence, to climb from  $L(s, nh)$  to  $L(s, (n + 1)h)$ , one has to move in the  $(x, y)$ -plane a distance of  $d = h/(p^2 + q^2)^{1/2} = hf^{-1}(x, y)$  in a direction perpendicular to the curve tangent. We therefore have



FIG. 25. (Art-Play-1) Image generated using a 3D equidistance curve evolution rule.

$$L(s, (n+1)h) = L(s, nh) + hf^{-1}(x, y) \left| \frac{\partial L(s, nh)}{\partial s} \right|^{-1} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \frac{\partial L(s, nh)}{\partial s},$$

leading, in the  $h \rightarrow 0$  limit to the initial value partial differential equation:

$$\frac{\partial L(s, t)}{\partial t} = f^{-1}(x, y) \left| \frac{\partial L(s, nh)}{\partial s} \right|^{-1} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \frac{\partial L(s, nh)}{\partial s}.$$

The algorithms in this paper are direct implementations of this “level set” approach to find solutions of eikonal equations.

*Note added in proof.* Image halftoning continues to present many challenges, both in research and in implementation. Recently, the idea that image content could and should be exploited to generate better halftones

simultaneously appeared in several papers. See, e.g., [24, 39–41]. In the summers of 1990 and 1991, D. Mitchell and A. M. Bruckstein ran a series of experiments with “directional error diffusion” halftoning algorithms, in an attempt to produce engraving-like effects. The surprising result of these experiments was that the classical, nonadaptive, error diffusion methods usually yielded better halftones, hence this direction of work was not pursued further. The present revival of activity in this direction seems to indicate that good effects can in fact be obtained by a combination of directional diffusion and texture analysis, see e.g., [39, 40]. The work presented in this paper took a radically different approach to halftoning from the work mentioned above, and some may even argue that it is more a direction of getting special effects than faithfully reproducing gray-level images.

#### ACKNOWLEDGMENTS

The authors thank Ron Kimmel for discussions on many aspects of curve evolution, among them the Eulerian formalism of Setian and Osher. We also thank G. Elber who suggested (with the CAD/CAM application in mind) the 3D equidistance evolution rule.

Figure 6 is reproduced by permission of MIT Press. The gray-level images used as inputs to our algorithm(s) were taken from [36].



FIG. 26. (Art-Play-2) Image generated using a variable terrain 3D equidistance curve evolution rule.

## REFERENCES

1. D. Anastassiou and S. Kollias, Progressive halftoning of images, *Electronics Lett.* **24**(8), 1988, 489–490.
2. S. H. Algie, Resolution and tonal continuity in bilevel printed picture quality, *Comput. Vision Graphics Image Process.* **24**, 1983, 329–346.
3. J. P. Allebach, Visual model-based algorithms for halftoning images, *SPIE* **310**, 1981, 151–158.
4. D. Anastassiou, Neural net-based digital halftoning of images, *Proc. IEEE ISCAS*, 1988, 507–510.
5. B. E. Bayer, An optimum method for two level rendition of continuous-tone pictures, *Proc. IEEE Int. Conf. Communication*, pp. (26-11)–(26-15), 1973.
6. H. Blum, Biological shape and visual science, *J. Theor. Biol.* **38**, 1973, 205–287.
7. G. Borgefors, Distance transformations in digital images, *Comput. Vision Graphics Image Process.* **34**, 1980, 344–371.
8. A. M. Bruckstein, On shape from shading, *Comput. Vision Graphics Image Process.* **44**, 1988, 139–154.
9. P. Carnevali, L. Coletti, and S. Patarnello, Image processing by simulated annealing” *IBM J. Res. Dev.* **29**(6), 1985, 569–579.
10. P. E. Danielsson, Euclidean distance mapping, *Comput. Vision Graphics Image Process.* **14**, 1980, 227–248.
11. R. W. Floyd and L. Steinberg, An adaptive algorithm for spatial grey scale, *Proc. SID* **17**(2), 1976, 75–77.
12. R. L. Gard, Digital picture processing techniques for the publishing industry, *Comput. Vision Graphics Image Process.* **5**, 1976, 151–171.
13. B. K. P. Horn, *Robot Vision*, MIT Press, Cambridge, MA, 1987.
14. W. M. Ivins Jr., *Prints and Visual Communication*, MIT Press, Cambridge, MA, 1985.
15. W. M. Ivins Jr., *How Prints Look: An Illustrated Guide*, Murray, London, 1988.
16. J. F. Jarvis, C. N. Judice, and W. H. Ninke, A survey of techniques for the display of continuous-tone pictures on bi-level displays, *Comput. Graphics Image Process.* **5**, 1976, 13–40.
17. J. F. Jarvis and C. S. Roberts, A new technique for displaying continuous-tone images on a bilevel display, *IEEE Trans. Commun.* **COM-24**, 1976, 891–898.

18. S. Kollias and D. Anastassiou, A unified neural network approach to digital image halftoning, *IEEE Trans. Signal Process.* **39**(4), 1991.
19. R. Kimmel, A. Amir, and A. M. Bruckstein, Finding shortest paths on graph surfaces, Technion-Israel Institute of Technology, CIS Report 9301, 1993.
20. R. Kimmel and A. M. Bruckstein, Shape offsets via level-sets, *CAD* **25**(5), 1993, 154–162.
21. MacLeod, Pictorial output with a line printer, *IEEE Trans. Comput.* **19**, 1970, 160–162.
22. S. J. Osher and J. A. Sethian, Fronts propagating with curvature dependent speed: Algorithms based on Hamilton–Jacobi formulations, *J. Comp. Phys.* **79**, 1988, 12–49.
23. E. Peli, Multiresolution, error-convergence halftone algorithm, *J. Opt. Soc. Am.* **8**(4), April 1991, 625–636.
24. Y. B. Pnueli and A. M. Bruckstein, **DigiDürer**—A digital engraving system, accepted for publication in *Visual Comput.*; also Technion-Israeli Institute of Technology, CIS Report 9210, June 1992.
25. Y. B. Pnueli, Gridless halftones—Beyond grid-based algorithms, Ph.D. thesis (in Hebrew), Technion-Israeli Institute of Technology, Haifa, Israel, June 1993.
26. P. G. Roetling, Halftone method for edge enhancement and Moire suppression, *J. Opt. Soc. Am.* **66**, 1976, 985–989.
27. M. Schroeder, The eikonal equation, *Math. Intelligencer* **5**(1), 1983, 36–37.
28. J. A. Sethian, Curvature and the evolution of fronts, *Comm. Math. Phys.* **101**, 1985, 487–499.
29. J. A. Sethian, A review of recent numerical algorithms for hypersurfaces moving with curvature dependent speed, *J. Diff. Geom.* (33), 1989, 131–161; *Comm. Math. Phys.* **101**, 1985, 487–499.
30. W. Skarbek, Generalized Hilbert Scan in image printing, in *Theoretical Foundations of Computer Vision* (R. Klette and W. G. Kropatsch, Eds.), Mathematical Research Vol. 69, pp. 47–58, Akademie Verlag, Berlin, 1992.
31. G. Sapiro, R. Kimmel, D. Shaked, B. B. Kimia, and A. M. Bruckstein, Implementing continuous-scale morphology via curve evolution, *Pattern Recognit.* **26**(9), 1993.
32. J. Smoller, *Shock Waves and Reaction-Diffusion Equations*, Springer-Verlag, Berlin/New York, 1983.
33. P. Stucki, MECCA—A multiple error correcting computation algorithm for bilevel image hardcopy reproduction, Research Report RZ1060, IBM Research Lab. Zurich, Switzerland, 1981.
34. R. A. Ulichney, *Digital Halftoning*, MIT Press, Cambridge, MA, 1987.
35. L. Velho and J. Gomes, Digital halftoning with space filling curves, *Proc. SIGGRAPH Comput. Graphics* **25**(4), 1991, 81–90.
36. A. Weber, Image data base, USCIP Report 1070, Image Proc. Inst. Univ. of Southern California, 1983.
37. G. Wyvill and C. McNaughton, Three plus five makes eight: A simplified approach to halftoning, in *Scientific Visualization of Physical Phenomena (Proc. CGI'91)*, pp. 379–392, Springer-Verlag, Berlin/New York, 1991.
38. I. H. Witten and R. M. Neal, Using peano curves for bilevel display of continuous tone images, *IEEE Comput. Graphics Applic.*, May 1982, 47–52.
39. K. R. Sloan and Y. Wang, Exploiting texture effects in dithering, in *SPIE Proceedings, Human Vision, Visual Processing and Digital Display III*, Vol. 1666, 1992.
40. K. R. Sloan and D. Campbell, Texture as information, in *SPIE Proceedings, Human Vision, Visual Processing and Digital Display IV*, Vol. 1913, 1993.
41. M. Hammerin and B. Kruse, Adaptive screening, in *IST Symposium on Prepress, Proofing and Printing*, Nov. 1993.
42. W. Leister, Computer generated copper plates, *Comput. Graphics Forum* **13**(1), 1994, 69–77.
43. R. Geist, R. Reynolds, and D. Suggs, A Markovian framework for digital halftoning, *ACM Trans. Graphics* **12**(2), 1993, 136–159.