COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING 42, 192–205 (1988)

Parsing Scale-Space and Spatial Stability Analysis

WALTER F. BISCHOF* AND TERRY CAELLI†

Department of Psychology and Alberta Centre for Machine Intelligence and Robotics, The University of Alberta, Edmonton, AB, T6G 2E9 Canada

Received December 5, 1986; revised October 16, 1987

The scale-space $S(\mathbf{x}, \sigma)$ of a signal $I(\mathbf{x})$ is defined as the space of the zero-crossings from $\{\nabla^2 G(\sigma) * I(\mathbf{x})\}$, where G is a Gaussian filter. We present a new method for parsing scale-space, spatial stability analysis, that allows the localization of region boundaries from scale space. Spatial stability analysis is based on the observation that zero-crossings of region boundaries remain spatially stable over changes in filter scale. It is shown that spatial stability analysis leads to an edge detection scheme with good noise resilience characteristics and that it can lead to improvements in "shape from texture" methods. © 1988 Academic Press, Inc.

1. INTRODUCTION

One goal of a visual system is to construct a representation of the visual input in terms of objects and their three-dimensional structure, i.e., their shape, spatial relationships, and surface properties. Recent work within artificial intelligence has shown that there exists a variety of methods that allow the successful recovery of information about the three-dimensional world from one or more two-dimensional images. These methods include, for example, the recovery of shape from stereo [21, 10], of shape from motion [13], of shape from shading [12], or of shape from texture [16, 14].

One property common to most "shape from" methods seems to be the use of multiple representations of image information, usually delimited by the outputs of multiple spatial or spatio-temporal filters. Further, the choice of such filters seems to be task dependent along with the associated comparison rules which generate estimates of categorical quantities such as edges, etc. [19, 7]. The use of multiple image representations is motivated by the fact that the physical phenomena of the world can lead to events in the filter outputs over a large range of resolutions. No single filter can, however, respond optimally over the whole range of such resolutions.

The analysis of images at several resolution levels leads to the problem of how image information contained within each level is combined. More specifically, some physical phenomenon may be limited to a certain resolution band and thus will produce an event in only a subset of the layers, and it is within this context that one is faced with the problem of matching corresponding events between layers. One example of such a matching rule is the "spatial coincidence assumption" [20] which is used in matching the zero-crossings of differently sized $\nabla^2 G$ filters where it is assumed that zero-crossings localized identically over adjacent filters are due to a single physical phenomenon.

*Supported by Grant 81.166.0.84 of the Swiss National Science Foundation.
†Supported by grant A2568 from the Natural Sciences and Engineering Research Council of Canada.

0734-189X/88 $3.00

As a generalization of such processes, Witkin [26, 27] has considered how the complete space of zero-crossings of continuously scaled filters may be adequate to determine images, per se, or to define specific image properties. It is this generalization which is of focal interest to this paper. Witkin's approach is based on the notion of scale-space which, for one-dimensional signals, is defined as follows. Let

$$E(x, \sigma) = \frac{\partial^2 G(x, \sigma)}{\partial x^2} * I(x) \tag{1}$$

correspond to the convolution $(*)$ of the input signal $I(x)$ with the second-partial derivative of a Gaussian filter:

$$G(x; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2(x/\sigma)^2}. \tag{2}$$

The filter "scale" is determined by σ, being the standard deviation of the (unit area) Gaussian filter, and, as will be seen, the differentiation process has the effect of converting this low pass kernel (2) into a band-pass filter. The *scale space* $S(x, \sigma)$ of the signal is then defined as the set of all *zero-crossings* (or level-crossings), where

$$S(x, \sigma) = \left\{(x, \sigma) | E(x, \sigma) = 0 \text{ and } \frac{\partial E}{\partial x} \neq 0, \sigma > 0\right\}. \tag{3}$$

These points in scale space form continuous curves that are either open contours (extending from a neighbourhood of $\sigma = 0$ to ∞: great circles) or are convex extending from an open neighbourhood of $\sigma = 0$ to a finite maximum, $\sigma_{\max}$.

Witkin's notion of scale-space has created considerable interest in the literature and there are several important results or properties of the scale-space that apply both to one- and two-dimensional signals. Yuille and Poggio [29] proved that the scale-space of the zero- (or level-) crossings, of almost all signals filtered by a Gaussian, determines the signal uniquely up to a scaling constant. Further, Rotem and Zeevi [22] have shown under what conditions and how 2D-signals can be recovered from their zero-crossings. The importance of this lies in the fact that—for almost all signals—no information is lost by working in scale-space rather than in the image domain. Yuille and Poggio [29] and Koenderink [17] proved that the Gaussian does not create zero-crossings as the scale σ increases, and that the Gaussian is the only filter with this "nice" scaling behaviour. This, in turn, has important consequences on the way the scale-space of images can be analyzed, as is discussed below.

2. PARSING SCALE-SPACE

To evaluate the usefulness of scale-space as an advanced image representation, we must know how easy it is to locate and extract "important" image properties from scale-space. What "important properties" are is largely determined by what subsequent "shape from" and related processes require. In the case of "shape from texture," for example, we are interested in locating boundaries of objects or, more generally, "inside" boundaries of more or less homogeneous image patches. The notion of scale-space is thus useful only in conjunction with rules for *parsing scale-space* into more advanced representations used by later processing modules.

Witkin [23, 27] has proposed a method for parsing scale-space, *stability analysis*, that allows the extraction of primitive events that occur over a large range of scales and to organize them into a qualitative signal description capturing major events in the signal. Every zero-crossing can be described by a scale-value, the σ-coordinate of its apex point, and an interval bounded by the two positions of the zero-crossing curve for $\sigma \to 0$. Assuming that zero-crossing curves never cross each other, one can define the stability of a zero-crossing curve as the difference between its scale value and the largest scale value of zero-crossing curves contained in its interval. Reducing the scale-space to zero-crossing curves whose stability exceeds a certain threshold then leads to a signal description that—intuitively—captures the "major features" of the signals, as can be seen in Witkin [27].

One problem with Witkin's stability analysis is that it is based on the assumption that zero-crossing curves do not cross each other, an assumption that has recently been shown to be false [15]. Another problem occurs when one tries to extend Witkin's stability analysis to two dimensions:

Two-dimensional zero-crossing surfaces can, and in practice often do, split and merge as the scale σ is changed [24]; a nice example of which is shown in Babaud, Witkin, and Duda [1]. There is, consequently, no topologically simple region associated with a zero-crossing surface and so Witkin's stability criterion can no more, if at all, be defined in a simple way. Further, tracing a two-dimensional zero-crossing surface from its peak downwards across scales becomes computationally much more difficult than the simple curve tracing for one-dimensional signals.

3. SPATIAL STABILITY ANALYSIS

We propose an alternative definition of zero-crossing stability, *spatial stability*, which does not require the "nice" scaling behaviour of Witkin's stability criterion. The *spatial stability*, $S_\rho(\chi)$, of a position χ in an image $I(\chi)$ is defined as the greatest depth of zero-crossings over σ in scale-space with respect to a given neighbourhood of radius ρ about χ. Our definition is based on the assumption that important physical events can be conceived of as (generalized) boundaries and, for stability, these boundaries must satisfy two constraints:

(A1) A boundary is a region of steep gradient and high contrast.

(A2) A boundary is well-defined if it has no neighbouring boundaries.

Assumption A1 implies that boundary edges have a broad spectrum and so the corresponding zero-crossing will exist at multiple scales. Assumption A2 guarantees that the zero-crossing position of a boundary edge is not affected by neighbouring boundaries.

Without loss of generality, in this case, support for (A1) comes from the observation that the Fourier transform ($\mathcal{F}$) of the second derivative of the one-dimensional Gaussian filter ((1) and (2)) is

$$\mathcal{F}\left(d^2G/dx^2\right) = -4\pi^2u^2\exp\{-2\pi^2\sigma^2u^2\}, \tag{4}$$

where u corresponds to the spatial frequency (in picture cycles) coordinate. It is readily seen from (4) that the peak frequency of this filter occurs at

$$d\mathcal{F}\left(d^2G/dx^2\right)du = 0,$$

that is, when

$$u = \frac{\sqrt{2}}{2\pi\sigma}. \tag{5}$$

Consequently, by decreasing σ the filter (peak) moves up the spectrum and so images with broad spectra are more likely to have broader ranges ("deeper" in scale-space) of zero-crossings.

The neighbourhood constraint (A2) comes about via the well-known observation that in scale space (due to the different filter scales) spatially proximate zero-crossing contours are more likely to dynamically change each other to form new curves, particularly those of the convex type: ideally, spatially stable edges should have stable zero-crossing contours. This spatial stability criterion of zero-crossings is an extension of the "spatial coincidence assumption" [20, 19]:

> If a zero-crossing segment is present in a set of independent $\nabla^2 G$ channels over a contiguous range of sizes, and the segment has the same position and orientation in each channel, then the set of such zero-crossing segments indicates the presence of an intensity change in the image that is due to a single physical phenomenon (a change in reflectance, illumination, depth, or surface orientation). [19, p. 70]

Not only do we assume that zero-crossings localized identically over adjacent filters are due to a single physical phenomenon, but that the *range* of filters, as measured by the spatial stability index, is an appropriate indicator of the importance of that particular physical phenomenon for describing the signal.

3.1. *Implementation of Spatial Stability Analysis*

Before we discuss further properties of (and extensions to) spatial stability analysis, we present the algorithm for computing spatial stability and a few examples.

It is obvious that spatial stability analysis cannot be implemented with a continuous variation of the filter size and that the σ-space has to be sampled at regular, small intervals. We have obtained satisfactory results using σ-sampling intervals of $\frac{1}{4}$- or $\frac{1}{8}$-octaves. Further, it has already been found in a number of psychophysical experiments [5, 3] and spectral compression algorithms [6] that the human visual system cannot apparently discriminate spatial information produced from spatial components closer than this $\pm \frac{1}{8}$ octave distance.

All analyses were done using an Imaging Technology image processing system on a PDP 11/23 computer. Scenes containing natural objects were digitized, via a frame grabber, as 512×512, 8-bit pixel images, which then were analyzed to first extract zero-crossings and then spatial stability via parallel pipeline pixel processor operations.

Since this project was specifically aimed at investigating two-dimensional images, we have used the (usual) Laplacian or ∇^2 operator to represent the second derivatives of an isotropic Gaussian since it is also an isotropic operator (see [23]). Accordingly, the image was first filtered by a digital approximation of the $\nabla^2 G(x, y, \sigma)$ filter (see [18] for details):

$$\nabla^2 G(x, y, \sigma) = \left(\frac{r^2 - 2\sigma^2}{2\pi\sigma^6}\right)\exp\left(\frac{-r^2}{2\sigma^2}\right), \qquad r^2 = x^2 + y^2 \tag{6}$$

for image cartesian coordinates (x, y). σ was varied in a 1–16 pixel range in both $\frac{1}{4}$ and $\frac{1}{8}$ octave increments yielding 17 and 33 scale space "slices," respectively. The discrete approximation to $\nabla^2 G$ was chosen in such a way as to maximize the resolution and support of the kernel without producing an overflow in the 16-bit output of the pipe-line pixel processor. The kernel sizes varied from $7*7$ pixels ($\sigma = 1$ pixel) to $85*85$ pixels ($\sigma = 16$ pixels). Zero-crossings were then located in each scale space slice. Zero-crossings with small gradients may be located within extended zero-value areas due to the limited resolution of the convolution. These zero areas were therefore removed by alternatively growing positive and negative regions until all pixels were classified as being either positive or negative. This ensured that no zero-crossing contour was incomplete. Figure 1 shows zero-crossing images obtained with filter sizes of $\sigma = 1$, $\sigma = 4$, and $\sigma = 16$ pixels (Fig. 1b–d).

The computation of the spatial stability index $S_\rho(x, y)$ from the (zero-crossing) scale-space is straightforward. Let

$$z_\rho(x, y, \sigma) = \begin{cases} 1, & \text{if there exists a zero-crossing} \\ & \text{in the } \rho\text{-neighbourhood} \\ & \{(x', y')/((x', y') - (x, y))^2 \le \rho^2\} \\ & \text{of } (x, y) \text{ for filter size } \sigma, \\ 0, & \text{otherwise.} \end{cases}$$

For a sequence of "zero-crossing slices" $Z_\rho(x, y, \sigma_1) \cdots Z_\rho(x, y, \sigma_n)$ the spatial stability index $S_\rho(x, y)$ then corresponds to the length l of the longest subsequence $Z_\rho(x, y, \sigma_i) \cdots Z_\sigma(x, y, \sigma_{i+l-1})$ such that

$$\prod_{\sigma_i \le \sigma' \le \sigma_{i+l-1}} Z_\sigma(x, y, \sigma') = 1.$$

Thus to determine $S_0(x, y)$ we simply compute the maximum number of zero-crossing slices along the σ direction having a zero-crossing at position (x, y). To determine $S_\rho(x, y)$ for $\rho > 0$ we first compute $Z_\rho(x, y, \sigma)$ by OR-ing $Z_0(x, y, \sigma)$ within the neighbourhood ρ and then compute the maximum number of adjacent σ-slices with $Z_\rho(x, y, \sigma) = 1$ as before. The resultant spatial stability image is shown in Fig. 1e, where the intensity at any pixel is proportional to its spatial stability.

3.2. *Properties of Spatial Stability Analysis*

The computation of spatial stability within a fixed neighbourhood does not take account of the fact that the accuracy of zero-crossing localization decreases with filter size and that even the zero-crossing of an isolated edge embedded in white noise may be be spatially shifted. To overcome this problem we introduce an aperture for position uncertainty which is proportionate to σ.

Let I be an isolated step edge of contrast C embedded in white noise of variance n^2, and let $\text{var}(Z_\sigma)$ be the variance of the zero-crossing position $(d^2/dx^2)G_\sigma * I$. It can be shown [8] that

$$\text{Var}(Z_\sigma) \propto \sigma \frac{n^2}{c^2}. \tag{7}$$

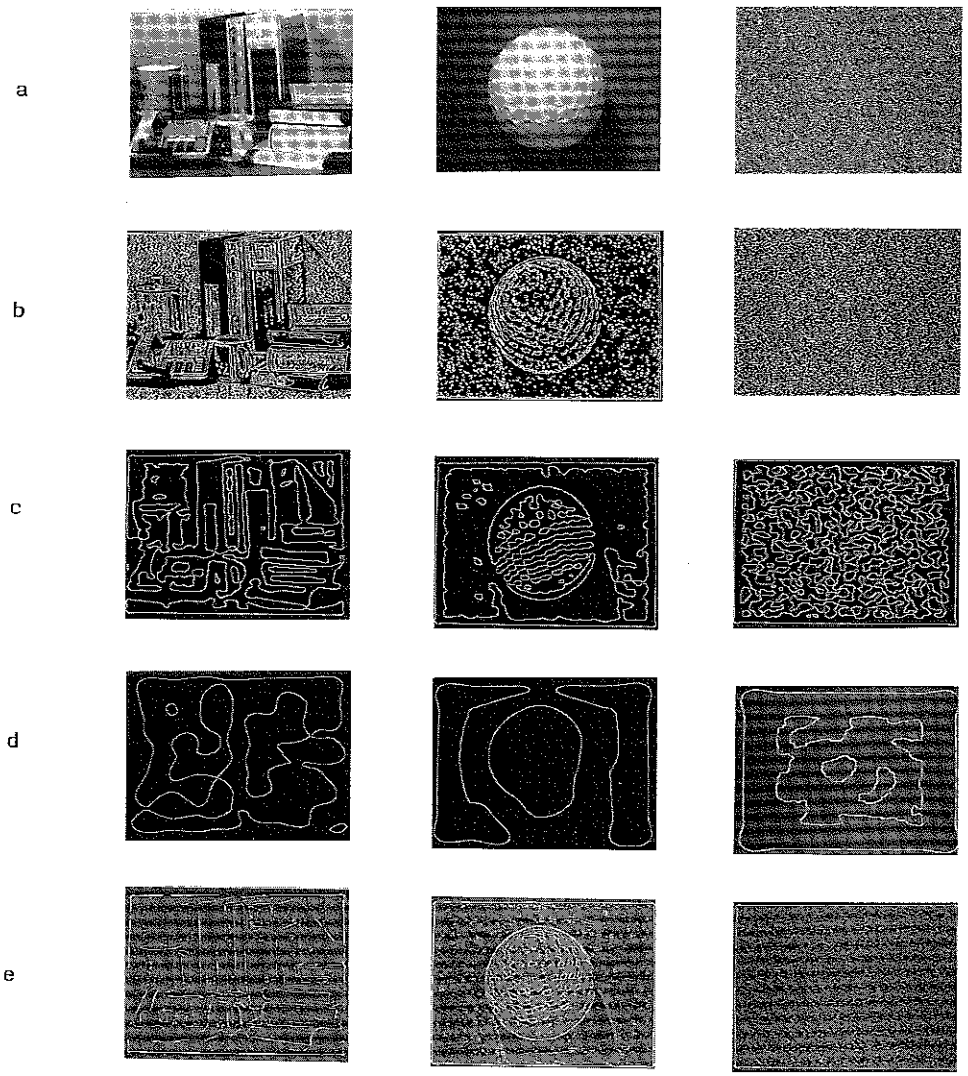

FIG. 1. (a) Two natural (golf ball and indoor scene) and one white noise image of size 256 × 256 pixels at 8-bit resolution. (b) Zero-crossings obtained with filter size $\sigma = 1$ pixels. (c) Zero-crossings obtained with filter size $\sigma = 4$ pixels. (d) Zero-crossings obtained with filter size $\sigma = 16$ pixels. (e) Spatial stability images for neighbourhood radius $\rho = 0$, with pixel intensity being proportional to the spatial stability value.

One way to account for this localization uncertainty is to vary the stability neighbourhood with σ, i.e., to blur the zero-crossing image with a filter whose size varies with σ, before computing spatial stability. Figure 2 shows the spatial stability image with spatial stability neighbourhood varying proportionally to σ^0, $\sigma^{1/2}$, and σ^1.

It is interesting to see that these "cones of uncertainty" have little effect on the spatial stability images, mainly due to the fact that white noise is, by definition,

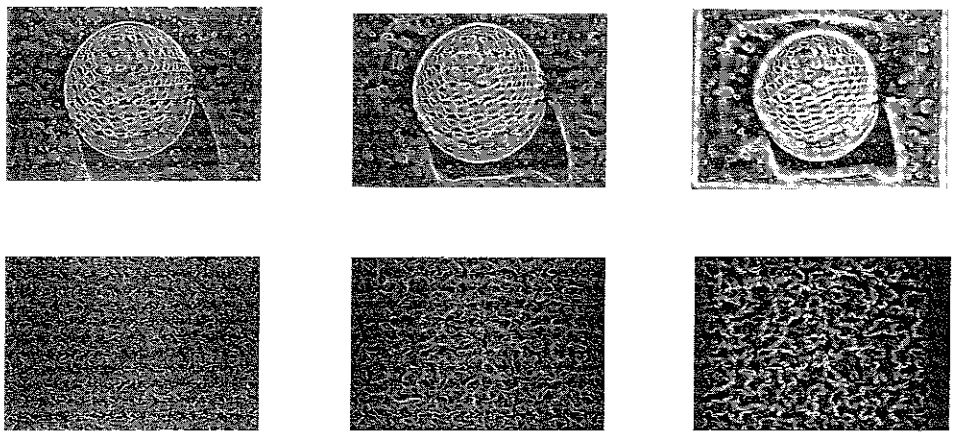

FIG. 2. (a) Spatial stability image with stability neighbourhood radius varying proportionally to σ^0. (b) Spatial stability image with stability neighbourhood radius varying proportionally to $\sigma^{1/2}$. (c) Spatial stability image with stability neighbourhood radius varying proportionally to σ.

uncorrelated. Hence estimates of zero-crossing presence is not improved by considering edge presence in broader (uncorrelated) neighbourhoods. Conversely, we found this criterion not to be useful with unperturbed shapes due to the degree of correlations present. That is, in both situations there is a trade-off between misses and false alarms over increasing window sizes.

In general the differences between stability in white noise and more naturally occurring images is illustrated in Fig. 3. As expected the white noise images are less stable and stable points are randomly distributed over the image.

To show that spatial stability is resilient to additive white noise we note that it is a result of the correlation between the various scale-space slices of the image. It is therefore possible to relate the shapes of spatial stability maps to the ensemble of cross-correlations between filtered versions of the input. We first note that the $\nabla^2 G(\chi, \sigma)$ operation is isotropic and so the cross-correlation between two such versions (σ_1, σ_2 space constants or scales) is determined by the Fourier series (inverse discrete Fourier transform)

$$C_{12}(\chi) = \int\int A_1(\mathbf{u}) \cdot A_2(\mathbf{u}) \cdot A^2(\mathbf{u}) \exp\{2\pi(\mathbf{u} \cdot \chi)\}\, d\mathbf{u}, \tag{8}$$

where $\chi \equiv (x, y)$ and $\mathbf{u} \equiv (u, v)$ are the image and spatial frequency domain co-ordinate systems, respectively. A_1 and A_2 correspond to the Fourier amplitude spectra of the two $\nabla^2 G$ filters (having zero phase spectra) while $A^2(\mathbf{u})$ corresponds to the image power spectrum. Clearly, as A_1 and A_2 become more spectrally disjoint, as they do as σ_1 and σ_2 differ, then the cross-correlation image attenuates. Further, for white noise, $A^2(\mathbf{u})$ is a constant, and so the "disjointness' of the filters is more emphasized as the σ's differ relative to that for normal images where $A^2(\mathbf{u})$ is generally monotonically decreasing from the spectral (DC) centre.

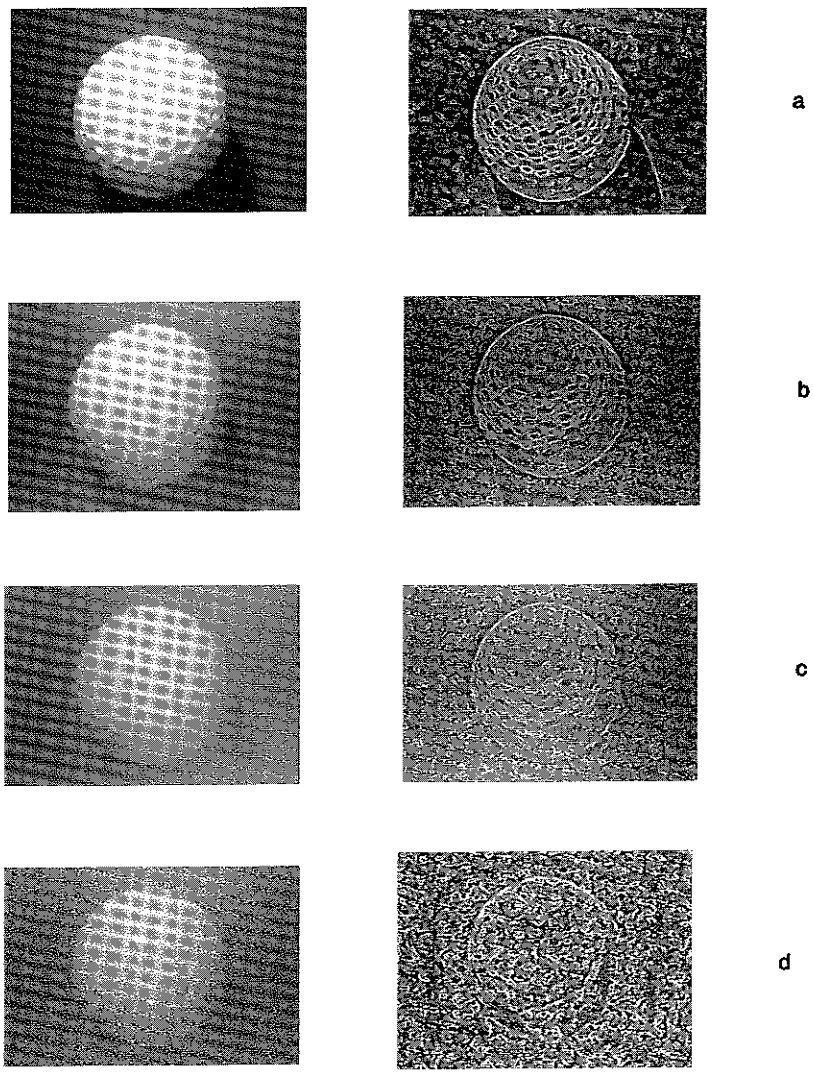

FIG. 3. Spatial stability images of patterns embedded in white noise: (a) SNR = ∞; (b) SNR = 5; (c) SNR = 3; (d) SNR = 1.

Alternatively (8) demonstrates that the cross-correlation image consists of the cross-correlation between the filters and the image autocorrelation image, due to the commutability and associativity of the cross-correlation process:

$$\begin{aligned} C_{12}(\mathbf{X}) &= \left(\nabla^2 G(\chi, \sigma_1) * I\right) * \left(\nabla^2 G(\chi, \sigma^2) * I\right) \\ &= \left(\nabla^2 G(\chi, \sigma_1) * \nabla^2 G(\chi, \sigma_2)\right) * (I * I). \end{aligned} \tag{9}$$

Now, for white noise $I * I \equiv \delta(x)$, the Dirac δ-function, reducing $C_{12}(\chi)$ to

$$C_{12}(\chi) = \nabla^2 G(\chi, \sigma_1) * \nabla^2 G(x, \sigma_2), \tag{10}$$

and using the unit area version of $\nabla^2 G$ (2) this, in one dimension, reduces to

$$C(x) = \frac{1}{\sigma^5\sqrt{2\pi}}(x^4 - 6x^2 + 3)e^{-x^2/2\sigma^2}, \qquad \text{for } \sigma^2 = \sigma_1^2 + \sigma_2^2, \tag{11}$$

the direct cross-correlation between the filters whose likelihood of correspondence decreases with the separation of theier spectral peaks, or σ^2 in (11).

For these reasons we can see that spatial stability has a "noise cleaning" effect, or is resilient to additive white noise in the sense that any process that combines "labelled" filter outputs (e.g., zero-crossings) as a function of their *correspondence* (correlation) will have weak and randomly distributed responses to white noise.

4. PARSING SCALE-SPACE FOR TEXTURAL ANALYSIS

In Section 2 we argued that the notion of "scale-space" cannot be justified independently of rules for parsing scale-space into more advanced representations suitable for later processes. Our line of justification is based on the fact that spatial stability analysis allows the extraction and localization of region boundaries independent of the level of resolution at which they occur. Further, one can show [4] that efficient pattern learning and classification can be accomplished by means of cross-correlational processes based on the stability map of images. Second, the stability map of images can be used to construct qualitative image descriptions capturing major image events, using methods similar to Witkin's [24] stability analysis.

In this section we outline two ways to parse scale-space in conjunction with spatial stability analysis, aimed at constructing an image representation suitable for "shape from texture" methods. These methods are based on the assumption that object surfaces are homogeneous with respect to some local geometric measure. The spatial variation of this measure can then be used to obtain constraints on the orientation of local surface patches, given the additional assumption that discontinuities in surface orientation occur rarely. First, in the case of regular patterns, constraints on local surface orientation are derived from the spatial variation of size, orientation, or distortion of the texture elements that are assumed to cover the surface [16]. Second, in the case of arbitrary textures, constraints on local surface orientation may be derived from some other local geometric measure [25, 9].

In the following we outline how both these approaches can be improved using spatial stability analysis. First, in the case of regular patterns, we discuss how to use spatial stability analysis for the extraction of texture elements from images, independent of their level of resolution. Second, in the case of arbitrary textures, we outline how spatial stability analysis can be used to improve the use of local geometric measures, again independent of their level of resolution.

4.1. *Regular Patterns*

For regularly textured patterns two problems must be addressed before any "shape from texture" method can be applied—problems which have rarely been addressed in the literature. First, how are the texture elements extracted from an image? Second, the boundary of a textured region must be determined in order to confine the interpretation of textural variation to that region. Both the textured boundaries and the texture elements may arise at any level of resolution, and in the

following we attempt to solve both problems by parsing scale-space in conjunction with spatial stability analysis.

Rephrasing the problems, we have to parse scale-space into a representation making explicit boundaries of homogeneous regions and making explicit the texture elements contained within each region. The texture elements can, in turn, be considered as textured regions, containing smaller texture elements: the notion of "texture" is *relative* to the enclosing boundary.

To express region containment we define a topological *region nesting level* $T(x, y)$ as follows:

$$T(x, y) = \begin{cases} 0, & \text{if } (x, y) \text{ belongs to the absolute image boundary,} \\ 1, & \text{if } (x, y) \text{ is contained "inside" the first zero-crossing contour,} \\ n, & \text{if } (x, y) \text{ is contained "inside" the } n\text{th zero-crossing contour.} \end{cases}$$

The computation of the region nesting level is straightforward in the case of a zero-crossing map, i.e., one scale-space slice. Starting with level $T = 0$ on the image boundary, the current nesting level is propagated with a region growing algorithm up to a zero-crossing contour. Once this region growing cycle terminates, T is then increased by one and the new nesting level is propagated. The algorithm terminates when all pixels have been classified. The topological region nesting level $T(x, y)$ can then be used to define the texture elements to some region R: under ideal conditions, $T(x, y) \geq t$ for all pixels of region R, then $T(x, y) \geq t + 1$ for all texture elements of that region. This straightforward extraction of texture elements requires, however, knowledge of which level of resolution the texture elements arise, as it is based on a single filter size.

If the level of resolution is not known then spatial stability analysis can be used to extract texture elements, assuming that both the texture region and its texture elements are bounded by boundaries satisfying assumptions (A1) and (A2). We now outline how a modified topological region nesting level $T^*(x, y)$ can be used to extract texture elements, independent of the level of resolution.

In the case of a single zero-crossing map, the topological region nesting level $T(x, y)$ is well defined since all zero-crossing curves are either closed or cross the image boundary [24]. The latter does not occur with a finite image and filter size and with a large enough zero-crossing map. However, the situation becomes more complex if the topological region nesting level is defined for regions with *stable* boundaries only. Thresholding boundaries with some minimum stability value does not guarantee that the boundaries are closed curves. One way to solve this problem is to complete stable contours by means of an interpolation scheme. There is, however, another way to define a topological region nesting level on the spatial stability map.

Assume that the scale-space of some image has been sampled in n scale-space slices. For each of the slices $\{S_i(x, y), i = 1, \ldots, n\}$ the topological region nesting level $\{T_i(x, y), i = 1, \ldots, n\}$ is well defined. Then we define the *modified topological*

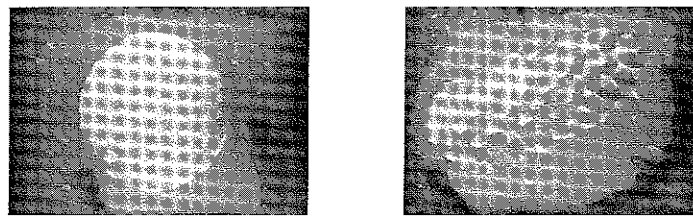

FIG. 4. Topological region nesting level map of two images, a golf ball and a pine cone. The pixel intensity is proportional to $T^*(x, y)$ (see text).

region nesting level over the whole scale space, $T^*(x, y)$ as

$$T^*(x, y) = \frac{1}{n} \sum_{i=1}^{n} T_i(x, y). \tag{12}$$

The region nesting level map of two images is shown in Fig. 4, with pixel brightness being proportional to $T^*(x, y)$. The maps suggest that this analysis can successfully recover homogeneous regions and texture elements contained in each region. It turns out, however, that the region nesting level $T(x, y)$ defined for stable regions, cannot be easily recovered from the T^*-map. First, the T^*-map does not meet the requirement that these regions be bounded by stable edges, and hence regions may be created in the T^*-map that are not due to a single physical phenomenon, such as the upper and lower half of the golf ball in Fig. 4a, that are segregated due to different illumination conditions. Second, for some region that is not detected at all levels of resolution $T^*(x, y) < T(x, y)$. For both reasons, $T(x, y)$ can, at best, be obtained from $T^*(x, y)$ through some nonlinear transformation. Like others, to this stage we cannot see how such topological measures can be further used, per se, to parse image information unless extra knowledge is introduced or such results are, in various ways, compared to our stability results or even the differential geometry of the image. Indeed, computations dependent on "closed" or "regular" features are unreliable in image understanding and, as with stability analysis, our approach is to develop statistical measures for the presence of structures physically present rather than impose artificial object interpretation schemes on the image.

4.2. *Arbitrary Textures*

In the more general case of arbitrarily textured patterns, "shape from texture" methods can still be applied, assuming that object surfaces are invariant with respect to, at least, one local geometric measure, and assuming that this measure can be identified. Further, the principle of textural homogeneity can be weakened to the assumption that the *distribution* of some local geometric measure remains invariant over a surface. One approach to textural interpretation that is based on the latter assumption is Witkin's [25] analysis of the texture edge orientation distribution. Under the assumption that the distribution of edge orientations is isotropic one can derive (probabilistic) constraints on the local surface orientation.

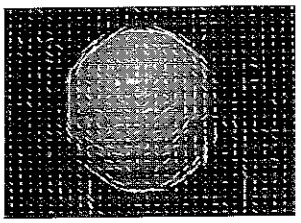

FIG. 5. Image with superimposed edge orientation lines. For display purposes the edge orientation lines are shown on a coarse grid. The length of the edge orientation lines is proportional to their spatial stability value.

As in the case of texture element identification for regular patterns, we are faced with two problems that precede the application of "shape from texture" interpretations. First, the appropriate geometric measure can arise at any level of resolution, and second, we have to guarantee that the measure is determined by a single physical phenomenon. In this case of edge orientation distributions, for example, the appropriate level of resolution is related to the granularity of the surface texture and surface orientation constraints should be derived only from edges that are clearly related to the surface granules. Both problems can again be dealt with by parsing scale-space using spatial stability analysis that allows the identification of texture-related edges independent of the level of resolution at which they arise.

The usefulness of spatial stability analysis for the generalized texture analysis is based on the assumption that edges near the appropriate level of resolution correspond closely between adjacent scale-space slices, whereas edges at other levels of resolution are more or less unrelated since they do not correspond to a single physical phenomenon. Thus spatial stability for texture edges related to the natural texture of an object is expected to be higher than that for other texture edges.

The approach we followed consisted essentially of extracting local edge orientation from the spatial stability map, weighting the orientation map linearly with their spatial stability value. Figure 5 shows the edge orientation map overlayed on the original image. The length of the edge orientation lines is proportional to their spatial stability value. For display purposes, edges have been sampled on a very coarse grid.

5. DISCUSSION

As has been pointed out in Section 1, scale-space determines the image uniquely, up to a scaling constant [28]. It is, however, not obvious how to recover the image, per se, or more generally, how to recover a given image property from scale-space. Spatial stability analysis has been shown to lead to successful recovery of region boundaries, without relying on "nice" scaling properties of scale-space. Further, it has been shown that spatial stability analysis leads to an edge-only based image representation characterized by resolution independence and good noise resilience. Both these characteristics support its benefits for pattern recognition applications. Within the framework of "shape from texture" methods we have provided evidence that spatial stability analysis solves some problems that have hardly been addressed at all in literature.

Spatial stability analysis can be regarded, in some way, as an exercise in how to recover image information from scale-space. In a more general framework it may, however, turn out that *extending* the scale-space with descriptions obtained from the image directly may be more appropriate than deriving these descriptions from scale-space.

This claim holds trivially for the zero-crossing gradient, a piece of image information which is lost in the scale-space representation. This inherent ambiguity of scale-space based image representations, or, in fact any edge-only based image representation, can, for example, lead to deterioration or even breakdown in pattern matching in situations where pattern information is carried mainly in contrast symmetries. In our current work we have *augmented* scale-space by local geometric properties of the input image at the zero-crossing positions, as principle directions and curvature information. Such additions are particularly useful for "disambiguating" stable edges extracted from coarsely sampled scale-space.

The space of possible enhancements of edge-only based image descriptions is large; examples include the "topographic" labeling of the image intensity surface [11] or the enhancement of image descriptions with differential geometry primitives as discussed by Besl and Jain [2]. Both approaches require, however, further investigations with respect to their suitability within a multi-resolution (e.g., scale-space based) framework. Further, these approaches can only be properly evaluated in terms of their usefulness for *later processes* operating on these representations.

REFERENCES

1. J. Babaud, A. Witkin, and R. Duda, Uniqueness of the gaussian kernel for scale-space filtering, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**, 1986, 26–33.
2. P. J. Besl and R. C. Jain, Invariant surface characteristics for 3D object recognition in range images, *Comput. Vision Graphics Image Process.* **33**, 1986, 33–80.
3. T. M. Caelli, On the specification of coding principles for visual image processing, in *Figural Synthesis* (P. C. Dodwell and T. M. Caelli, Eds.), Erlbaum, Hillsdale, NJ, 1984.
4. T. M. Caelli, W. F. Bischof, and Z.-Q. Liu, Filter-based models for pattern classification, submitted for publication, 1987.
5. T. M. Caelli, H. Brettel, I. Rentschler, and R. Hilz, Discrimination thresholds in the two-dimensional spatial frequency domain, *Vision Res.* **23**, (No. 2), 1983, 129–133.
6. T. M. Caelli and M. Hübner, Coding images in the frequency domain: Filter design and energy processing characteristics of the human visual system, *IEEE Trans. Systems Man Cybernet.* **SMC-13**, 1983, 1018–1021.
7. T. M. Caelli and G. Moraglia, On the detection of signals embedded in natural scenes, *Percept. Psychophys.* **39** (No. 2), 1986, 87–95.
8. J. F. Canny, *Finding Edges and Lines in Images*, Technical Report TR-720, MIT Artificial Intelligence Laboratory, Boston, 1983.
9. L. Davis, L. Janos, and S. Dunn, Efficient recovery of shape from texture, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-5**, 1983, 485–492.
10. W. E. L. Grimson, *From Images to Surfaces: A Computational Study of the Human Early Visual System*, MIT Press, Cambridge, 1981.
11. R. M. Haralick, L. T. Watson, and T. J. Laffey, The topographic primal sketch, *Int. J. Rob. Res.* **2**, 1983, 50–72.
12. B. K. P. Horn and M. J. Brooks, The variational approach to shape from shading, *Comput. Vision Graphics Image Process.* **33**, 1985, 174–208.
13. B. K. P. Horn and B. G. Schunck, Determining optical flow, *Artif. Intell.* **17**, 1981, 185–203.
14. K. Ikeuchi, Shape from regular patterns, *Artif. Intell.* **22**, 1984, 49–75.
15. I. Katz, *Coaxial Stereo and Scale-Based Matching*, Tech. Rep. 85-13, Department of Computer Science, University of British Columbia, Vancouver, BC, Canada, 1986.

16. J. R. Kender, *Shape from Texture*, Tech. Rep. CMU-CS-81-102, Department of Computer Science, Carnegie-Mellon University, Pittsburgh, 1980.
17. J. J. Koenderink, The structure of images, *Biol. Cybernet.* **50** 1984, 363–370.
18. W. H. H. J. Lunscher and M. P. Beddoes, Optimal edge detector coefficients. II. Coefficients quantization, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**, 1986, 178–187.
19. D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman, San Francisco, 1982.
20. D. Marr and E. Hildreth, Theory of edge detection, *Proc. R. Soc. London B* **200**, 1980, 269–294.
21. D. Marr and T. Poggio, A computational theory of human stereo vision, *Proc. R. Soc. London B* **204**, 1979, 301–328.
22. D. Rotem and Y. Y. Zeevi, Image reconstruction from zero crossings, *IEEE Trans. Acoustics Speech Signal Process.* **ASSP-34**, 1986, 1269–1277.
23. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Vol. 1, 2nd ed., Academic Press, New York, 1982.
24. V. Torre and T. A. Poggio, On edge detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**, 1986, 147–163.
25. A. P. Witkin, Recovering surface shape and orientation from texture, *Artif. Intell.* **17**, 1981, 17–45.
26. A. P. Witkin, Scale-space filtering, in *Proceedings, Int. Joint Conference on Artificial Intelligence, Karlsruhe, West Germany, August 1983*, pp. 1019–1022, Kaufmann, Palo Alto, CA, 1983.
27. A. P. Witkin, Scale-space filtering: A new approach to multi-scale descriptions, in *Image Understanding 1984* (S. Ullman and W. Richards, Eds.), Ablex, Norwood, NJ, 1984.
28. A. L. Yuille and T. A. Poggio, Fingerprint theorems for zero crossings, *J. Opt. Soc. Amer. A* **2**, 1985, 683–692.
29. A. L. Yuille and T. A. Poggio, Scaling theorems for zero-crossings, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**, 1986, 15–25.