A Transform for Multiscale Image Segmentation by Integrated Edge and Region Detection

Narendra Ahuja, *Fellow, IEEE Computer Society*

Abstract—This paper describes a new transform to extract image regions at all geometric and photometric scales. It is argued that linear approaches such as convolution and matching have the fundamental shortcoming that they require a priori models of region shape. The proposed transform avoids this limitation by letting the structure emerge, bottom-up, from interactions among pixels, in analogy with statistical mechanics and particle physics. The transform involves global computations on pairs of pixels followed by vector integration of the results, rather than scalar and local linear processing. An attraction force field is computed over the image in which pixels belonging to the same region are mutually attracted and the region is characterized by a convergent flow. It is shown that the transform possesses properties that allow multiscale segmentation, or extraction of original, unblurred structure at all different geometric and photometric scales present in the image. This is in contrast with much of the previous work wherein multiscale structure is viewed as the smoothed structure in a multiscale decimation of image signal. Scale is an integral parameter of the force computation, and the number and values of scale parameters associated with the image can be estimated automatically. Regions are detected at all, a priori unknown, scales resulting in automatic construction of a segmentation tree, in which each pixel is annotated with descriptions of all the regions it belongs to. Although some of the analytical properties of the transform are presented for piecewise constant images, it is shown that the results hold for more general images, e.g., those containing noise and shading. Thus the proposed method is intended as a solution to the problem of multiscale, integrated edge and region detection, or low-level image segmentation. Experimental results with synthetic and real images are given to demonstrate the properties and segmentation performance of the transform.

Index Terms—Image segmentation, representation, scale-space, edge detection, region detection, perceptual structure, pyramids, medial axis, nonlinear image analysis, texture.

———————————— ✦ ————————————

1 INTRODUCTION

THIS paper is concerned with the problem of low level image segmentation, or partitioning of an image into regions, that represent low level image structure. A region is characterized as possessing a certain degree of interior homogeneity and a contrast with the surround which is large compared to the interior variation. This is a satisfactory characterization from both perceptual and quantitative viewpoints. Homogeneity and contrast may be defined differently: A region may be uniform, in which case its contrast with the surround must be large; alternatively, a region may be shaded, in which case the local contrast across a boundary point must be large compared to the interior variation on each side. The sizes, shapes, types of homogeneity, and contrast values of regions in an image are a priori unknown. To illustrate, consider the image shown in Fig. 10a, which shows people at a 3D movie. Within the lower of the two windows shown, the lenses and the rest of the head form two regions having homogeneous intensities and large contrasts with their surrounds.

• *The author is with Beckman Institute, Coordinated Science Laboratory, and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801.*
E-mail: ahuja@stereo.ai.uiuc.edu.

If the amount of acceptable variability and the contrast value of the regions are reduced, the frames of the eyeglasses and the ear emerge as new, smaller regions. If the homogeneity and sensitivity to contrast are increased further, nose and lips separate as regions from the rest of the face. Thus, a decrease in the acceptable contrast leads to increased, hierarchical decomposition, which culminates in constant-value regions at the bottom of the hierarchy. In another window in the same image, e.g., the upper window shown in Fig. 10a, the depth of the hierarchy and the region homogeneity, contrast, shape and size parameters associated with the different levels are unrelated to the corresponding values in the lower window. Thus, in the tree representing the entire image, features such as depth and branching factor are unrelated across subtrees, each solely determined by the image and therefore a priori unknown. The homogeneity and contrast parameters associated with different image regions will be said to form the set of photometric scales present in the image, while the region shapes and sizes will be said to define the geometric scales present.

Finding a solution of the low level segmentation problem poses two main challenges. First, a valid image region must be detected regardless of its shape, size, type of homogeneity, and contrast. Second, all geometric and photometric scales at which regions happen to occur across an image must be identified. If these two problems are solved,

the result will be a segmentation tree representing the multiscale, low level, image structure. To obtain such a tree for an arbitrary gray-scale image is the objective of image segmentation pursued in this paper.

Limited work has been done to meet both of the above challenges. Much of the previous work on multiscale analysis is concerned with a scale-space decomposition of the image signal, determined by a single scale parameter. The decomposition amounts to a blurring of the image to different degrees. The image structure is present across this scale-space continuum, and methods for extracting image regions of different sizes and contrasts from this continuum are not available. Further, even if they were extracted, the regions in the different decompositions would be correspondingly deformed. Automatic estimation of scale parameters is typically not addressed. Even at a given scale, robust detection of a region continues to be an area of active investigation, mainly through the work on edge detection. Region detection such that the detected boundaries are closed and coincident with the true region boundaries regardless of region parameters is not a solved problem. Most methods are linear and often use restrictive region models, e.g., allowed geometric and photometric complexity of edges. Although these models simplify processing, they cause fundamental limitations in the detection accuracy and sensitivity achieved which is partly why the problem of region and edge detection continues to evade a satisfactory solution.

A central theme of this paper is to meet both of the aforementioned challenges, namely, accurate detection of regions without using rigid, geometric, and photometric models and automatic estimation of all scales associated with an image. This is accomplished by introducing a new transform which has two major properties. First, for a specific pair of scale values, the transform leads to well-defined signatures of corresponding image regions which are easy to detect. The transform definition incorporates the duality of interior- and edge-based descriptions of regions. Thus, the transform performs integrated edge and region detection, and can be viewed as a multiscale blob and edge detector at the same time. Second, all scale values for which regions occur in the image can be computed by analyzing the results of the transform as the scale values used are varied.

The transform computes affinities among image *points* or *pixels* for grouping with other pixels, letting the structure emerge bottom-up from "interactions" among the pixels instead of imposing a priori chosen models of region edges and interior. The emergent region geometry is not restricted, since pixels can group together to form any connected set. Specifically, the transform computes a family of force fields for a given image where the force vector at a point denotes its gray level similarity to the rest of the image. On either side of a region boundary, the pixels have high affinities, but there is little affinity between pixels across regions. The strength of interaction between pixels, and consequently their affinities, depend upon their distances and contrasts, and this allows association of the computed affinities and segmentation with spatial and photometric scales. Since the transform allows interaction between a pixel and all other pixels, it can be viewed as

collecting globally distributed evidence for image structure and making it available locally, e.g., at the locations of region edges and medial axes (The medial axis of a region is defined as the locus of points inside the region which are equidistant from two or more points on the region boundary [6], [16].) The regions are encoded in the force field via distinct signatures amenable to robust, local identification. In this sense, the transform performs Gestalt analysis.

This paper introduces the transform and shows how it can be used for segmentation (the basic idea of the transform can be found in [1]). It does not present specific segmentation algorithms. The segmentation is intended to represent low-level image structure at all scales, thus with applicability to textured as well as smooth images. To analyze and illustrate the basic properties of the transform, we model regions, whenever necessary, as possessing uniform gray levels and step edges. However, the transform properties and segmentation results are shown to apply to images containing general types of regions as discussed above, e.g., having shading and noise. Section 2 discusses some basic desired characteristics of segmentation and how they motivate the proposed approach. Section 3 describes the transform, describes some of its properties of interest, and shows how these properties facilitate multiscale segmentation. Section 4 analyzes the segmentation performance of the transform, and Section 5 describes some experiments conducted to demonstrate this performance. Section 6 presents concluding remarks.

2 BACKGROUND AND OBJECTIVES

In this section, we first discuss past work on the two major subproblems of image segmentation: structure detection at a single scale and multiscale analysis (Section 2.1). This leads us to formulate the characteristics desired in a satisfactory segmentation (Section 2.2).

2.1 Two Aspects of Image Segmentation

We will first review the past work on photometric and geometric models of a region used for segmentation at a single scale. Mumford and Shah [25] estimate a 2D functional that minimizes a cost function comprised of the difference between the images and estimated intensity values, the length of detected edges, and the variation in the functional away from edges, which are combined using a priori chosen relative weights. Morphological methods are used in [22] to detect regions as intensity hills in grayscale landscape. More commonly, region detection has been performed through the detection of their boundaries (or edges). An edge separates two different regions and thus two different types of gray level populations. Edge detection methods use different models of edge geometry, and gray level variation along edge as well as within region. These models are fitted to local pixel populations to determine if an edge is present or not. Such local responses are then combined to derive a more global segmentation. Clearly, the validity of the models of the edge as well the gray level populations are critical factors in achieving a valid segmentation. We will now review some models used in the previous work. It is common to treat the problem of edge detection as mainly that of selecting a point

along the intensity profile across edge, assuming such a profile can be extracted from the image. Accordingly, a model of the intensity profile is used to precisely define an edge and to optimally detect its location. Different types of intensity models of an edge have been proposed, according to the nature of the two populations and the spatial profile of the transition from one to the other across the edge [29], [28], [5], [10], [13]. To meet the assumption that edge profile through a pixel can be identified, it is common in edge detection work to implicitly or explicitly use a model of edge curvature. The use of such geometric model constrains the number of possible, different subdivisions of the pixel neighborhood (e.g., a 3×3 or a 5×5) into two regions which must be analyzed to detect the presence of an edge in the neighborhood. For example, the assumption of local straightness of edge is common which makes it very easy to select neighborhoods on the two sides of the edge. Marr and Hildreth [21] assume that the edge is locally straight (and that the intensity changes linearly along a direction parallel to the edge.) Nalwa and Binford [26] assume straightness to extract a sample edge profile. Even the computation of gradient which is common to many edge detectors, e.g., the diffusion based methods [20], [32], implicitly assumes local edge straightness. The same can be said about Laplacian based edge detectors. The use of straightness is very explicit in the different types of discrete edge masks each of which is meant to detect a different edge orientation [30]. To detect intensity facets meeting at an edge [12], a model of edge geometry is required so candidate neighborhoods from each side of the edge can be identified. The work on optimal edge detection (e.g., [8]) is also subject to the validity of the assumed model of the edge geometry. In short, the work on edge detection has lead to different approaches to estimate edge location and orientation for edges having some (implicitly or explicitly) assumed local curvature properties. Image edges do not always conform to these assumptions, and deviations lead to detection errors. Examples of such errors incurred using the Laplacian-of-Gaussian operator for different edge geometries can be found in [4]. To avoid some of these problems, Geman and Geman [11] use the Markov random field model to obtain an estimate similar to that in [25] but allow for end points, corners, and junctions in the edge models used. An approach that avoids the dependence on geometric models of edge is given in [27] where interpixel correlations in spatiotemporal space are considered instead of interwindow correlations.

The second major aspect of segmentation is related to scale. As we stated earlier, scale as defined in this paper is associated with both geometric and photometric sensitivity to detail. Thus, a pixel may simultaneously belong to different regions each having a different contrast value and size, giving rise to the tree representation mentioned earlier. Large regions may be said to have a coarse spatial scale while smaller sizes may be said to be associated with finer spatial scales. Analogously, an edge contour which separates two regions of a given contrast scale may not be detected at a higher scale associated with a larger contrast. The exact number and parameters of scales for a given image are a priori unknown. Therefore, multiscale segmentation must automatically estimate these parameters and de-

tect the corresponding regions. Although the general notion of multiscale operators has been examined for a long time [31], there has been limited work on definition, analysis and automatic estimation of multiscale image structure as pursued in this paper. Our objective here is to separate *original* (unsmoothed) image structure at different scales (regions with different sizes and contrasts), as well as identify the spatial and topological relationships among the regions. We obtain a multiscale structure-preserving decomposition of the image, unlike the more common multiscale structure-decimation of the image signal [17], [33], [34], [19] where coarse scale structure is detected from blurred images and is therefore a smoothed version of the original image structure. In addition to accruing edge displacement error [4], the latter leads to artifacts such as phantom edges [9]. Among other approaches to multiscale segmentation, multiscale blob detection using morphological methods is described in [7], [18] and computation of multiscale medial axis representation is discussed in [24], [2], [16].

2.2 Desired Characteristics and Objectives
The above discussion leads us to the following desired characteristics of multiscale structure detection and segmentation.

2.2.1 Shape and Topological Invariance
The regions should be correctly detected regardless of their shapes and relative placement. For example, an edge point must be detected at its true location, regardless of whether the edge in the vicinity of the point is straight, curved or even contains a corner or a vertex where multiple regions meet.

2.2.2 Photometric Scaling
It should be possible to detect all regions which are in contrast to their surround, regardless of the actual degree of intra-region homogeneity and the value of the contrast. Regions having large contrast may be associated with higher scales.

2.2.3 Spatial Scaling
It should be possible to detect all regions regardless of their shapes and sizes. Higher scales may be associated with larger regions.

2.2.4 Stability and Automatic Scale Selection
Image structures associated with different scales correspond to segmentations that are locally invariant to changes in geometric and contrast parameters. Since the contrasts and sizes of regions contained in an arbitrary image are a priori unknown, they should be identified automatically.

The transform presented in this paper has been motivated by the objective of achieving these desired characteristics. Specifically, the objective is to derive multiscale segmentation of the image and represent it through a hierarchical, tree structure in which the different image segments, their parameters, and their spatial interrelationships are made explicit. The bottom (leaf) nodes of the hierarchy correspond to regions consisting of individual pixels or connected components of constant gray level, and the path

from a leaf to the root node specifies how the leaf regions recursively merge with adjacent regions to form larger regions each of which is homogeneous with respect to its surround and is characterized by its own contrast. Alternate representations of the same image structure and contrast information are also possible, e.g., by ordering regions according to contrast. In this paper, we will not dwell on the different possible data structures that could be used for representation. Rather, we will demonstrate how the transform extracts information about the a priori unknown region geometries, homogeneities and contrasts, and associates this structural information with each image pixel. Any specific image representation may be constructed from such annotated pixel arrays. Since the efficacy of segmentation is ultimately determined by its perceptual validity, performance evaluation will be done through assessment of perceived quality of real image segmentations in addition to quantitative measures computed for synthetic images.

3 THE TRANSFORM AND IMAGE SEGMENTATION

In this section we first discuss how the problems with the previous methods and the desired characteristics of the segmentation motivate an approach such as that underlies the proposed transform (Section 3.1). We then introduce the transform (Section 3.2) and present its properties (Section 3.3) that demonstrate how the transform makes explicit the image structure and facilitates image segmentation, and why the resulting segmentation possesses the desired characteristics listed in Section 2.2. Section 3.4 describes how a given image region appears in the vector field computed by the transform. In Section 3.5, we discuss the estimation of the unknown scale parameters associated with an image, which are required to extract the unknown structures present in different parts of the image. Section 3.6 describes the hierarchical image structure that is extracted as the final result of segmentation. Whenever necessary, we will consider constant-value regions to analytically and qualitatively describe the transform behavior. However, the properties of the transform responsible for its segmentation capability remain valid for images containing more general types of regions, such as those having noise (statistical constancy) and shading (smoothly varying values or higher-order homogeneity) as will be explained in Section 4. Thus, the transform is proposed for use in segmentation of piecewise smooth images.

3.1 Overcoming Limitations of Linear Processing via the Transform

To consider the limitations of previous linear approaches, consider the edge-based detection of regions. In the literature, different models of edge profile (step, ramp, roof) and their validity have been investigated. However, the limitations and impact of the assumptions made about edge geometry have received limited attention. Since any convolution kernel for edge detection must incorporate a template for the expected edge geometry, no linear, convolution based approach can avoid the limitations resulting from the use of geometric models of edge. In the digital case, one could attempt to circumvent the problem by enumerating all possible edge geometries in a neighborhood. But the number of resulting kernels will fast increase with neighborhood size and will be prohibitively large for any reasonable size neighborhood.

The inspiration for the proposed solution comes from physics where microscopic homogeneity of physical properties leads to islands of, say, similar particles or molecules. An island shape is congruent with the space occupied by a set of contiguous, similar particles, whatever the complexity of the boundary! The particles group together and coalesce into regions based on the similarity of their intrinsic properties, regardless of their relative locations. The common property of the particles then characterizes the region they form. As an alternate analogy, the grouping process is like the alignment of microscopic domains over an area of ferromagnetic material. The key process is that of interaction among particles which leads to bindings among similar particles.

The problem of segmentation has similarities to the above physical process. The goal is to find a partition of the image, regardless of the boundary complexity, such that each cell of the partition has a characteristic property, say, homogeneity of gray level. This analogy suggests a formulation of the segmentation process in terms of a suitably defined method of interpoint interaction—one that would group, bottom-up, each set of points of the same property to form a region having a boundary of any complexity. Being a parameter of the grouping process, different acceptable degrees of the presence of the property within a region would yield groupings over different regions, making scale an integral part of structure detection.

In the next section, we introduce a transform which achieves the above grouping.

3.2 The Transform

The transform converts the image I into a vector field **F**. The vector $\mathbf{F}_p$ at an image location **p** is defined as

$$\mathbf{F}_p = \int_{q \neq p} d_s\left(\mathbf{r}_{pq}, \sigma_s(\mathbf{p})\right) d_g\left(\Delta I, \sigma_g(\mathbf{p})\right) \hat{\mathbf{r}}_{pq} d\mathbf{q} \qquad (1)$$

where

$\hat{\mathbf{r}}_{pq}$ = unit vector in the direction from **p** to another image location **q**;

$\sigma_s(\mathbf{p})$ = spatial scale parameter at **p**; denotes the shortest distance to region boundary; all valid $\sigma_s(\mathbf{p})$ values are computed automatically;

$\sigma_g(\mathbf{p})$ = photometric scale parameter at **p**; denotes contrast of region with surround; all valid $\sigma_g(\mathbf{p})$ values are computed automatically;

ΔI = absolute gray level difference between image points under consideration;

$d_s(a, b)$ = a nonnegative, nonincreasing function of $\|\mathbf{a}\|$, not identically 0 for $\|\mathbf{a}\| \leq b$, and 0 for $\|\mathbf{a}\| > b$, and

$d_g(a, b)$ = a nonnegative, nonincreasing, and symmetric function of a, not identically 0 for $a \leq b$, and 0 for $a > b$.

Since $d_g(a, b)$ as defined above cannot be a linear function of a for unrestricted values of a, the transform does not obey the principle of superposition and hence is nonlinear.

3.2.1 Formulation of the Transform

Having stated the definition of the transform, we will now explain the motivation behind this formulation and construct the definition from basic principles.

Consider a region R and a point P inside it. It is desired that the similarity of P to all the other points within R, relative to those outside R, be recognized, regardless of the distance to and curvature of the nearest region boundary. To do this, the transform defines a neighborhood around P which is sufficiently small so that the pixels within the neighborhood which have the most influence on the computation at P are within R. The similarity is estimated by comparing the gray level at P with those of the points in the neighborhood, rather than testing some (position independent) gray level statistics of points within the neighborhood, or comparing the statistics of different sets of pixels near P. An estimate of the local shape of R near P, specifically of the nearest region border, is computed. This is achieved by a vector integration of the results of pairwise comparisons of points at different orientations, instead of computing a scalar, weighted average at P as is done in linear methods. Contiguous points having compatible estimates of local shape of R are grouped together. This results in the detection of an arbitrarily large region of arbitrary shape.

To capture the local region geometry, the transform computes an attraction-force field over the image wherein the force at each point denotes its affinity to the rest of the image. The force vector points in the direction in which the point experiences a net attraction from the points in the rest of the image. For example, a point inside a region would experience a force towards the interior of the region. This force is computed as the resultant of attraction-forces due to all other image points. If $F(\mathbf{p}, \mathbf{q})$ denotes the magnitude of the force vector $\mathbf{F}(\mathbf{p}, \mathbf{q})$ with which a pixel P at location $\mathbf{p}$ is attracted by another pixel Q at location $\mathbf{q}$, then the transform is given by

$$\mathbf{F}(\mathbf{p}, \mathbf{q}) = F(\mathbf{p}, \mathbf{q})\hat{\mathbf{r}}_{\mathbf{pq}},$$

where $\hat{\mathbf{r}}_{\mathbf{pq}}$ denotes the unit vector in the direction from P to Q, i.e.,

$$\hat{\mathbf{r}}_{\mathbf{pq}} = \frac{\mathbf{q} - \mathbf{p}}{\|\mathbf{q} - \mathbf{p}\|}.$$

In the real image plane, an image is transformed into a continuous vector field. The vector $\mathbf{F}_{\mathbf{p}}$ at point P is given by

$$\mathbf{F}_{\mathbf{p}} = \int_{\mathbf{q} \neq \mathbf{p}} F(\mathbf{p}, \mathbf{q})\hat{\mathbf{r}}_{\mathbf{pq}} d\mathbf{q} \qquad (2)$$

where $\mathbf{q}$ can be any image location other than $\mathbf{p}$. In the discrete case,

$$\mathbf{F}_{\mathbf{p}} = \sum_{\mathbf{q} \neq \mathbf{p}} F(\mathbf{pq})\hat{\mathbf{r}}_{\mathbf{pq}}$$

We need to specify what forms the force function $F(\mathbf{p}, \mathbf{q})$ could take. We will do so by identifying the characteristics that any such function must possess, to yield the correct segmentation for a given pair of spatial and photometric scales as well as exhibit appropriate behavior across multiple scales. Specific choices of force functions having these characteristics will then define different instances of the transform.

Since the presence of an edge of a region at any given spatial scale must be determined by its adjoining regions rather than by distant points across other intervening regions, the force F exerted on a given pixel P by another pixel Q should be a nonincreasing function of the distance between P and Q. Further, a pixel should be attracted more to a pixel within its own region than to one in a different region, and the attraction between two pixels should depend only on the magnitude of their gray level difference and not its sign. This is accomplished by making F to be a nonincreasing and symmetric function of the difference between the gray levels of P and Q.

Since both spatial and photometric vicinities are relative to the scales of interest, let us now consider how to integrate the scale information in the computation of F. Consider a region W which exists amidst many other regions in an image (Fig. 1). To detect any structural characteristics of W, the computation of F at a point must not extend to nonlocal parts of W or to other regions, since otherwise the result will depend upon nonlocal structure of W or multiregion structure. To be specific, first consider the situation where a point P is inside W but not on its medial axis. Let $\delta(\mathbf{p})$ denote the distance to the unique border point (or border segment) of W closest to P, and let D denote a disk of radius $r = \delta(\mathbf{p})$ centered at P. If r is increased, for some value $r = \delta'(\mathbf{p}) > \delta(\mathbf{p})$, D's intersection with W's surround will begin to consist of multiple disconnected regions or D will begin to intersect with other regions near W (e.g., regions Y and Z in Fig. 1). Now consider the second situation where the point P is along the medial axis of W. For any such point, $\delta'(\mathbf{p}) = \delta(\mathbf{p})$ since a disk of radius greater than $\delta(\mathbf{p})$ will intersect W's surround in multiple disconnected components, and possibly other regions. In order that $\mathbf{F}_{\mathbf{p}}$ reflects the structure of W surrounding P, it must be ensured that the contribution to F at $\mathbf{p}$ from points at a distance greater than or equal to $\delta'(\mathbf{p})$ is negligible. Such spatial locality of computation is enforced by including a spatial (or geometric) scale parameter σ_s in the definition of F. $\sigma_s(\mathbf{p})$ associates a "cut-off distance" with $\mathbf{p}$ such that the points farther than this distance make negligible contribution to F value at $\mathbf{p}$. In particular, we define the spatial scale parameter $\sigma_s(\mathbf{p})$ at each point $\mathbf{p}$ in some region R as having a value given by $\sigma_s\mathbf{p} = \delta(\mathbf{p}) + \alpha(\delta'(\mathbf{p}) - \delta(\mathbf{p}))$, where $0 < \alpha < 1$. Thus, $\delta(\mathbf{p}) < \sigma_s(\mathbf{p}) < \delta'(\mathbf{p})$ at all points inside R except along the medial axis where $\sigma_s(\mathbf{p}) = \delta(\mathbf{p}) = \delta'(\mathbf{p})$. To achieve the desired monotone dependence of F on the spatial scale as well as on the distance as discussed in the previous paragraph, we make it proportional to a function $d_s(\mathbf{r}_{\mathbf{pq}}, \sigma_s(\mathbf{p}))$, where $d_s(\mathbf{r}_{\mathbf{pq}}, \sigma_s(\mathbf{p}))$ is a nonnegative, nonincreasing function of $\|\mathbf{r}\|$, not identically 0 for $\|\mathbf{r}\| \leq \sigma_s(\mathbf{p})$, and 0 for $\|\mathbf{r}\| > \sigma_s(\mathbf{p})$.
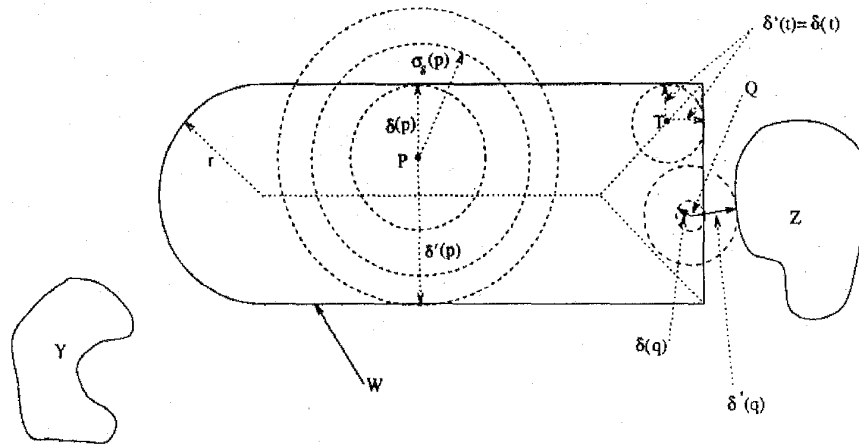
Fig. 1. The medial axis of the region W is the locus of points equidistant from multiple points on W's border, shown here by dotted lines. A point P's distance to the nearest border point is given by $\delta(\mathbf{p})$. A disk D of radius $> \delta$' may intersect with W's surround in more than one connected component (e.g., point P), or it may intersect with other regions as well (e.g., point Q). For points along the medial axis, $\delta' = \delta$ (e.g., point T). The spatial scale parameter σ_s at any point is defined as a value between δ and δ' (radius of the middle circle for point P).

Just as the spatial scale parameter is chosen to ensure that F at **p** depends on spatial structure in the neighborhood of **p**, the photometric (contrast, or gray level) scale parameter at P is determined by the degree of gray level homogeneity and contrast of a specific region of interest containing P. It ensures that F at **p** is determined by points within the region, i.e., the point experiences negligible attraction to another point having gray level difference larger than a cut-off value characteristic of within-region gray level variability. To achieve this, we make F proportional to the function $d_g(\Delta I, \sigma_g(\mathbf{p}))$, where ΔI denotes the absolute gray level difference between points P and Q, and $\sigma_g(\mathbf{p})$ is the photometric scale parameter. $d_g(\Delta I, \sigma_g(\mathbf{p}))$ is a nonnegative, nonincreasing, and symmetric function of ΔI, not identically 0 for $\Delta I \leq \sigma_g(\mathbf{p})$, and 0 for $\Delta I > \sigma_g(\mathbf{p})$.

Thus, $\mathbf{F_p}$ can be written as defined earlier in (1), namely,

$$\mathbf{F_p} = \int_{q \neq p} d_s(\mathbf{r_{pq}}, \sigma_s(\mathbf{p})) d_g(\Delta I, \sigma_g(\mathbf{p})) \hat{\mathbf{r}}_{pq} d\mathbf{q}$$

Two observations follow from the above definition of F. First, the scale parameters at different points within a region R are mutually dependent since they are all determined by the structure of R. In particular, σ_s at a point P inside a region R depends on P's location relative to R's boundary, and all points inside R have the same σ_s value which corresponds to the contrast of R with its surround. For example, in a piecewise constant image, σ_s varies continuously and σ_g is piecewise constant. Second, since the region structure is to be determined in the first place, σ_s and σ_g are a priori unknown to the segmentation algorithm and must be computed at each point. Since a point in general belongs to many different regions at different scales, σ_s and σ_g will have multiple values at a point.

Note that when a homogeneous region (or background) is at least partly enclosed by the image border, the computation of F will be undefined at those image points P whose nearest border point is on image boundary. This is because the computation of $\sigma_s(\mathbf{p})$ involves points at distances $> \delta(\mathbf{p})$ from **p** (Fig. 1) and some such points are outside the image. To resolve this problem, we will treat the entire image as surrounded by a hypothetical, constant-value region whose contrast relative to the given image is infinite. Accordingly, for computational purposes, the points outside the image will be assumed to be accessible but having a gray level of infinity. This would yield the given, finite size image as the largest and least homogeneous region within the hierarchical segmentation of the hypothetical, infinitely large image.

3.3 Properties

In this section, we present some properties of F which collectively describe the relationship between the spatial structure of F and the image structure, and consequently, suggest F as a means of image segmentation. For brevity, the proofs are outlined for some of the properties; the rest can be found in [3].

3.3.1 Null Response

Suppose an image contains a constant-value disk, an arbitrary number of arbitrarily located other regions not intersecting with the disk, and a constant-value background. Then the value of F is zero at the disk center.

PROOF. Let D denote the constant-value disk and P denote the center of D. Let X denote the constant-value background adjacent to D. By definition, $\sigma_s(\mathbf{p})$ has a value such that only points of D and X are contained within a distance $\sigma_s(\mathbf{p})$ from **p**. Thus F at **p** is computed from a neighborhood comprised of the disk D surrounded by an annulus of X, which makes the neighborhood gray level distribution to be isotropic around **p**. Therefore, for each point Q, there is another point Q' such that Q and Q' are located radially symmetrically about P, and Q and Q' exert equal and opposite forces at P. Therefore the net force $\mathbf{F_p}$ at P is zero.   □

3.3.2 Inward Attraction

Let P be a point inside a homogeneous region W. Let D be the disk of radius $\sigma_s(p)$ *centered at P which intersects with W and its border with a homogeneous region X of contrast C. Let M and N be the two points where the borders of D and W intersect (there will be exactly two such points by definition of* σ_s*). Then, for* $\sigma_g < C$*, there exists a point S along W's border within D and between the points M and N such that the direction of* F_p *at P is given by the vector from S to P, i.e.,* $F_p = k\hat{r}_{sp}$*, where k is a positive constant and* $\hat{r}_{sp}$ *denotes the unit vector from S to P. Thus* F_p *points inward.*

PROOF. Let L denote the region of intersection between D and X (Fig. 2). If we construct another region L' such that L and L' are symmetric about P, with the segment M'N' being symmetric to MN, then the force at P due to the neighborhood (D-L-L') will be 0 by symmetry (see Property 3.3.1). Therefore, the force F_p at p is due to the points in the regions $L \cup L'$ and is given by

$$F_p = F_{pL} + F_{pL'} \tag{3}$$

where F_{pL} and $F_{pL'}$ denote the forces at p due to regions L and L', respectively. Now for all $q_1 \in L$, it is given that $\Delta I(p, q_1) > \sigma_g$. Therefore, $d_g(\Delta I(p, q_1), \sigma_g) = 0$ from the definition of d_g. For all $q_2 \in L'$, $\Delta I(p, q_2) = 0 < \sigma_g$. Since d_g is a nonincreasing function of ΔI not identically 0, and $\Delta(p, q_2 \in L') = 0 < \sigma_g < \Delta I(p, q_1 \in L)$, we have

$$d_g\big(\Delta I(p, q_2), \sigma_g\big) > 0 = d_g\big(\Delta I(p, q_1), \sigma_g\big). \tag{4}$$

Thus $F_{pL} = 0$, and consequently, $F_{pL} = F_{pL'}$. Since d_g is not identically zero and nonnegative, it follows from (1), (3), and (4) that $F_{pL'} > 0$. Therefore, $F_p = F_{pL'} > 0$. Since $F_{pL'}$ is the net force on P due to all points $q_2 \in L'$, the direction of $F_{pL'}$ must be given by the unit vector $\hat{r}_{ps''}$ for some point S'' $\in$ L'. If S' denotes a point of intersection of the line joining P and S'' with the segment M'N', then the direction of Fp is given by the unit vector $\hat{r}_{ps'}$. Therefore, $F_p = k\hat{r}_{ps'}$, where k is a constant. But for each point s' on segment M'N', there is another point s on segment MN where the line PS intersects the segment MN, and $\hat{r}_{ps'} = \hat{r}_{sp}$. Therefore, $F_p = k\hat{r}_{sp}$. □
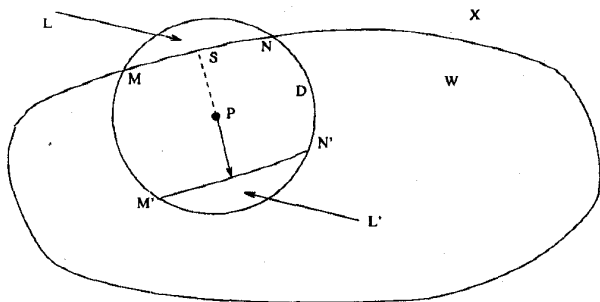
Fig. 2. The force at a point P inside a homogeneous region points inward (Property 3.3.2).

3.3.3 Divergence

Consider a pair of points P and P' inside and outside, respectively, a homogeneous region W, and infinitesimally close to W's boundary. Then $F_p X F_{p'} = 0$, i.e., the force vector undergoes direction reversal across the boundary, regardless of the shape of the boundary.

PROOF. Let P' belong to region X (Fig. 3). Consider the line *l* through P and P', and the forces at P and P' due to points along *l* and within a distance $\sigma_s(p)$ (= $\sigma_s(p')$). Let $F_{IW}(P)$ denote the total attraction force at P along *l* due to and towards the points inside W. Let $F_{IX}(P)$ denote the total attraction force at P along *l* due to and towards the points inside X. Analogously, let $F_{IX}(P')$ and $F_{IW}(P')$ denote the total attraction forces on P' along *l* due to and towards the points inside X and W, respectively. Then the total attraction force $F_{IW}(P)$ at P towards W is given by $F_{IW}(P) = (F_{IW}(P) - F_{IX}(P))\hat{w}$ where $\hat{w}$ is the unit vector along *l* and towards W. Similarly, the total attraction force $F_{IX}(P')$ at P' towards X is given by $F_{IX}(P') = (F_{IX}(P') - F_{IW}(P'))\hat{x}$ where $\hat{x}$ is the unit vector along *l* and towards X. Since each of W and X is a constant-gray-level region, we have $F_{IW}(P) = (F_{IX}(P')$ and further since F is an even function of gray level difference, $F_{IX}(P) = (F_{IW}(P')$. Therefore, $F_{IW}(P) = -F_{IX}(P')$. Now the total force $F_W(P)$ at P towards W is given by $F_W(P) = \int_0^\pi F_{IW}(P)d\theta$ and the total force $F_X(P')$ at P' towards X is given by $F_X(P') = \int_0^\pi F_{IX}(P')d\theta$. From the previous three equations, we see that $F_W(P) = -F_X(P')$, which means that there is a directional discontinuity of magnitude π between the force vectors at P and P'. □

3.3.4 Orthogonality

Consider two points P and Q inside a homogeneous region W such that the unit vector $\hat{r}_{pq}$ is orthogonal to the region boundary at the point S where the line PQ intersects region

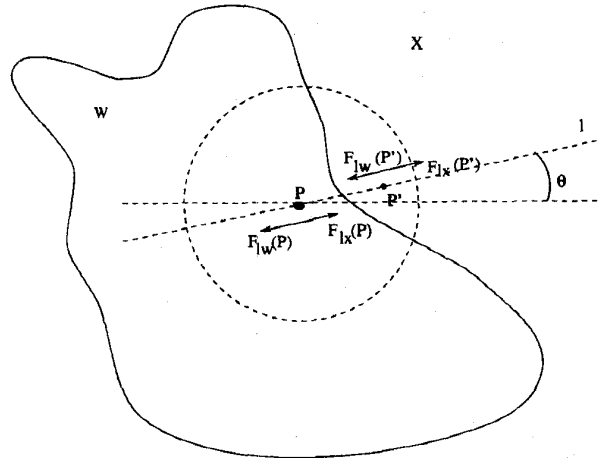Fig. 3. Divergence. If a point P in region W and another point P' in region X are infinitesimally far from each other, and hence also from the boundary between W and X, then the force vectors at P and P' are equal and opposite (Property 3.3.3).

boundary and is directed into W. If the intercept of W's boundary with the disk D of radius $r = \sigma_s(\mathbf{p})$ centered at P is symmetric about the line PQ, then the direction of $\mathbf{F}_p$ is given by $\hat{\mathbf{r}}_{pq}$, i.e., $\mathbf{F}_p = k\hat{\mathbf{r}}_{pq}$ for some positive constant k, i.e., $\mathbf{F}_p$ is orthogonal to the boundary.

PROOF. By definition of σ_s, the force $\mathbf{F}_p$ at P does not depend on points outside D. Since the boundary of W within D is symmetric about P, line PQ divides the region of intersection of W and D into two symmetric halves. For each point M in one half, there exists another point M' in the other half such that the components of the force at P due to M and M' and orthogonal to PQ are equal and opposite. Thus, the net force at P is along PQ and, from Property 3.3.2, in the direction $\hat{\mathbf{r}}_{pq}$. Since PQ is orthogonal to the region boundary, so is $\mathbf{F}_p$.   □

3.3.5 Medial Axis

At all points P along the medial axis of a region, $\mathbf{F}_p = 0$.

PROOF. This follows directly from the definition of σ_s. For any point P along the medial axis, $\delta(\mathbf{p}) = \delta'(p)$. Therefore, $\sigma_s(\mathbf{p}) = \delta(\mathbf{p})$. Since the force on P due to points farther than $\sigma_s(\mathbf{p})$ is negligible (from the definition of $\sigma_s(\mathbf{p})$ and since all points within the distance $\sigma_s(\mathbf{p})$ form a homogeneous region, it follows from Property 3.3.1 that $\mathbf{F}_p = 0$. Note that here the detected axis is medial in the sense of [6]. Other related shape axes are based on local symmetery [35] or inertia [36].   □

3.3.6 Convergence

At points near and on either side of the medial axis of a region R, F is directed toward the medial axis. That is, F points away from the closest border segment in the sense described in Property 3.3.2.

PROOF. Consider a point Q near the medial axis. Then $\delta'(q) > \delta(q)$ and $\sigma_s(q) > \delta(q)$. From Property 3.3.2, therefore, $\mathbf{F}_q$ points inward (away from the region border). As Q approaches a medial axis segment from either side, $\delta'(q)$ approaches $\delta(q)$ and therefore, $\sigma_s(q)$ approaches $\delta(q)$. Thus, the direction of F points toward the medial axis which becomes the locus of force convergence.   □

3.3.7 Smoothness

If $d_g(\Delta I, \sigma_g)$ is a continuous function of ΔI, and $d_s(\mathbf{r}, \sigma_s)$ is a continuous function of $\mathbf{r}$ and σ_s, then F is a spatially continuous function at all nonboundary points of a region.

PROOF. Consider a point P inside a completely homogenous region R, and another point T inside R and arbitrarily close to P. Then,

$$\mathbf{F}_p = \int_{q \neq p} d_s(\mathbf{r}_{pq}, \sigma_s(\mathbf{p})) d_g(\Delta I(\mathbf{p}, \mathbf{t}), \sigma_g(\mathbf{p})) \hat{\mathbf{r}}_{pq} d_q$$

and

$$\mathbf{F}_t = \int_{q \neq t} d_s(\mathbf{r}_{tq}, \sigma_s(\mathbf{t})) d_g(\Delta I(\mathbf{t}, \mathbf{q}), \sigma_g(\mathbf{t})) \hat{\mathbf{r}}_{tq} d_q$$

Now for any third point Q, $\Delta I(\mathbf{p}, \mathbf{q}) = \Delta I(\mathbf{t}, \mathbf{q})$ since P and T have the same gray level. Further, since P and T are arbitrarily close, $(\mathbf{r}_{tq} - \mathbf{r}_{pq})$ is arbitrarily small. Also, $\sigma_s(\mathbf{t}) - \sigma_s(\mathbf{p})$ is arbitrarily small from the definition of σ_s. Therefore, for any given choice of $\sigma_g(\mathbf{p})$ ($= \sigma_g(\mathbf{t})$), $(\mathbf{F}_p - \mathbf{F}_t)$ is arbitrarily small. Consequently, $\mathbf{F}_p$ is continuous everywhere within a region.   □

3.3.8 Closure

For any piecewise constant image, the contours along which F exhibits divergence are closed.

PROOF. From Property 3.3.3, a region edge appears as a directional discontinuity of π in F. Further, from Property 3.3.7, F varies smoothly on either side of the edge. Therefore, for each point of directional discontinuity, there is another point in its vicinity which is also a point of directional discontinuity. Since the boundary of any region in a piecewise homogeneous image is a closed contour, the points of directional discontinuity form closed contours with smoothly varying F on each side.   □

3.4 A Region's Signatures

The above properties of the transform collectively suggest the following structure of F associated with an image region R, or the signatures of R in the F-field. When a σ_g value corresponding to R's contrast is used to compute F, it leads to much attraction between pixels within R and little attraction between pixels across R's boundary. Further, if the σ_s value at each point is chosen corresponding to R, then R is characterized by a spatially smooth and inward force flow, where the force lines emanate from the region border and converge at R's medial axis given by $\mathbf{F} = 0$. The direction of F undergoes a divergent discontinuity of magnitude π across the entire region border except at those border points where $\mathbf{F} = 0$, i.e., border points which also lie on the medial axis such as corner points. Thus F's magnitude along a region's border varies but is 0 at corners. Fig. 4 illustrates the signatures of rectangular regions.

A test of F's validity can be performed by considering the nature of $\mathbf{F}_p$ at a point P inside an arbitrary shaped region R as R shrinks in size uniformly and vanishes. Let us first consider the case where P belongs to the medial axis of R. Here, $\delta(\mathbf{p}) = \sigma_s(\mathbf{p}) = \delta'(\mathbf{p})$. Since $\mathbf{F}_p$ is computed over a constant-value disk of radius $\sigma_s(\mathbf{p})$, $\mathbf{F}_p = 0$. Now consider a point P off the medial axis of R. Then, $\delta(\mathbf{p}) < \sigma_s(\mathbf{p}) < \delta'(\mathbf{p})$. However, as R continues to shrink, the area of the region L' (= area of region L) in Fig. 2 approaches 0. Since the net force at P is that due to L' (see the proof of Property 3.3.2),

$$\lim_{Area(R) \to 0} \mathbf{F}_p = 0. \tag{5}$$

Thus the influence on the F field caused by an image region vanishes as the area of the region vanishes, as is to be expected.

3.5 Estimation of Scale Parameters

Recall that image regions are in general recursively embedded,

each standing in contrast with its surround and characterized by its own gray level homogeneity. An image point is associated with multiple regions, and therefore with multiple degrees of homogeneity and contrast scales. Since in the proposed transform an increasing value of σ_g corresponds to increasingly nonuniform regions, σ_g comprises one index into the structural hierarchy. Further, since each region has its own shape and size, and the spatial scale parameter at a point depends on its location with respect to region boundary, each image point is also associated with a number of spatial scales (σ_s values). Therefore, each image point is associated with a number of (σ_s, σ_g) pairs, corresponding to the different regions that it belongs to. In real images, a point is contained in only a small number of regions, and is therefore characterized by a small number of scale pairs. F makes explicit at each image location regions corresponding to all scales present at that location. A particular selection of regions across the image corresponds to a specific cutset of the segmentation hierarchy.

The signatures of a region in the F field, as described in the previous subsection, are obtained assuming that F at each point is computed using the appropriate pair of (σ_s, σ_g) values. We will now explain how these unknown scales can be estimated to yield the multiscale image structure. For estimation of these values, we will treat σ_s and σ_g as variables and identify their values that correspond to image regions. Suppose that for all images of interest, $\left(\sigma_s\right)_{min} \leq \sigma_s \leq \left(\sigma_s\right)_{max}$ and $\left(\sigma_g\right)_{min} \leq \sigma_g \leq \left(\sigma_g\right)_{max}$. That is, the ranges of sizes and contrasts to be encountered in images have known bounds. If no specific information is available for images to be processed, the image size and the maximum gray level can be used as $\left(\sigma_s\right)_{max}$ and $\left(\sigma_g\right)_{max}$, respectively, while $\left(\sigma_s\right)_{min} = \left(\sigma_g\right)_{min} = 1$ can be used as the minimum size and contrast a region could have. Suppose the transform is used to compute the force at each point for $\left(\sigma_s\right)_{min} \leq \sigma_s \leq \left(\sigma_s\right)_{max}$, and $\left(\sigma_g\right)_{min} \leq \sigma_g \leq \left(\sigma_g\right)_{max}$. Suppose that at a point P, the pair of values $\left(\sigma_{s1}, \sigma_{g1}\right)$ and $\left(\sigma_{s2}, \sigma_{g2}\right)$ correspond to two regions R1 and R2 at two adjacent scales which contain P, with R2 containing (>) R1, and $\left(\sigma_{g2} > \sigma_{g1}\right)$. That is, there is no other region R3 such that R1 < R3 < R2. By our definition of scale given in Section 1, R1 and R2 represent two adjacent natural scales occurring in the image at P. Let V_1 and V_2 denote, respectively, the values of F corresponding to $\left(\sigma_{s1}, \sigma_{g1}\right)$ and $\left(\sigma_{s2}, \sigma_{g2}\right)$. As the value of the variable σ_s is increased from σ_{s1}, the image area responsible for a nonzero value of F at P (area L in Fig. 2) will increase, resulting in an increase in the magnitude of F. Now from Property 3.3.2, as the value of σ_s at a point P increases beyond $\sigma_s(p)$, the shape of L changes gradually and s moves along the subregion boundary, resulting in a gradual change in $\hat{r}_{sp}$. A similar gradual change in F will also be associated with an increase in σ_g. Thus, F will slowly deviate from V_1 as (σ_s, σ_g) values increase. However, for a sufficiently large value of σ_s, i.e., $\sigma_s > \delta'(p)$, the value of

F will begin to depend on multiple disconnected components of R1's surround which are subregions of R2. Since the direction $\hat{r}_{sp}$ is in general different for R1 and R2, V_1 and V_2 are in general different. Similarly, for $\sigma_{g1} << \sigma_g << \sigma_{g2}$, the subregions of R2 will make significant contribution to F value at P. Consequently, F will change with (σ_s, σ_g). As (σ_s, σ_g) approach $\left(\sigma_{s2}, \sigma_{g2}\right)$, F will assume the relatively stable value of V_2. Therefore, in the $\sigma_s\sigma_g$-space, the locations where F is stable will be scattered, associated with structures at different pairs of scale values. Somewhere between each pair of nearby locations of stable points in the $\sigma_s\sigma_g$-space corresponding to R1 and R2, F will make a sharper transition from the value V_1 to V_2. Fig. 5 illustrates the different stable values of F for a simple synthetic image. By traversing the $\sigma_s\sigma_g$-space, computing F using all σ_s, σ_g values at all image points, and identifying those parts in the $\sigma_s\sigma_g$-space where F is locally stable (has locally minimal variation), we can determine all the scales associated with P. The scale values at all image points can be estimated jointly, because together the values comprise the signatures of regions as explained in the previous subsection. Such scale estimation is robust for two reasons. First, at each point only qualitative changes in F are detected. Second, the qualitative changes at different points are analyzed jointly to detect the spatial signatures of a region. This further suppresses any noise in F, which is already low because of the large neighborhoods used in the computation of F.

The signatures of a region in the F field, as described in the previous subsection, are obtained assuming that F at each point is computed using the appropriate pair of (σ_s, σ_g) values. We will now explain how these unknown scales can be estimated to yield the multiscale image structure. For estimation of these values, we will treat σ_s and σ_g as variables and identify their values that correspond to image regions. Suppose that for all images of interest, $\left(\sigma_s\right)_{min} \leq \sigma_s \leq \left(\sigma_s\right)_{max}$ and $\left(\sigma_g\right)_{min} \leq \sigma_g \leq \left(\sigma_g\right)_{max}$. That is, the ranges of sizes and contrasts to be encountered in images have known bounds. If no specific information is available for images to be processed, the image size and the maximum gray level can be used as $\left(\sigma_s\right)_{max}$ and $\left(\sigma_g\right)_{max}$, respectively, while $\left(\sigma_s\right)_{min} = \left(\sigma_g\right)_{min} = 1$ can be used as the minimum size and contrast a region could have. Suppose the transform is used to compute the force at each point for $\left(\sigma_s\right)_{min} \leq \sigma_s \leq \left(\sigma_s\right)_{max}$, and $\left(\sigma_g\right)_{min} \leq \sigma_g \leq \left(\sigma_g\right)_{max}$. Suppose that at a point P, the pair of values $\left(\sigma_{s1}, \sigma_{g1}\right)$ and $\left(\sigma_{s2}, \sigma_{g2}\right)$ correspond to two regions R1 and R2 at two adjacent scales which contain P, with R2 containing (>) R1, and $\left(\sigma_{g2} > \sigma_{g1}\right)$. That is, there is no other region R3 such that R1 < R3 < R2. By our definition of scale given in Section 1, R1 and R2 represent two adjacent natural scales occurring in the image at P. Let V_1 and V_2 denote, respectively, the values of F corresponding to $\left(\sigma_{s1}, \sigma_{g1}\right)$ and $\left(\sigma_{s2}, \sigma_{g2}\right)$.
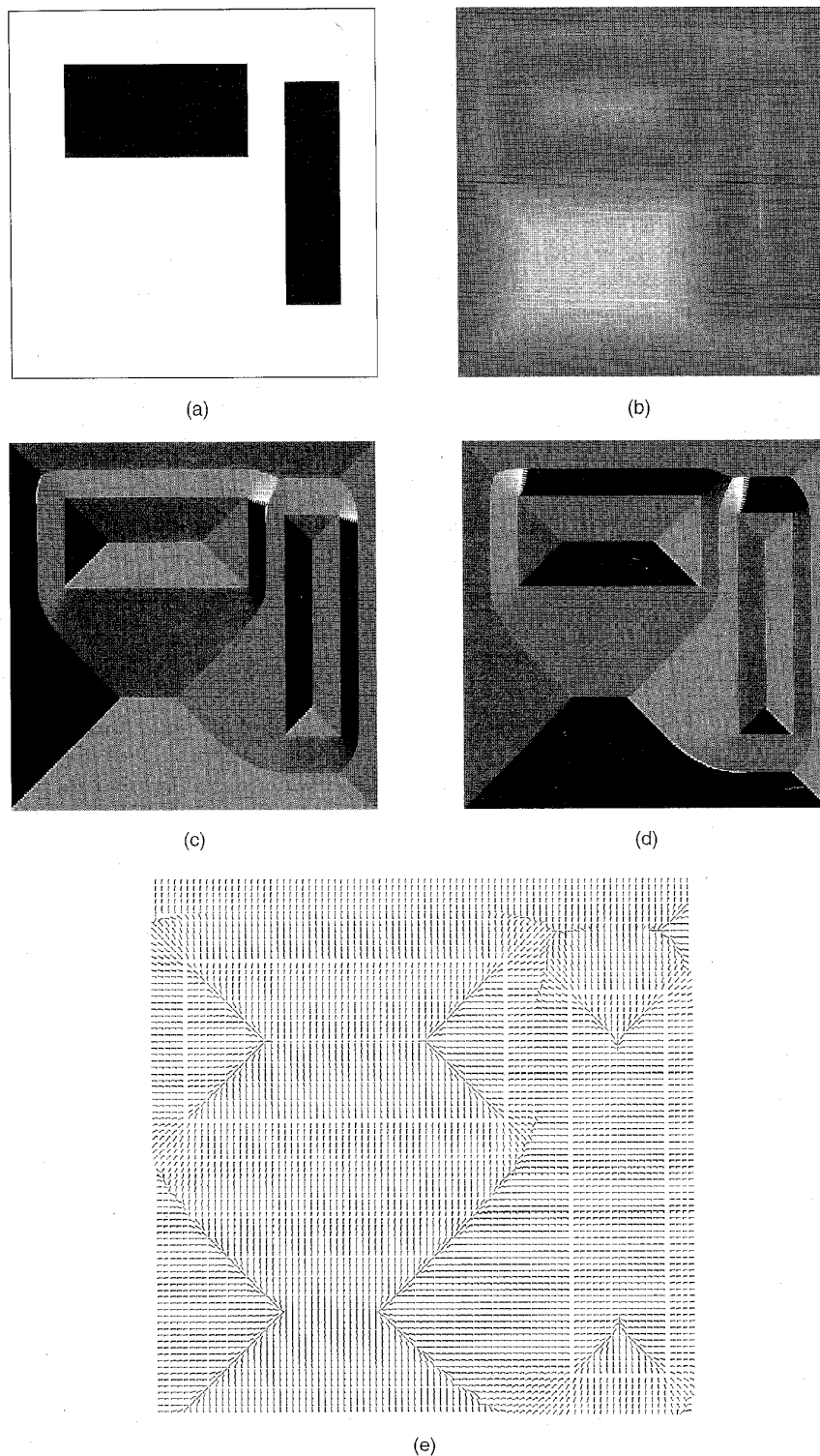
(a)

(b)

(c)

(d)

(e)

Fig. 4. Demonstration of the properties of the transform and the region signatures through a simple synthetic image. (a) Two rectangles having a constant intensity of 1, on a background of intensity 255. (b) Intensity coded image of σ_s values chosen to correspond to the rectangular regions. Image brightness is proportional to σ_s value. (c) Force vectors obtained after the transform has been applied using the σ_s values shown in (b), and $\sigma_g = 20$. The vector directions are intensity coded so that the brightness is proportional to the clockwise angle of the force vector from the positive x-axis. Some intensity discontinuities are artifacts of intensity based (linear) coding of the cyclic direction values. (d) Same as (c), but with vector directions rotated 90° clockwise and then intensity coded. This shifts the bright-dark artifacts resulting from the intensity coding, allowing one more easily to verify the field in (c) near areas where such artifacts occur. (e) The vectors in (c-d) shown as line segments. The length of the line segment represents the vector magnitude, and the tail of the vector is indicated by a small square.
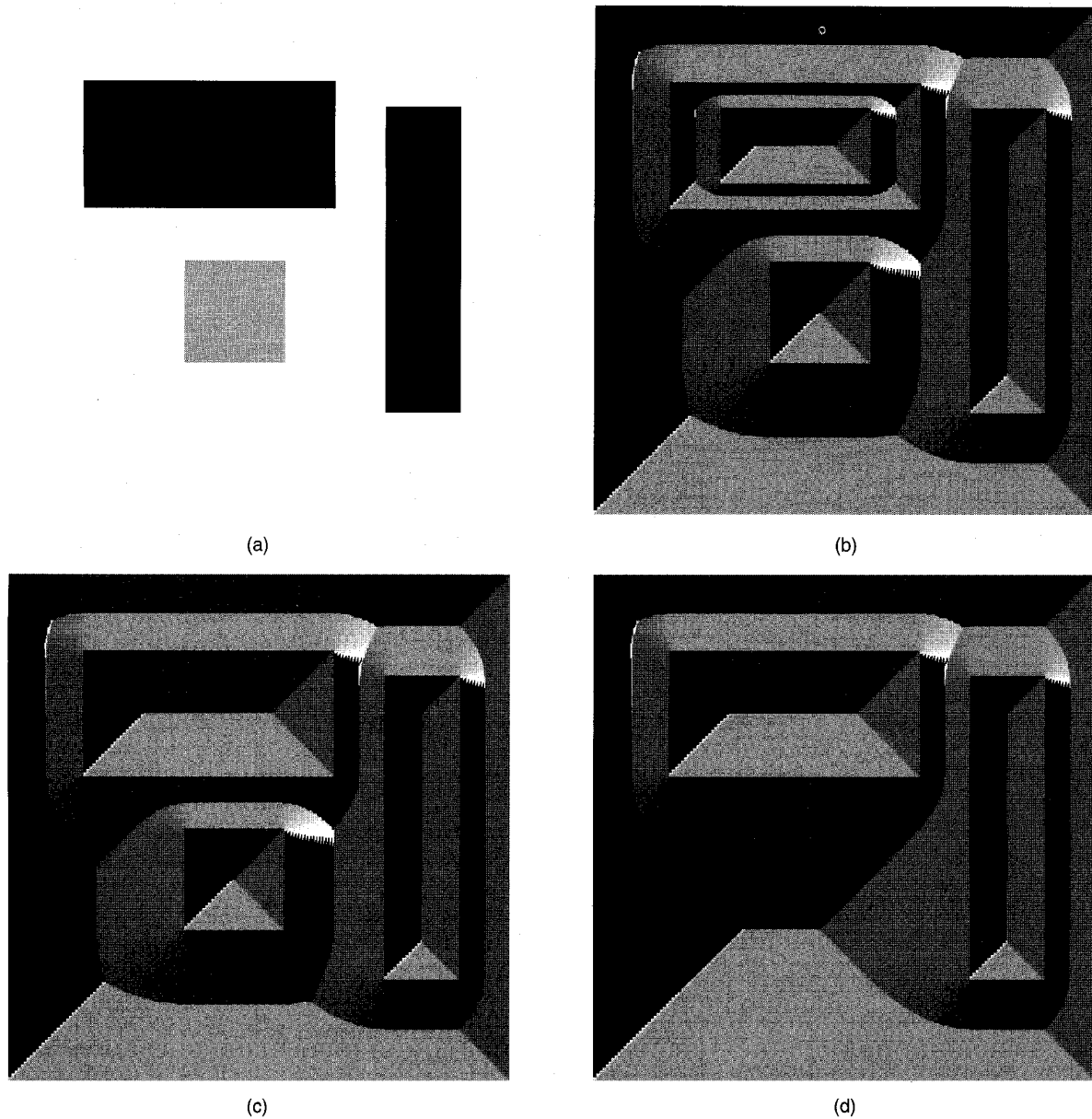
Fig. 5. Demonstration of the multiscale capability of the transform. (a) Same as Fig. 4a, except that two new, lower contrast regions have been added. The new small rectangle has a gray level of 30, and the new square has a gray level of 200. The three contrasts present are 29, 55, and 254. (b) Intensity coded vector direction image using $\sigma_g = 10$. Because 10 is less than all contrasts, all four regions are detected. (c) Analogous to (b) using $\sigma_g = 40$. Because 40 > 29, the new rectangle is lost. (d) Analogous to (b) using $\sigma_g = 60$. Because 60 > 29, 55, both new regions are lost, leaving a result identical to Fig. 4c.

As the value of the variable σ_s is increased from σ_{s1}, the image area responsible for a nonzero value of F at P (area L in Fig. 2) will increase, resulting in an increase in the magnitude of F. Now from Property 3.3.2, as the value of σ_s at a point P increases beyond $\sigma_s(p)$, the shape of L changes gradually and s moves along the subregion boundary, resulting in a gradual change in $\hat{r}_{sp}$. A similar gradual change in F will also be associated with an increase in σ_g. Thus, F will slowly deviate from V_1 as (σ_s, σ_g) values increase. However, for a sufficiently large value of σ_s, i.e., $\sigma_s > \delta'(p)$, the value of F will begin to depend on multiple discon-

nected components of R1's surround which are subregions of R2. Since the direction $\hat{r}_{sp}$ is in general different for R1 and R2, V_1 and V_2 are in general different. Similarly, for $\sigma_{g1} \ll \sigma_g \ll \sigma_{g2}$, the subregions of R2 will make significant contribution to F value at P. Consequently, F will change with (σ_s, σ_g). As (σ_s, σ_g) approach $\left(\sigma_{s2}, \sigma_{g2}\right)$, F will assume the relatively stable value of V_2. Therefore, in the $\sigma_s\sigma_g$-space, the locations where F is stable will be scattered, associated with structures at different pairs of scale values. Somewhere between each pair of nearby locations of stable

points in the $\sigma_s \sigma_g$-space corresponding to R1 and R2, F will make a sharper transition from the value V_1 to V_2. Fig. 5 illustrates the different stable values of F for a simple synthetic image. By traversing the $\sigma_s \sigma_g$-space, computing F using all σ_s, σ_g values at all image points, and identifying those parts in the $\sigma_s \sigma_g$-space where F is locally stable (has locally minimal variation), we can determine all the scales associated with P. The scale values at all image points can be estimated jointly, because together the values comprise the signatures of regions as explained in the previous subsection. Such scale estimation is robust for two reasons. First, at each point only qualitative changes in F are detected. Second, the qualitative changes at different points are analyzed jointly to detect the spatial signatures of a region. This further suppresses any noise in F, which is already low because of the large neighborhoods used in the computation of F.

The point pattern in $\sigma_s \sigma_g$-space defined by the locations of the actual scale values corresponding to any image point is unique for the point and the image, and represents the a priori unknown multiscale structure determined by the algorithm at the point. This structure is restricted for common images because an image point does not have multiple contrasts associated with the same spatial scale although it may be contained in multiple regions of the same contrast. Therefore, it would suffice to perform a linear sweep of the $\sigma_s \sigma_g$-space in the σ_g direction, while identifying all (if any) σ_s values which yield locally stable F for each σ_g.

3.6 Region Detection

For each region in the image, occuring at any scale, the automatic scale estimation process computes at each pixel in and around the region a σ_s-σ_g pair of values which correspond to the region's shape (σ_s) and contrast (σ_g). When the F field is computed using these σ_s-σ_g values, the field contains the signatures of the region. The detection of regions would thus require partitioning of the F field such that each cell has the signatures of a region.

A simple approach to finding candidate regions is to locate contours of force divergence. This is easy and the result robust since the directional discontinuity across such contours is known to be π. All characteristics (listed in Section 3.4) comprising the region's signature may then be matched jointly with the local F-field around the candidate regions to test the region hypothesis. The details of such hypothesis formation and testing are outside the scope of this paper. The results of some experiments that demonstrate the detection performance are given in Section 5.

3.7 Segmentation Hierarchy

The properties of the transform and the capability of automatic estimation of the scale parameters discussed in the previous subsections allow the construction of a hierarchical representation of segmentation. The computed scale values for nearby pixels are mutually compatible in that for all points within a region, σ_s values are continuous and σ_g values are constant.

The regions of a single gray level form the smallest regions. These may be defined as the leaf nodes of the hierar-

chy. Homogeneous regions having larger sizes or gray level variations are used to define higher levels. For example, the hierarchy may be based on spatial containment relationship; thus, the regions containing a particular leaf region are arranged in increasing order of size to define the path from the leaf node to the root. Nearby pixels merge into increasingly large regions as the path to the root is traversed. The subtree below any node in the hierarchy is unrelated to any other disjoint subtree, i.e., the structure, path lengths and σ_s-σ_g values associated with the nodes are unrelated across the subtrees. They reflect the a priori unknown spatial structure within an image. In this sense, the hierarchy define a recursive partition of the image into arbitrarily shaped regions, analogous to the irregular pyramid representation of [23].

Specific algorithms to compute the scale parameters and to obtain different hierarchical representations are not within the scope of this paper; these will be reported in subsequent publications.

4 PERFORMANCE ANALYSIS

Let us first review the overall performance of the transform with respect to the desired characteristics (Sections 2.2.1–2.2.4) listed in Section 2.2. Characteristic 2.2.1, invariance to local edge geometry and topology, serves as a key motivation for proposing the transform, and is central to its design. The discussion in Section 3 makes it clear how this characteristic is possessed by F. For example, the capability of multiscale segmentation holds even if more than two regions share a border point since the properties of the transform leading to region signatures are not affected by shape and adjacency characteristics of the regions. With regard to desired characteristic 2.2.2, the scale parameter σ_g provides a mechanism to accomplish contrast scaling. As σ_g increases, adjacent regions may merge. This is because the attraction of a point in one region from another point across region boundary may increase sufficiently so that the directional discontinuity in F responsible for the edge may vanish. Thus changing σ_g achieves the same result as contrast based split-and-merge of regions [14], [18], and therefore, the desired contrast scaling. Analogously, for any given σ_g, scale parameter σ_s helps achieve geometric scaling (desired characteristic 2.2.3). Larger σ_s values at a point correspond to more global structures having a given contrast σ_g, which results in the capability to detect different spatial scales. The desired characteristic 2.2.4 is of course met since scales can be automatically estimated as explained in Section 3.5.

In the rest of this section, we will divide the performance of the transform into two types. The first type is concerned with the capabilities of the transform to detect off-axis signatures of a region, specifically those represented by Properties 2, 3, 7, and 8. The second type consists of region signatures related to the medial axis, represented by Properties 5 and 6. Properties 1 and 4 do not directly contribute to the region signatures. Sections 4.1 and 4.2 examine the type 1 and type 2 performance, respectively.

In discussing each type of performance, we consider two types of deviations from the image model used in deriving

the transform's properties. First, we consider the case where image regions are characterized by a smooth variation of gray level instead of a constant gray level. Second, we consier the effects of additive noise. These deviations better characterize real images as stated earlier. We investigate the effect these deviations have on the region signatures, i.e., on the different properties. We then examine the impact of different choices of the function F, i.e., the functions d_s and d_g, on the transform's performance.

No detailed proofs are given for these claims as doing so will require the use of specific models of the deviations and further analysis which are beyond the scope of this paper. Results of further, experimental evaluation of the performance are presented in Section 5.

4.1 Off-Axis Signatures

In this section, we examine the first type of performance of the transform, namely, the impact of lack of piecewise constancy and noise on inward attraction, divergence, smoothness, and closure properties.

4.1.1 Deviation from Piecewise Constancy

In the discussion so far, we have assumed that the regions have a constant gray level. We will now explain how the segmentation performance of the transform extends to regions having other types of smooth variations. If the region does not have a constant intensity then the force at a point P in the region will include additional components due to differential rates of change of intensities in different directions away from P (unlike the case for Property 3.3.1). For the simple case of an intensity ramp, the changes in intensity around P are antisymmetric. Since force depends on the absolute intensity difference, P still experiences equal and opposite forces from radially symmteric locations within the region resulting in zero net force from these locations.

Regions in real images often contain shading which is more complex spatial variation of intensity than represented by the ramp considered above, e.g., given by a polynomial in image coordinates x and y. Consider a point P within such a region R and another point Q in R within a neighborhood of radius $\sigma_s(p)$ centered at P. Then, $d_g(\Delta I(p, q))$ will vary for different points Q, unlike was the case for constant-value regions. This variation will in general be nonlinear, partly due to the nonlinear variation in $(\Delta I(p, q)$. Now consider another point T within the neighborhood but not within R. If the range of $\Delta I(p, q)$ values is sufficiently small compared to $\Delta I(p, t)$ for all choices of Q and T, then many of the properties of the transform may still hold. The proofs of Properties 3.3.2, 3.3.3, 3.3.7, and 3.3.8 given earlier suggest that the boundaries of the regions may still be detected as before. Accordingly, at a region boundary, there will still be directional discontinuities because of the large gray level discontinuity; and in the process of finding the scale parameter values for the region, stable response of F will be found for the same values of σ_s and σ_g as if the region were homogeneous. Any directional discontinuities found at locations other than the region border, due to nonlinear variation in gray level, will not persist if σ_g is varied. Since the gray levels vary smoothly within the region, Property 3.3.7 suggests that F will still be continuous within the region although it will exhibit differences in F

values from the piecewise constant case. Therefore, regions with shading but in contrast with the surround should still have signatures similar to those for the piecewise constant case. Verification of the above extrapolation of the properties and the exact restatement of these and other properties for the general case would require exact models of within-region intensity variation, and will be omitted here.

4.1.2 Intensity Noise

We will now consider sensitivity to noise in intensity values. First, suppose that the regions have constant values but contain independently distributed, zero-mean, additive noise having a distribution which is symmetric with respect to the mean. Consider a point P and any other point Q within a neighborhood of radius $\sigma_s(p)$ around P. Then, $d_g(\Delta I(p, q))$ will have the same mean as for the case when the noise is absent. Since $d_g(\Delta I(p, q))$ is a symmetric function of $\Delta I(p, q)$, the expected value of $F(p, q)$ will remain unchanged compared to the case without noise. Therefore, the expected value of F_p due to all points q in the image is the same with or without noise. Now suppose that the regions exhibit ramp-like intensity variation which is contaminated by independent, zero-mean, additive noise. Again, because of gray level antisymmetry about P, and the symmetry of $d_g(\Delta I)$ with respect to ΔI, the region boundary will remain unchanged assuming the region contrast with the surround is high compared to within region variation.

For shaded regions, the noise effects will be anisotropic because the region intensities are asymmetric. For a given point P, consider two other points Q and T within a neighborhood of radius $\sigma_s(p)$ around P. Q is within the region R but T is across R's border. If the range of $\Delta I(p, q)$ values is sufficiently small compared to the range of $\Delta I(p, t)$ values, then resulting F will have limited differences relative to the noiseless case. That is, boundaries of regions with shading but in contrast with the surround will still coincide with direction discontinuities in F and will therefore still be detected. However, as for noiseless shaded regions, exact analysis is necessary to obtain the true characterization of the region signature and its dependence on noise which we will again omit in this paper.

4.1.3 Choices of F

While defining the transform (Section 3), we stated that $d_s(r_{pq}, \sigma_s(p))$ should be a nonincreasing function of the magnitude of r_{pq}, and $d_g(\Delta I, \sigma_g(p))$ should be a nonincreasing and symmetric function of ΔI. A variety of such functions could be used including pulse (box-car), Gaussian, exponential, and linear functions. For example, we may use a Gaussian for d_s as well as d_g, having standard deviations of $k_s\sigma_s$ and $k_g\sigma_g$, resepectively, where k_s and k_g are normalization constants. The choice of Gaussian for d_s and d_g results in optimal localization properties in both spatial and transform domains, in addition to others such as separability in computation. Then, the transform at image location p is given by

$$F_p = \int_{q \neq p} e^{-\frac{\|r_{pq}\|^2}{2(k_s\sigma_s)^2}} e^{-\frac{\Delta I^2(p,q)}{2(k_g\sigma_g)^2}} \hat{r}_{pq} dq \qquad (6)$$

The properties of the transform given in Section 3 hold for any choices of such functions. Although the exact values of the force vectors and the computational speeds depend on specific choices of the functions, the region signatures and segmentation hierarchy remain unchanged. We have verified empirically that this in fact is the case for the four choices of box-car, Gaussian, exponential, and linear functions (see Fig. 7 and [2]).

4.2 Axial Signatures

This section discusses the impact of the two kinds of deviations from piecewise constancy and the choices of F on the medial axis related signatures.

There are two ways in which the transform yields multiscale description of region shape. First, of course, is through the detection of region boundaries which may be used to estimate the medial axis following its definition, or using existing algorithms [6], [30]. The reliability of the region axis detected by this method directly depends on the corresponding reliability of detected region boundaries which we have discussed above. The second, more direct way in which the transform extracts multiscale region shape information is by making explicit the location of the medial axis in the field signatures of the region. We will now consider the performance of this method of medial axis detection with respect to shading and noise. As stated in Property 3.3.5, for constant-value regions the $F = 0$ curve within the region where F directions converge represents the medial axis. Now in the vicinity of locations where $F = 0$, the magnitude of F will change at a rate determined by how fast the neighborhood of radius σ_s begins to have significant intersection with an adjacent region, and thus on the local region shape. The steeper the F variation, the more accurate will be the detection of the $F = 0$ locations, and therefore, the medial axis. However, if the intensity value within the region is not constant, then the disk of radius $\sigma_s(p)$ centered at a point on the medial axis will not in general have isotropic intensity distribution about the point. The force at the point will not be 0, and the points where $F = 0$ may be off the medial axis. Therefore, when shading is present, the detection of medial axis is less reliable than the detection of region borders. With regard to noise, there will be no expected change in results for noisy piecewise constant regions for reasons analogous to those given for off-axis signatures. However, in shaded regions, noise will further increase the medial axis deviation beyond that already present due to shading alone. This is because the function d_g in the integral defining the transform is in general nonlinear, and therefore, in the presence of shading, uniform noise distribution around the point will cause unequal deviations in the F value in different directions, which will lead to deviation in the location of the $F = 0$ curve. Finally, let us consider the effect of different choices of F. For the piecewise constant case, changes in F will not cause any deviation in the mdeial axis because of symmetry. However, when shading is present different choices of the nonlinear function d_g will in general result in different distributions of force magnitudes and hence deviation in the $F = 0$ curve.

5 EXPERIMENTS

This section describes some of the experiments we have performed with the transform. We estimated scale parameters at all pixels as described in Section 3.5. A detailed treatment of algorithms for automatic extraction of scale parameters is outside the scope of this paper, and will be presented in the future. The purpose of the experiments reported here is to illustrate: the various properties of the transform (Section 3.3); the F-field signatures of regions (Section 3.4); the stability of the region signatures at all naturally occurring scales (Section 3.5); the detection of hierarchical image structure amidst a priori unknown and spatially varying number of scales (Section 3.6); and finally the robustness of the above performance of the transform with respect to a number of factors (Section 4).

Fig. 4 demonstrates the properties of the transform and the region signatures in the F-field through a simple synthetic image. Fig. 4a shows two rectangular regions having a constant intensity of 1, on a background of intensity 255. Estimates of σ_s were obtained and used to derive the F-field. Any errors in σ_s will result in errors in the F-field. Consequently, the performance of the transform as demonstrated in our experiments serves as a lower bound on the true performance. Fig. 4b shows an intensity coded image of the estimates of σ_s values used for the rectangular objects in Fig. 4a; the higher the σ_s is, the brighter the point appears. The transform was applied using the pulse (boxcar) function for each of d_s and d_g. Specifically, the function shown in Fig. 4a was used where σ indicates $\sigma_s(\sigma_g)$ value used by $ds(dg)$. Fig. 4c shows intensity coded directions of the force vectors computed using $\sigma_g = 20$, and the σ_s values shown in Fig. 4b. Since 20 is smaller than the contrast value 254 of the rectangles, the signatures of the rectangles as well as the background region of Fig. 4a can be seen in Fig. 4c. In producing Fig. 4c, the cyclic values of the directions in the interval [0 degrees, 360 degrees] are linearly mapped onto the intensity interval [0-255]. The 0-degree direction is defined to coincide with the positive x-axis, and is arbitrarily mapped onto a value in the intensity interval, e.g., to a value near 0. Thus, the brightness of a point is proportional to the counterclockwise angle the force vector makes from the reference, say, 0-degree direction. The black to white transitions in Fig. 4c between neighbor gray levels of 0 and 255 correspond to force direction transitions between nonnegative and negative angles. Since directions are continuous across the 0-degree direction, the associated intensity transition is an artifact of the mapping. For example, the black to white transitions across the 45-degree, diagonal lines in Fig. 4c, e.g., the one starting from the bottom left corner, are such artifacts. Other diagonal lines (e.g., at 135-degree) do not suffer from similar intensity discontinuities because they do not represent transitions across the reference direction. To help recognize and disregard these artifacts, an additional, redundant intensity coded direction image is shown in Fig. 4d. It still uses linear mapping from directions to intensities but the reference direction is now placed at 90 degrees instead of 0 degrees as above. Consequently, those transitions across

the 90 degree direction now result in the artifactual, 0 to 255 intensity transitions, instead of transitions across the 0-degree direction. The artifacts can be identified by comparing the direction coded images in Figs. 4c and 4d since the same direction transition will not appear as an artifact in both images. If directions were color coded, these visual artifacts would not occur. However, intensity coding is less expensive but workable and hence it is used here. Fig. 4e avoids the mapping completely and draws the vectors explicitly, although because it takes more space to draw vectors the vectors are shown at every fifth pixel along rows and columns. In each of Figs. 4c, 4d, and 4e, the force directions at points inside the regions demonstrate the inward attraction property while their gradual spatial variation is in accordance with the smoothness property. The force divergence and convergence can be seen across region boundaries and medial axes, respectively. Along the medial axes, the values of F can be seen to be 0 at points where the nearest boundary segments are parallel. This is particularly clear in Fig. 4e where the length-0 vectors appear only as dots. (Due to some limitations of our current implementation, σ_s value at other medial axis points is greater than δ and not equal to δ, as defined in Section 3. This is the reason for nonzero F values at medial axis segments associated with the corners.) The orthogonality property is demonstrated by the 90-degree angle between a rectangle's edge away from corners, and the force vectors at points near a rectangle's edge. The closed contours formed by points where force vectors diverge are consistent with the closure property.

Fig. 5 demonstrates the multiple F-fields provided by the transform using $\sigma_s \sigma_g$ values corresponding to the different scales present in the image (Section 3.5). Fig. 5a shows a modified version of the image in Fig. 4a, where a smaller rectangular region has been inserted within the horizontal rectangular region, and a square region has been inserted within the background. This leads to four foreground regions. The new rectangle has a gray level of 30 and hence a contrast of 29, and the new square has a gray level of 200 and hence a contrast of 55. Thus, there are three different contrasts present within the image, one (254) corresponding to the two old rectangles, and the two new values of 29 and 55. Figs. 5b, 5c, and 5d show intensity coded direction images of F-field obtained using σ_g = 10, 40, and 60, respectively. Since 10 is lower than all three contrasts present, all four foreground regions are detected as can be seen from their signatures present in the F-field which is obtained using σ_s values which take into account the new rectangle. For σ_g = 40, which is higher than the contrast of the new rectangle, the signatures of the new rectangle are absent in the F-field. This is shown in Fig. 5c. Finally, Fig. 5d shows the intensity coded direction image obtained using σ_g = 60, which is higher than both new contrasts, and the σ_s values used are the same as those shown in Fig. 4b, corresponding to the two old rectangles. The result is that neither of the new regions is detected. This can be seen by comparing the intensity coded direction images in Fig. 4c and Fig. 5d which are identical. An arbitrary point in the image assumes as many stable values of F as the number of different regions containing it. For a point within the square, the F value changes from one stable value V_1 corresponding to the square, to another value V_2 for the case when the square is missing. If $\sigma_g > 255$ were used, the F-field computed using σ_s values corresponding to the whole image as a single region would show the signatures of the square image with no other region. Thus, for a point within the new rectangle, there will be three stable points in the $\sigma_s \sigma_g$-space, corresponding to the three regions in which it is contained. For increasing values of (σ_s, σ_g), the value changes from V_1 corresponding to the embedded rectangle, to V_2 corresponding to the embedding rectangle, on to V_3 corresponding to the entire image.

Fig. 6 illustrates the invariance of the transform to boundary curvature and topology, which is a problem with the existing approaches and a major motivation for the proposed transform. Fig. 6a contains constant-value regions with a complex boundary structure. The maplike region in the foreground has a jagged boundary. The checkerboard background contains vertices where boundaries of four adjacent square regions meet. Four of the checkerboard squares near the top of the image have been further divided into two triangles each, leading to a common vertex where the boundaries of eight triangular regions meet. Two triangular regions with very sharp corners are also included as a further challenge to the curvature invariance of transform. The smaller of the two sharp angles has the low value of about 5 degrees. All region boundaries have a contrast of at least 15. Fig. 6b shows intensity coded direction images computed using σ_g = 10 < 15. Therefore, the σ_s value used for each point is the one corresponding to the smallest region containing it, regardless of the contrast of the region. Undistorted signatures of all the regions can be seen in Fig. 6b, except for the artifacts caused by the direction to intensity mapping. As before, a redundant direction image is provided in Fig. 6c, which is obtained using the alternate mapping using reference direction of 90 degrees. From Figs. 6b and 6c, all regions can be seen to have been identified without distortion. For example, the contours of force divergence coincide with the jagged edges of the maplike region. This will be difficult to achieve with the conventional methods that apply a priori models of boundary structure over pixel neighborhoods, since the actual edge structure could be different or more intricate than that incorporated in the model. The sharp corners of the triangular regions demonstrate insensitivity to boundary complexity more emphatically since in the vicinity of the sharp corner of an unknown angle any a priori model of local geometry is likely to be violated. The performance of the transform near the vertices is more notable since a priori models of vertex topology are even less likely to be valid. The artifacts of direction-intensity coding are almost unavoidable near the sharp corners since the force directions here span almost the entire 360-degree range making it very likely for the 0 degree to 360 degree transitions to occur. The medial axes of the regions can also be seen as force convergence contours.
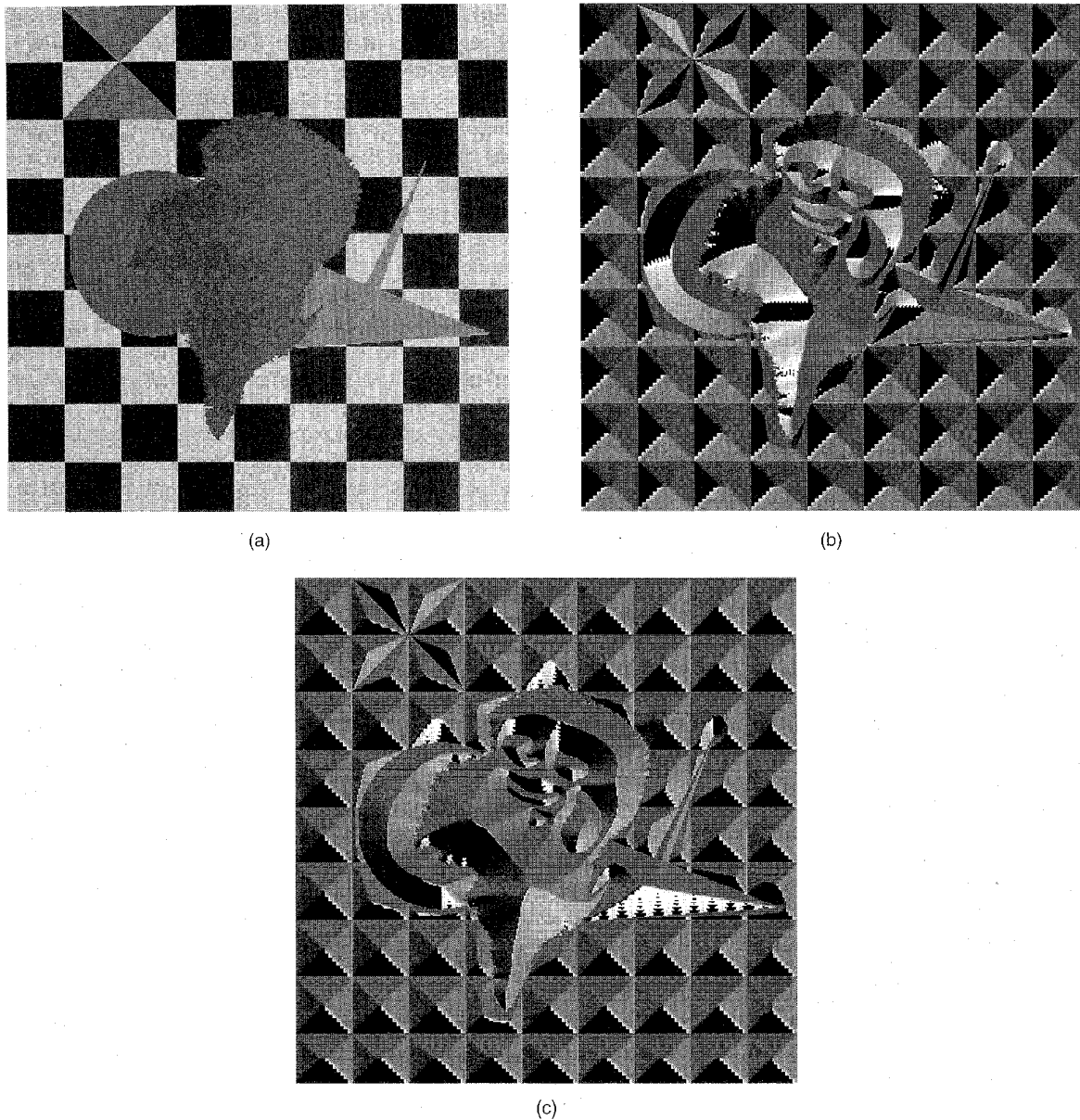
(a)

(b)

(c)

Fig. 6. Demonstration of the insensitivity of the transform to boundary curvature and topology. (a) Constant value regions having complex boundary structure. The boundaries contain smooth as well as high-curvature segments including corners. Some of the boundary points are vertices where more than two regions meet. All region boundaries have a contrast of at least 15 grey levels. (b) Intensity coded directions computed by the transform using σ_s values chosen corresponding to the region structure for $\sigma_g = 10$. (c) Same as (b), but with vector directions rotated 90° clockwise and then intensity coded. The regions are represented in (b-c) by their force field signature without distortion.

Fig. 7 shows the insensitivity of the transform performance to the choice of the function F, i.e., the functions d_s and d_g. Figs. 7a-7d show intensity coded directions of F for the image of Fig. 6a using four different types of ds and dg:

(a) boxcar (pulse),
(b) Gaussian,
(c) exponential, and
(d) linear.

The same function is used for both d_s and d_g. The signatures of the rectangles as described in Section 3.4 can be clearly seen in each case. The direction images for cases (b), (c),

and (d) are visually almost identical. Clearly, the magnitudes of the vectors and, to a lesser extent, even their directions, cannot be expected to be identical because the integrand in (1) is different for the different cases. But what is important is that the spatial features of F which determine the relevant image structural properties (Section 3.3) be the same. Consequently, the same image structure is seen in each of the cases (a-d) despite the differences that must exist among the F-fields. Some minor differences can be seen between the direction image for case (a) and those for cases (b-d). This is because the fluctuations in F-values computed in the different cases lead to direction transitions across the
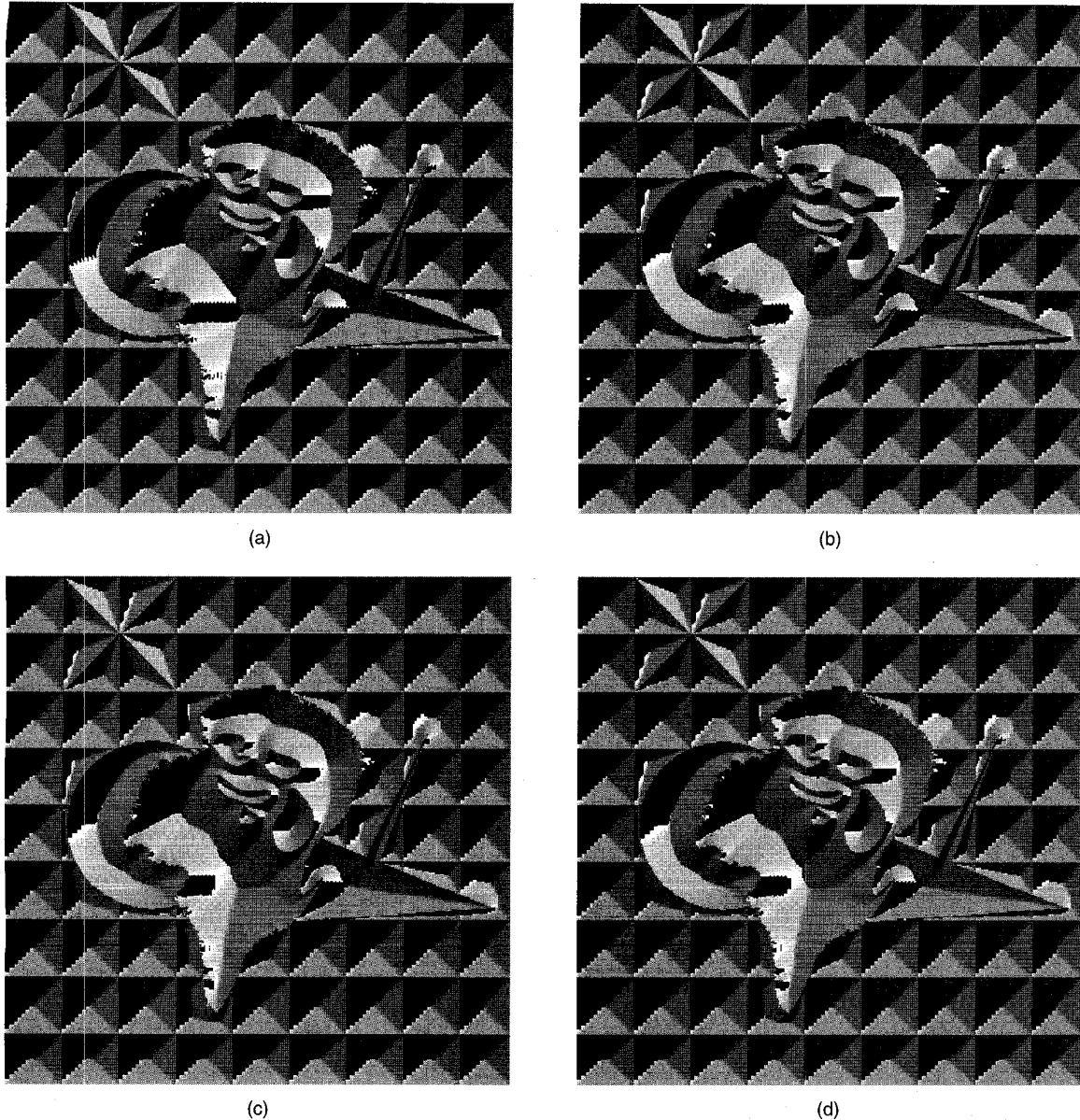
(a)

(b)

(c)

(d)

Fig. 7. Demonstration of the insensitivity of the transform to the choice of the functions $d_g(\cdot)$ and $d_s(\cdot)$. (a-d) Intensity coded directions of Fig. 6a computed by the transform using σ_s values chosen corresponding to the region structure for $\sigma_g = 10$. The forms used for $d_g(\cdot)$ and $d_s(\cdot)$ are box-car, Gaussian, exponential, and linear, respectively. The region structure is represented within the field without distortion in all cases.

reference direction near the sharp triangle corners, and within the checkerboard square in the fourth row and column from bottom left square), which lead to visual differences that are predominantly mapping artifacts. Any remaining differences in F values are minor and due to fluctuations in F-values which arise primarily from fluctuations in the values of d_g; the values of d_s do not vary much across (a-d). These minor differences in F can be minimized by suitably normalizing the different d_s/d_g functions with respect to each other. Such calibration may involve scaling to find the optimum values of the various parameters of the different functions, for example, the values of k_s and k_g mentioned in Section 4.2.2. The small differences among (a-d)

suggest that the segmentation performance of the transform is relatively independent of the exact choice of F. Therefore, the distinguishing aspect of the transform which is responsible for its performance is $\hat{r}$, which is the only remaining part of the integrand in (1) not considered so far. The use of $\hat{r}$ reflects the fact that vector integration is being performed, and results in the bottom-up estimation of image structure which, as we argued in Section 3.1, is a central theme of the proposed approach and represents a major deviation from the previously used methods.

Fig. 8 illustrates the sustained performance of the transform when the regions are no more piecewise constant in gray level but contain shading. Fig. 8a shows the same im-

(a)

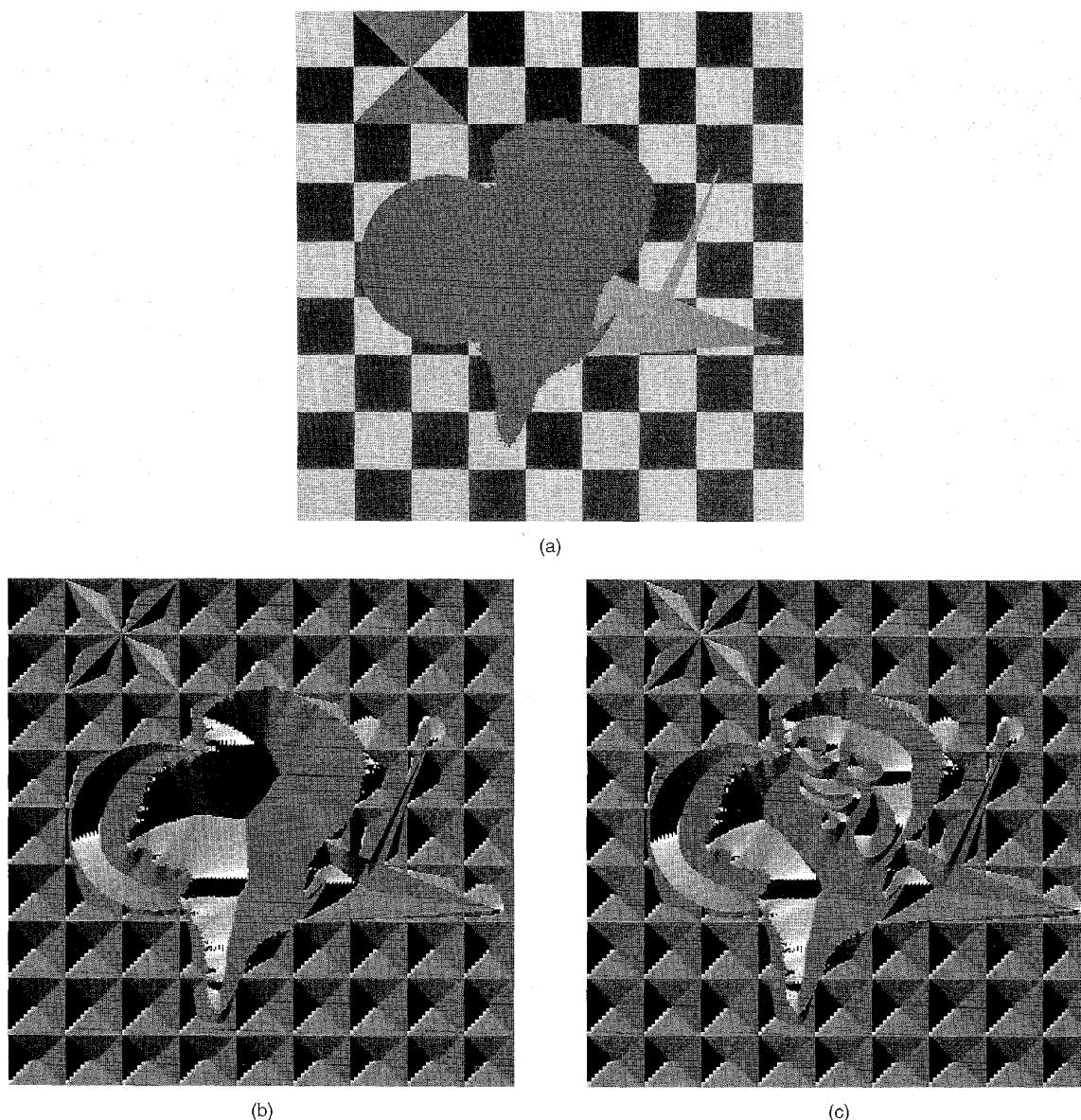(b)   (c)

Fig. 8. Demonstration of the insensitivity of the transform to shading. (a) Same image as in Fig. 6a, but different regions now have spatially varying grey levels. Here, the disk contains linearly increasing grey levels from left to right, whereas the polygonal region has quadratically increasing grey levels from right to left. The irregularly shaped center region contains quadratically varying grey levels which increase from bottom to top. (b) Intensity coded force directions after the transform has been applied using $\sigma_g = 10$. (c) Same as (b) but with $\sigma_g = 25$. At this value of σ_g, the microstructure within the irregular center region has disappeared.

age as in Fig. 6a but the gray levels within every region are not constant but spatially vary linearly or quadratically. The intensity coded force directions shown in Figs. 8b and 8c demonstrate that the inward attraction, divergence, convergence, smoothness, and closure properties are preserved despite the violation of the constant gray level assumption.

Fig. 9 shows the effect of noise on the performance of the transform using an image with three regions whose boundaries contain sharp corners as well as smooth curves. Fig. 9a shows the image with additive Gaussian noise having a standard deviation of 50. This results in the worst case signal to noise ratio of about 7 db. Figs. 9b

and 9c show that regions are detected without distortion. Fig. 9d contains additive Gaussian noise of standard deviation 100, i.e., a worst case signal to noise ratio of 0 db. The worst ratio occurs at the border between the polygonal region and the background, which is particularly difficult to overcome near the sharp corner. The F-field still contains all regions, including the noisy sharp corner. In addition, signatures of some small spots are also present, some of which can be seen in Figs. 9b and 9c as artifacts of mapping. But other small spots represent new small regions formed by random clumping of noisy pixels. However, as stated in Section 3.5, the final image structure
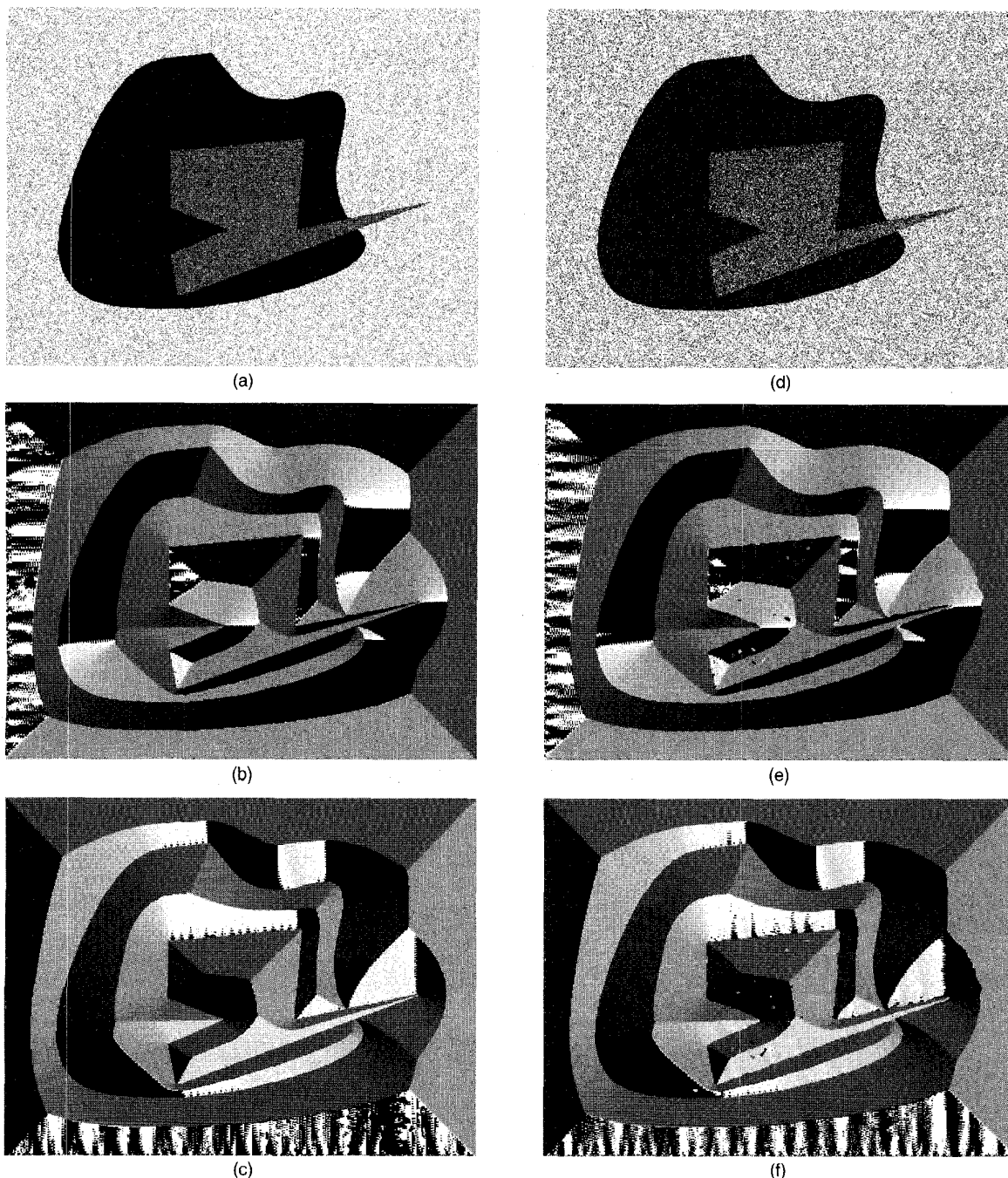
Fig. 9. Demonstration of the insensitivity of the transform to additive, white Gaussian noise. (a) Image containing two regions of intensity 0 and 100 on a background of intensity 200, and with zero-mean noise of standard deviation $\sigma = 50$. (b) Intensity coded directions computed using $\sigma_g = 50$. (c) Same as (b), but with vector directions rotated $90°$ counterclockwise and then intensity coded. (d-f) Same as (a-c), but with noise having $\sigma = 100$. Note that the structure boundaries are all present, have closed contours, and retain geometric fidelity. (For example, the sharp corner of the intensity 100 region has been preserved without smoothing, except, of course, where the region structure has been changed because of the noise, in which case the new structure is reflected.)

detected by the algorithm will be that which corresponds to stable points in the $\sigma_s\sigma_g$-space. Thus, only a subset of the identified structures that survives the dynamics in the $\sigma_s\sigma_g$-space would comprise the detected regions. As mentioned earlier, the details of such multiscale algorithm beyond Section 3.5 are not presented in this paper.

Figs. 10 to 13 show the results of the transform on four real images, chosen to form a diverse test set. Fig. 10a shows the first image which contains relatively homogeneous regions at multiple scales. For all points in the image, region signatures are obtained for two, relatively fine scales corresponding to σ_g values of 7 and 26. Figs. 10b and 10c
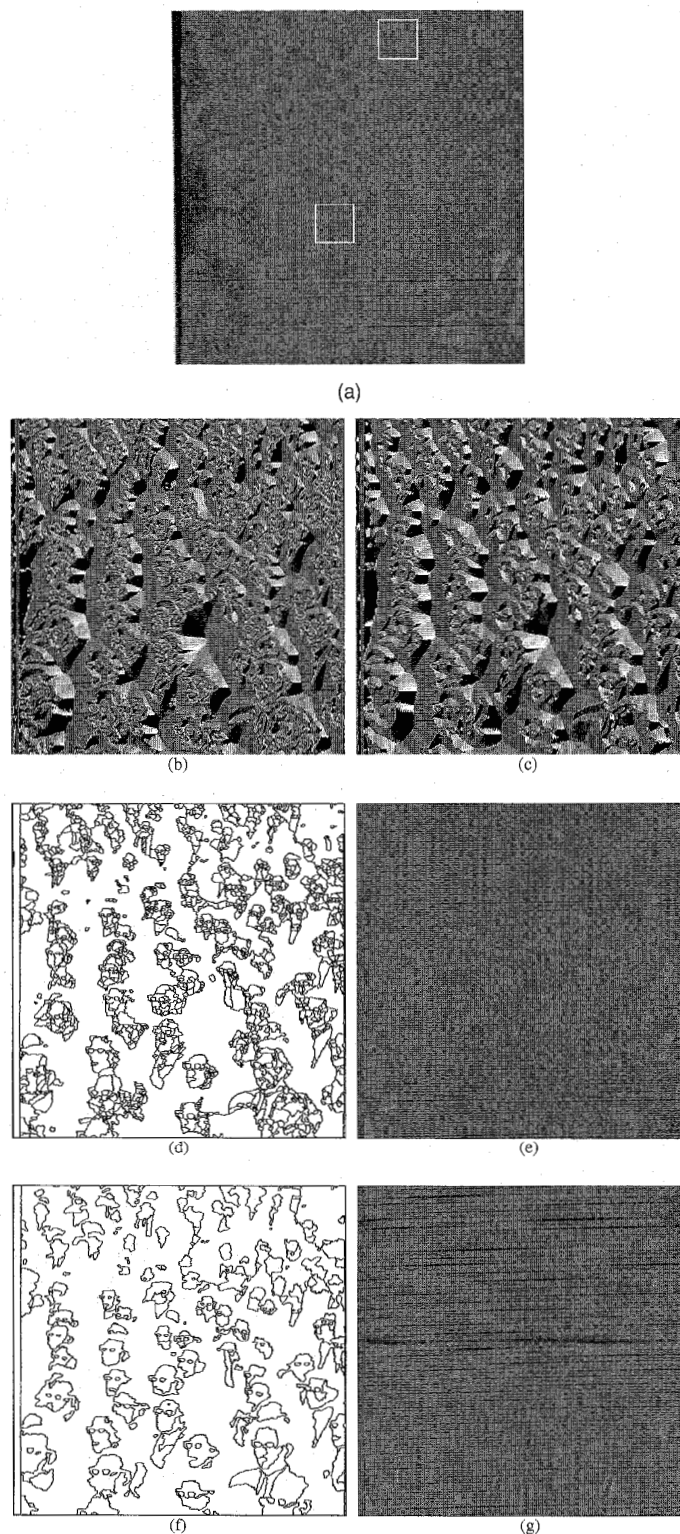
Fig. 10. Performance of the transform on a real image with several geometric and photometric scales. (a) The original image. The multiscale structure in the lower window is independent of that present in the upper window, i.e., the number of levels and branching factors in the segmentation subtrees containing the two windows are independent. In general, the structure of the multiscale segmentation hierarchy is determined solely by the image, and, therefore, is a priori unknown. (b-c) Intensity coding of the force directions computed for this image using $\sigma_g = 7$ and $\sigma_g = 26$, respectively. (d) Contours characterized by force divergence at $\sigma_g = 7$. (e) Pixels enclosed by each closed contour defined in (b) are replaced by the mean graylevel of the region, thus showing the segmentation at that scale. (f-g) Analogous to (d-e) but obtained from **F** for $\sigma_g = 26$.
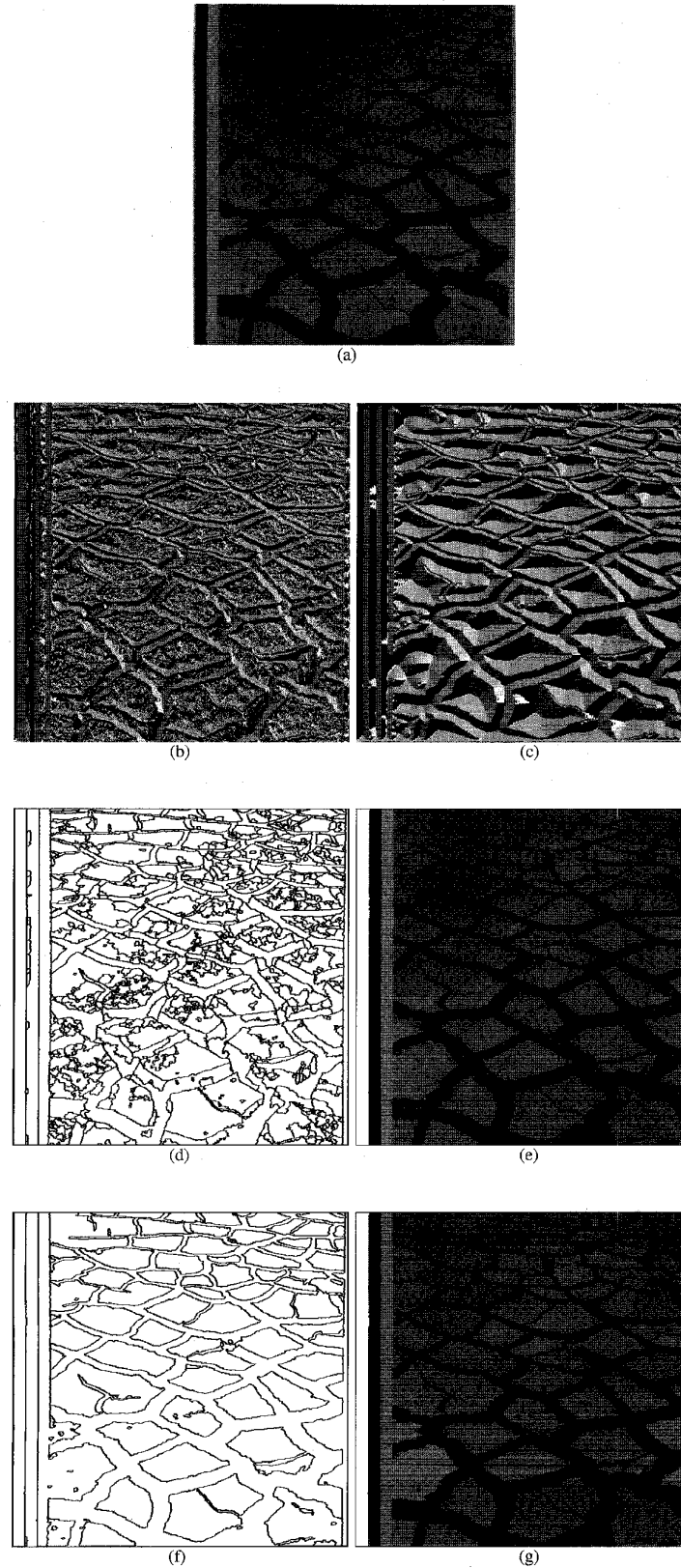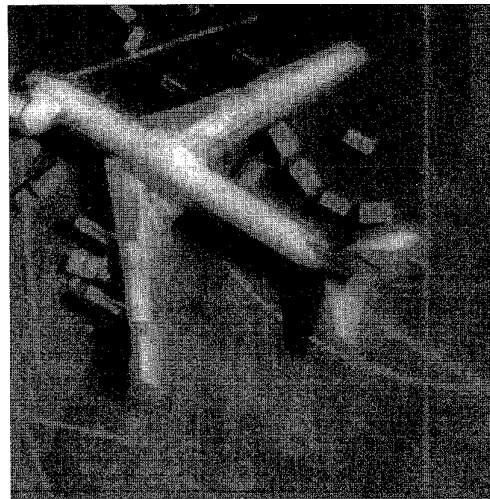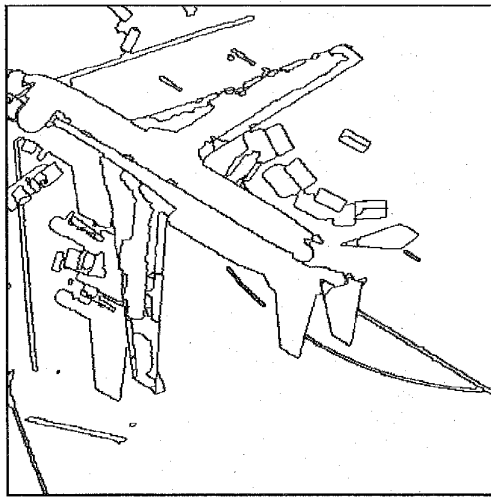
Fig. 11. (a) A real image containing structure at predominantly two scales, with jagged edges particularly visible inside the lighter, cellular regions. The structure is easy to follow visually, which makes it easy to test the transform performance with respect to boundary complexity. (b-g) Analogous to Figs. 10b-10g using $\sigma_g = 7, 26$, corresponding to the two major scales.
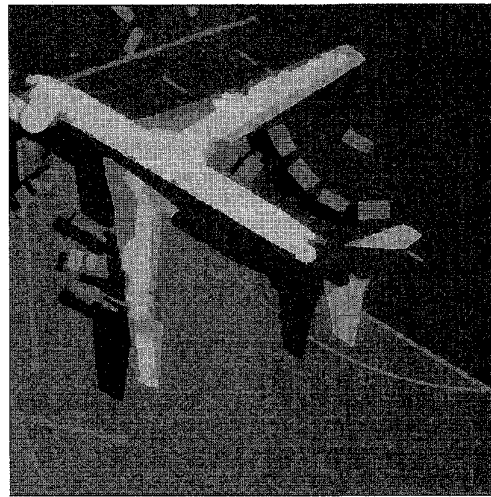
(a)

(b)

(c)

Fig. 12. Performance of the transform on another real image. The image contains an airplane with a large amount of intensity shading. (a) The original image. (b-c) Analogous to Fig. 10f-10g.

show the intensity coded direction images of F for the two scales. The signatures are shown in two ways. Figs. 10d and 10f show the contours of force divergence, i.e., the region border, for the two scales. Figs. 10e and 10g replace each pixel enclosed by a region contour by the mean gray level of the region, thus showing the segmentation at the corresponding scale. Fig. 11a (provided by Tom Binford of Stanford University) shows another real image which contains structure at predominantly two scales. The structures have jagged edges, particularly inside the cellular regions. This structure is easy to follow visually which makes it easy to test the transform performance with respect to the boundary complexity. Figs. 11b-11g show the results analogous to Figs. 10b-10g using the same two σ_g values as used in Fig. 10. Fig. 12a shows a real image which has significant amount of shading, e.g., across the airplane body. For brevity, the direction coded images are not shown, and the contours and segmentation are presented for only one σ_g value (Figs. 12b and 12c). The σ_g value used is the larger of the two used in Fig. 10. As

can be seen, regions are detected correctly despite the large amount of shading. Finally, Fig. 13a contains a laboratory scene, selected since it shows an indoor, human made environment. Results analogous to Figs. 12b and 12c using $\sigma_g = 17$ are shown in Figs. 13b and 13c.

6 SUMMARY

We have introduced a transform for integrated detection of image edges and regions for general-purpose, low-level image segmentation. A region is viewed as being homogeneous but in relative contrast with the surround. The main feature of the transform is that it lets the image structure emerge bottom-up by allowing pixels to group recursively to form regions, instead of fitting detailed models of region geometry and photometry to the data. The transform computes a force vector at each image point which represents the point's affinity to the rest of the image. The resulting force field contains distinct signatures of image regions. The signatures succinctly reflect the distributed evidence of
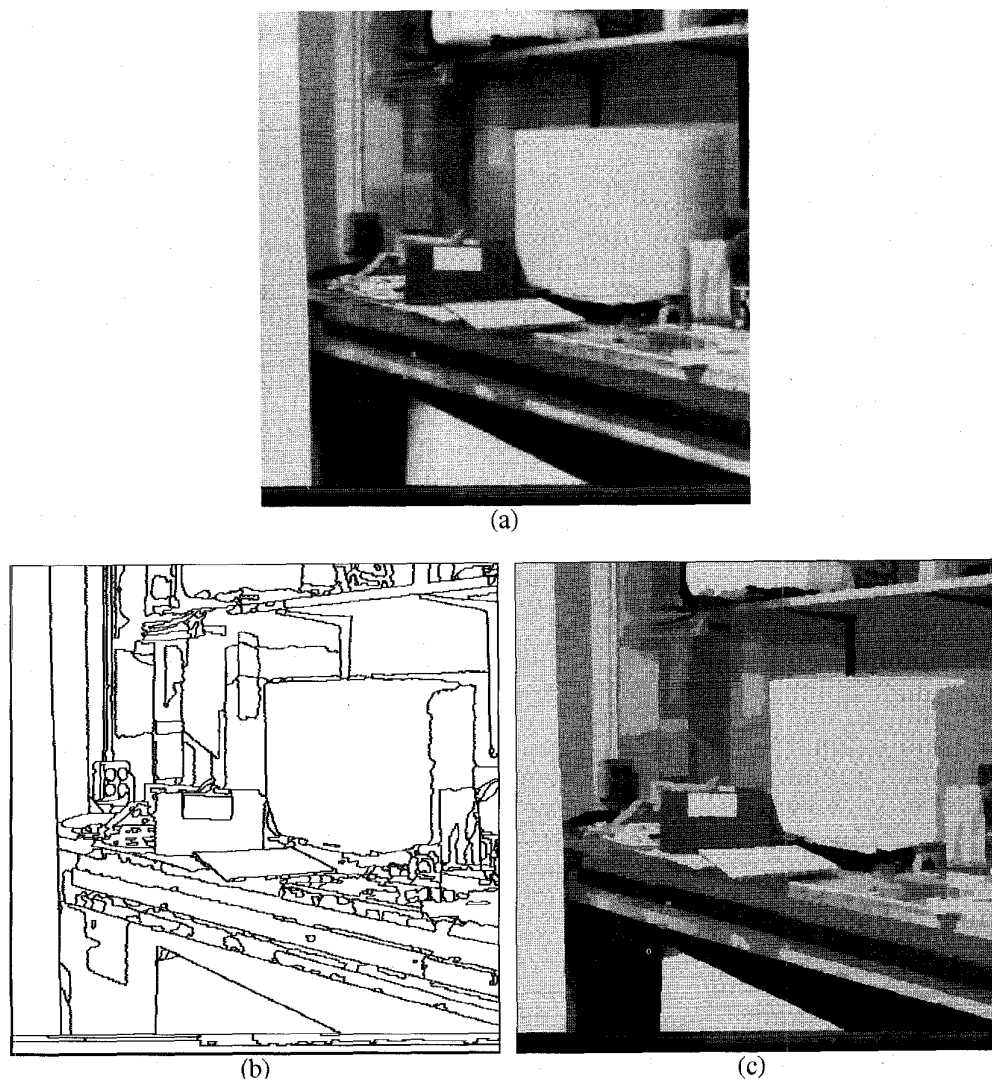
Fig. 13. Performance of the transform on another real image, an indoor scene of a workbench. (a) The original image. (b-c) Analogous to Fig. 10f-10g but for $\sigma_g = 17$.

the presence of a region and are characterized by a number of relatively qualitative properties which make their detection robust and easy. No critical thresholds are necessary. The transform-based segmentation has the following salient properties:

1) The regions are detected accurately, regardless of their size and shape. The detected region boundary has no dislocation errors, irrespective of the boundary curvature.

2) The regions are detected correctly, regardless of their topology, e.g., the edges are found without errors even around a vertex where a number of regions meet.

3) Valid regions are detected independent of the value of the contrast they have with the surround.

4) The notion of scale is formulated such that multiscale analysis sorts the regions in the image based on geometric and photometric criteria. This is in contrast with most previous work where multiscale analysis is synonymous with different degrees of image blurring which entails deformation of the original image regions.

5) Regions of all scales (sizes, shapes, homogeneity, contrast) and topology are detected without any advance knowledge of these parameters.

The detected regions are organized into a hierarchical, tree structure, which captures the image-specific and a priori unknown region configuration and scale values present in the image. Thus the use of the transform yields a tree representation of the image whose depth and structure correspond to the image at hand. This differs from the commonly used multiscale methods where the number and specification of the different degrees of blurring to be performed are selected a priori. I have demonstrated the above properties through experiments. The segmentations obtained in agreement with real/perceived structure even for

images with large noise, intricate geometric and topological structure, and shading.

Another important property of the transform is that the transform space is the same as image space, unlike, for example, the Fourier transform. The resulting multiscale representation is directly useful for analysis by humans, e.g., for image browsing, manipulation, and retrieval. Also, the transform could be used in other applications where determination of perceptually salient image structure is critical or advantageous, such as image compression, motion analysis, texture analysis, and perceptual grouping. My objective in this paper has been to present the transform and its capabilities; details and different applications are left to future work. I have not given specific algorithms for using the transform to automatically estimate the scales and identify region signatures. These will also be reported later. Since no assumptions specific to the sensing modality are made in defining the transform, it could be applied to other types of data, such as range images and synthetics aperture radar images. It will be interesting to compare the transform with force-based clustering algorithms [15], which we also plan to do.

ACKNOWLEDGMENTS

REFERENCES

[1] N. Ahuja, "A Transform for Detection of Multiscale Image Structure," *Proc. Computer Vision and Pattern Recognition,* pp. 780-781, New York, 1993.

[2] N. Ahuja and J. Chuang, "Shape Representation Using a Generalized Potential Field Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* to appear.

[3] N. Ahuja, "A Transform for Multiscale Image Segmentation," Beckman Institute Technical Note, no. 96-01, July 1995.

[4] V. Berzins, "Accuracy of Laplacian Edge Detectors," *Computer Vision, Graphics, and Image Processing,* vol. 27, pp. 195-210, 1984.

[5] T. Binford, "Inferring Surfaces from Images," *Artificial Intelligence,* vol. 17, pp. 205-244, Aug. 1981.

[6] H. Blum, "A Transformation for Extracting New Descriptors of Shape," *Models for the Perception of Speech and Visual Form,* edited by W. Dunn. Cambridge, Mass.: MIT Press, pp. 362-380, 1967.

[7] R. Boomgaard and A. Smeulders, "Towards a Morphological Scale-Space Theory," *Shape in Picture: Mathematical Descriptions of Shape in Grey-Level Images.* New York: Springer-Verlag, pp. 631-640, 1992.

[8] J.F. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 679-698, 1986.

[9] J.J. Clark, "Authenticating Edges Produced by Zero-Crossing Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 43-57, Jan. 1989.

[10] L.S. Davis, "A Survey of Edge Detection Techniques," *Computer Graphics and Image Processing,* pp. 248-270, 1975.

[11] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transac-*

[12] R.M. Haralick, "Digital Step Edges from Zerocrossings of Second Directional Derivative," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 58-68, 1984.

[13] R.M. Haralick and L. Shapiro, "A Survey of Image Segmentation Techniques," *Computer Vision, Graphics and Image Processing,* vol. 29, pp. 100-132, 1985.

[14] S.L. Horowitz and T. Pavlidis, "Picture Segmentation by a Directed Split and Merge Procedure," *Proc. Second Int'l Joint Conf. Pattern Recognition,* pp. 424-433, 1974.

[15] A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data.* Englewood Cliffs, N.J.: Prentice Hall, 1983.

[16] J.A. Koenderink, *Solid Shape.* Cambridge, Mass.: MIT Press, 1990.

[17] T. Lindeberg, *Scale-Space Theory in Computer Vision.* Boston: Kluwer Academic, 1994.

[18] T. Lindeberg and J.O. Eklundh, "Scale-Space Primal Sketch: Construction and Experiments," *Image and Vision Computing,* pp. 3-18, Jan. 1992.

[19] Y. Lu and R.C. Jain, "Behavior of Edges in Scale Space," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 11, pp. 337-356, Apr. 1989.

[20] J. Malik and P. Perona, "Edge Detection by Diffusion," technical report, Univ. of California, Berkeley, 1986.

[21] D. Marr and E. Hildreth, "A Theory of Edge Detection," Royal Society of London, vol. B-207, pp. 187-217, 1980.

[22] F. Meyer and S. Beucher, "Morphological Segmentation," *Vis. Comm. Image Representation,* vol. 1, no. 1, pp. 21-46, 1990.

[23] A. Montavert, P. Meer, and A. Rosenfeld, "Hierarchical Image Analysis Using Irregular Tesselations," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 13, pp. 307-316, 1991.

[24] B.S. Morse, S.M. Pizer, and A. Liu, "Multiscale Medial Analysis of Medical Images," *Proc. 14th Int'l Conf. Information Processing in Medical Imaging,* edited by H. Barrett and A. Gmitro. New York: Springer-Verlag, 1993.

[25] D. Mumford and J. Shah, "Boundary Detection by Minimizing Functionals, i." *IEEE Conf. Computer Vision and Pattern Recognition,* pp. 22-26, 1985.

[26] V. Nalwa and T. Binford, "On Detecting Edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 8, pp. 699-714, 1986.

[27] V.S. Nalwa, "Experiments with a Spatiotemporal Correlator," *IEEE Conf. Computer Vision and Pattern Recognition,* pp. 712-716, 1992.

[28] P. Perona and J. Malik, "Detecting and Localizing Edges Composed of Steps, Peaks and Roofs," *Third Int'l Conf. Computer Vision,* pp. 52-57, Osaka, Japan, Dec. 1990.

[29] J. Ponce and M. Brady, "Toward a Surface Primal Sketch, Technical Report MIT-AI-TR-824, Massachusetts Institute of Technology, 1985.

[30] A. Rosenfeld and A. Kak, *Digital Picture Processing.* New York: Academic Press, vol. 2, 1981.

[31] A. Rosenfeld and M. Thurston, "Edge and Curve Detection for Visual Scene Analysis," *IEEE Transactions on Computers,* pp. 562-569, 1971.

[32] R. Whitaker and S. Pizer, "Geometry-Based Image Segmentation Using Anisotropic Diffusion," *Shape in Picture: Mathematical Descriptions of Shape in Grey-Level Images.* New York: Springer-Verlag, pp. 641-650, 1992.

[33] A. Witkin, "Scale Space Filtering," *Int'l Joint Conf. Artificial Intelligence,* 1983.

[34] A. Yuille and T. Poggio, "Scaling Theorems for Zero Crossings," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 8, pp. 15-25, 1986.

[35] M. Brady and H. Asada, "Smoothed Local Symmetries and Their Implementation," *Int'l J. Robotics Research,* vol. 3, no. 3, pp. 36-61, 1984.

[36] J.B. Subirana-Vilanova, "Curved Inertia Frames and the Skeleton Sketch: Finding Salient Frames of Reference," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 702-708, 1990.

tions on Pattern Analysis and Machine Intelligence, vol. 6, no. 11, pp. 721-741, 1984.
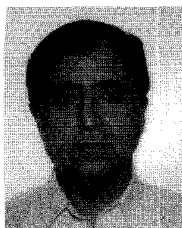
Narendra Ahuja (F'92) received the BE degree with honors in electronics engineering from the Birla Institute of Technology and Science, Pilani, India, in 1972; the ME degree with distinction in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1974; and the PhD degree in computer science from the University of Maryland, College Park, USA, in 1979.

From 1974 to 1975 he was Scientific Officer in the Department of Electronics, Government of India, New Delhi. From 1975 to 1979 he was at the Computer Vision Laboratory, University of Maryland, College Park. Since 1979 he has been with the University of Illinois at Urbana-Champaign, where he is currently (1988-) a professor in the Department of Electrical and Computer Engineering, the Coordinated Science Laboratory, and the Beckman Institute. His interests are in computer vision, robotics, image processing, image synthesis sensors, and parallel algorithms. His current research involves integrated use of multiple image sources of scene information to construct 3D descriptions of scenes; the use of integrated image analysis for realistic image synthesis; parallel architectures and algorithms and special sensors for computer vision; and use of the results of image analysis for a variety of applications, including visual communication, image manipulation, video retrieval, robotics, and scene navigation.

He was a Beckman associate in the University of Illinois Center for Advanced Study during 1990-1991. He received the University Scholar Award (1985), Presidential Young Investigator Award (1984), National Scholarship (1967-1972), and President's Merit Award (1966). He has coauthored the books *Pattern Models* (Wiley, 1983) and *Motion and Structure from Image Sequences* (Springer-Verlag, 1992). He is a fellow of the IEEE, American Association for Artificial Intelligence, International Association for Pattern Recognition, ACM, American Association for the Advancement of Science, and International Society for Optical Engineering. He is a member of the Optical Society of America. He is on the editorial boards of *IEEE Transactions on Pattern Analysis and Machine Intelligence; Computer Vision, Graphics, and Image Processing; Journal of Mathematical Imaging and Vision;* and *Journal of Information Science and Technology* and is a guest coeditor of the *Artificial Intelligence Journal*'s special issue on vision.