Scale Space Classification Using Area Morphology

Scott T. Acton, *Senior Member, IEEE,* and Dipti Prasad Mukherjee

***Abstract*—We explore the application of area morphology to image classification. From the input image, a scale space is created by successive application of an area morphology operator. The pixels within the scale space corresponding to the same image location form a scale space vector. A scale space vector therefore contains the intensity of a particular pixel for a given set of scales, determined in this approach by image granulometry. Using the standard k-means algorithm or the fuzzy c-means algorithm, the image pixels can be classified by clustering the associated scale space vectors. The scale space classifier presented here is rooted in the novel area open–close and area close–open scale spaces. Unlike other scale generating filters, the area operators affect the image by removing connected components within the image level sets that do not satisfy the minimum area criterion. To show that the area open–close and area close–open scale spaces provide an effective multiscale structure for image classification, we demonstrate the fidelity, causality, and edge localization properties for the scale spaces. The analysis also reveals that the area open–close and area close–open scale spaces improve classification by clustering members of similar objects more effectively than the fixed scale classifier. Experimental results are provided that demonstrate the reduction in intraregion classification error and in overall classification error given by the scale space classifier for classification applications where object scale is important. In both visual and objective comparisons, the scale space approach outperforms the traditional fixed scale clustering algorithms and the parametric Bayesian classifier for classification tasks that depend on object scale.**

***Index Terms*—Image classification, nonlinear filters, scale space.**

I. Introduction

IN CLASSIFYING image pixels within a digital image, the scale of a particular object, as well as the intensity, may be important in evaluating class membership. Conventional image classification techniques using the original input image at a fixed feature scale may result in intraobject classification errors (assigning two different class labels to pixels of the same object). These techniques can be sensitive to localized image noise that differs in intensity but is of limited area. In this paper, we propose an image classification approach that utilizes an image scale space, a coarse-to-fine collection of image representations to detect objects at different scales. Vectors in scale space that define pixel intensity through a range of scales are used to classify the image. The utilization of the scale space vectors for classification is grounded on the premise that no single scale representation is sufficient to properly label each pixel for classification.

The scale space used in this approach is based on *area morphology* operators [18]. The area morphology operators generate scaled images based on the area of connected components within the image level sets (thresholded versions of the grayscale image). Here, the *area open–close* and *area close–open* scale spaces are introduced and used in image classification. We show that the area open–close and close–open scale spaces do not introduce new features with increased scale, so that the multiscale representation does not produce classification errors due to processing artifacts. Moreover, we prove that edge locations, as defined by boundaries of connected components, will not move with increased scale, leading to faithful preservation of region boundaries in the image classification. This property, in particular, is elusive for scale spaces generated using standard morphology. Further, unlike standard morphology, the area operators do not impose any specific shape (as determined by the shape of structuring element) on the processed image. With standard morphology, the boundaries of connected components within the image level sets may be distorted according to the shape of the structuring element used. Boundary distortion can lead to misclassification in a classification procedure.

Maragos [11] provides a multiscale shape description using morphological filters. He has used standard morphological opening and closing [8] with structuring elements of varying shape and size to generate a scale space for shape representation. For increasing or decreasing scale, specific binary patterns are self-dilated or eroded and subsequently used in open or close operations. The scale parameter is governed by the degree of self dilation or erosion of a given pattern. In contrast, the area based scaling in our approach depends on the threshold decomposition of the gray level image into image level sets and the area of the connected components within these level sets. This area morphology based processing is used as an effective precursor to classification in scale space.

Other attempts in multiscale classification hinge on subsampled pyramidal representations and hierarchical processing techniques. Multiscale, multiresolution representations are generated by successively filtering and subsampling the image. Then, a multiscale data structure, such as the quad tree or an image pyramid, can be used in image classification. The multiscale classification is generally performed in a sequential manner, namely, by utilizing the classification result of a coarse scale, low-resolution image to guide classification at a finer scale and higher resolution. Bouman and Liu [3] report a successful classification scheme for textured images that utilizes a multiscale, multiresolution representation in a hierarchical approach. Another hierarchical classification

Manuscript received January 18, 1999; revised October 27, 1999. This work was supported in part by NASA. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Robert J. Schalkoff.

S. T. Acton is with the Oklahoma Imaging Laboratory, School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078 USA (e-mail: sacton@okstate.edu).

D. P. Mukherjee is with the Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta, India 700035.

Publisher Item Identifier S 1057-7149(00)02671-3.

method is developed in [14], based on adaptive clustering. In this case, the grayscale image is modeled as a Markov random field and a Bayesian technique, in conjunction with adaptive multiscale windowing, is used to classify the image. Bouman and Shapiro also apply a multiscale Markov random field model for Bayesian image segmentation [4]. In a different context, a multiscale classification scheme is implemented using a rule-based inductive learning algorithm [10]. Multiresolution, multiscale classification is explored in [5] for the specific application of radar signal classification.

Although not common in image classification, *watersheds* have been used for image segmentation [7], [19]. Typically, the image gradient magnitude is viewed as a topographic surface that is subdivided by ridges. Between the ridges are valleys (watersheds) that share a common basin or local minimum in gradient magnitude. The watersheds thus define a segmentation of the image in which closed regions with thin, unbroken boundaries are provided. The drawbacks of the watershed approach include over-segmentation (too many watersheds), computational expense, and sensitivity to noise/detail. Unlike the area morphology approach described in this paper, the scale or size of the resultant regions cannot be specified directly in the watershed transform, without the use of prefiltering. The regions that emerge from the watershed segmentation are not necessarily connected components within the image level sets, as with the area morphology method used here.

Hierarchical multiscale segmentation approaches have been proposed that exploit the watershed transform. In this case, both morphological and nonmorphological approaches are used to generate scale spaces [6], [9], [13], [22]. However, these watershed-based multiscale representations lack some of the desirable scale space properties. In case of the area morphology approach, we develop a scale space that satisfies both the causality and edge localization properties. With the watershed approach, causality and edge position through scale space are not guaranteed. Furthermore, the area of connected components in the image level sets is used as the scale parameter in the area morphology classification method. The multiresolution watershed methods do allow scaling of the image, so as to avoid over-segmentation, but do not allow exact specification of object scale (e.g., watersheds/objects with area greater than 47 pixels). Finally, the multiresolution watershed segmentation techniques are hierarchical (coarse-to-fine), while the scale space classification approach simultaneously utilizes scaled image representations.

The image classification method developed in this paper differs from previous multiscale classifiers in that the classifier does not utilize subsampled image representations (pyramids, quadtrees), and the classification technique is not hierarchical. The major contribution of this paper is the establishment of a classification paradigm within the area morphology scale space in which all scale representations are used simultaneously. As the results and analysis reveal, the scale space classifier yields reduced intraobject classification error, in comparison to traditional classification methods, for classification tasks where object scale is important.

The organization of the paper is as follows. First, the theory and analysis of the area open–close and close–open scale spaces are presented in Section II. The application of the scale space to classification is also discussed in Section II, along with an analysis of the scale space classifier properties. As the scale space classifier can be used in conjunction with any clustering technique, the fundamental clustering techniques are reviewed and adapted for scale space classification in Section III. Qualitative results in the form of images and quantitative results in the form of tabulated data are given in Section IV that demonstrate the efficacy of the scale space classifier.

II. Scale Space Classification: Theory and Analysis

A. Area Operators

We implement classification on an image scale space that is generated using area operators that manipulate connected components within image level sets. Our processing methods assume both discrete domain and discrete range – digital imagery. For a set S defined on domain $\Omega \subset Z^2$, we have members of the *on-set*: $(x, y) \in S$ and members of the *off-set*: $(x, y) \in S^c$ where S^c is the complement of S. Two points (x_1, y_1) and (x_2, y_2) are members of the same *connected component* S_i of a set S if both are members of the set, and there exists a connected path between the two points that only includes members of the set.

Area open and *area close* are operators [18] that can be used to generate a scale space. For the on-set S, the area open operation is denoted by $S\hat{\circ}s$ and removes all connected components S_i with area (cardinality) less than s. Area close is the complementary operation: $S\hat{\bullet}s$ removes all connected components of area less than s in the off-set S^c. So, the area s is the scale parameter of area morphology, similar to the structuring element size in standard morphology. The shape of an equivalent structuring element for area open and close is not defined. Hence, the area operators are amorphous.

For images, the area open and close operators are implemented via stacking. In a threshold decomposition of the image $\boldsymbol{I}$, an associated level set $L(\boldsymbol{I}, t)$ is a set obtained by thresholding the image intensity: $(x, y) \in L(\boldsymbol{I}, t)$ if $I(x, y) \geq t$. For a discrete domain of K intensities $i \in \{0, \cdots, K-1\}$, $I(x, y) = \sum_{i=0}^{K-1} 1_{(x, y) \in L(\boldsymbol{I}, i)}$ where $1_{[\,]}$ is the set indicator function. The stacking operation also allows $\boldsymbol{I}$ to be defined by

$$I(x, y) = \max[t\colon (x, y) \in L(\boldsymbol{I}, t)]. \tag{1}$$

To implement the area open and close operators, each level set can be processed independently. Then, the result image can be reconstructed by stacking using (1). Within an area open operation on an image $\boldsymbol{I}\hat{\circ}s$ or area close operation on an image $\boldsymbol{I}\hat{\bullet}s$, the order of level sets processed does not affect the final result. An area open operator on an image will remove all connected components within the level sets $L(\boldsymbol{I}, t)$ of $\boldsymbol{I}$ that do not have a minimum area of s. Similarly, the area close operator will remove connected components of the complemented (off-set) level sets $L^c(\boldsymbol{I}, t)$ of $\boldsymbol{I}$ that do not possess the minimum area. In this way, area open "flattens" small bright objects and area close flattens small dark objects in an image.

The concatenation of the area open and close operators leads to area open–close (AOC) $\boldsymbol{I}\hat{\circ}s\hat{\bullet}s$ and area close–open (ACO)

$\boldsymbol{I}\hat{\bullet}s\hat{\circ}s$. These operators control the scale of both positive-going bright objects and negative-going dark objects. However, AOC operation is not identical to the ACO operation. As with the traditional open–close and close–open operators, the AOC and ACO are contrast-biased.

AOC and ACO are connected operators [18] that will either remove or preserve connected components within the level sets $L(\boldsymbol{I}, t)$ and the complemented level sets $L^c(\boldsymbol{I}, t)$. In this way, two points within the same connected component are treated equally in scale space generation. This property is important in image classification, where we want two pixels in the same object to be classified into the same class.

Connected regions of constant intensity are called *flat zones* [18]. The AOC and ACO operators increase the area of these flat zones in the image, while reducing the total number of flat zones. Thus, the area operators increase the region homogeneity of the image $\boldsymbol{I}$ as s is increased. Furthermore, the coarseness or fineness of an image can be quantified by the nesting of the flat zones. Given two scaled representations $\boldsymbol{I}_a$ and $\boldsymbol{I}_b$ of the same image, image $\boldsymbol{I}_b$ is coarser than $\boldsymbol{I}_a$ if each flat zone of $\boldsymbol{I}_a$ is a subset of a single flat zone (of equal or greater area) in the same position in $\boldsymbol{I}_b$.

For classification, the concept of *edges* is important. We define edges as connected component boundaries within the image level sets. Let the neighborhood of (x_1, y_1) be denoted by $N(x_1, y_1)$. On the discrete domain Ω (with rectangular tessellation), the neighborhoods are typically defined via 4-connectivity or 8-connectivity. A point in the on-set $L(\boldsymbol{I}, t)$ is defined as an edge point if at least one of its neighbors is a member of the off-set $L^c(\boldsymbol{I}, t)$ (for the same level t). Therefore, we say that an edge exists in the image at (x, y) if there exists an edge within one of the level sets at (x, y). These definitions are used in the analysis of the AOC and ACO scale spaces. Specifically, we need a definition of edge position to evaluate the edge localization properties of the AOC and ACO scale spaces.

B. AOC and ACO Scale Spaces

Let an image scale space be denoted by $\{\boldsymbol{I}\}$ where $\boldsymbol{I}_s$ is the image representation at scale s. $\{\boldsymbol{I}\}$ is a set of such scaled representations $\boldsymbol{I}_s$ defined on the discrete (image position) domain Ω and discrete scale domain $\Omega_s \subset \{0, 1, 2, 3, \cdots\}$. The scales $s(t)$ in the scale domain can be parameterized by t, such that $s(0)$ is the finest scale and $s(|\Omega_s| - 1)$ is the coarsest scale in the scale space. In this study, we create image scale spaces using AOC and ACO operators. The AOC scale space $\{\boldsymbol{I}\}$ is given by

$$\boldsymbol{I}_{s(t)} = \boldsymbol{I}_{s(t-1)}\hat{\circ}s(t)\hat{\bullet}s(t) \tag{2}$$

while the ACO scale space is

$$\boldsymbol{I}_{s(t)} = \boldsymbol{I}_{s(t-1)}\hat{\bullet}s(t)\hat{\circ}s(t). \tag{3}$$

In both cases, $s(0) = 0$ and $\boldsymbol{I}_0 = \boldsymbol{I}$, the original input image.

Morel and Solimini [12] list desirable properties for scale spaces that include fidelity, Euclidean invariance, causality and strong causality. Of the four properties, the fourth, strong causality, has been elusive for image scale spaces. We will discuss these properties with respect to the AOC and ACO scale spaces and explain the impact of the properties on image classification.

The fidelity property ensures that the finest scale of the scale space contains the original input signal. With the AOC and ACO scale spaces, we have $\boldsymbol{I}_0 = \boldsymbol{I}$ exactly. This guarantees that each scale space is unique for a given unique input $\boldsymbol{I}$, so that a classification procedure using the scale space is based on the unique input $\{\boldsymbol{I}\}$. If a finer representation than $\boldsymbol{I}$ existed in the scale space, members of the same flat zone in the original image could be classified into two different classes, which is undesirable.

Given a discrete domain scale space, we cannot claim Euclidean invariance (invariance to translation and rotation). Because the area morphology operators modify the image based on connected component area alone, the scale spaces are Euclidean invariant in the continuous-domain case, since translation and rotation do not distort connected component areas. In the discrete case (the case examined in this paper), the scale spaces are invariant to integer-valued translations. With rotation, distortions will occur due to discretization.

In classification, we seek to simplify the image and increase intraregion homogeneity through scale. A scale-generating process that introduces new features with increased scale would be counterproductive and would lead to artificial objects and possibly erroneous classification. Causality in scale infers that a coarse scale representation can be recreated from any finer scale representation. Formally stated, we say that $\{\boldsymbol{I}\}$ is causal if $\boldsymbol{I}_s$ depends only on $\boldsymbol{I}_r$ for $s \geq r$, $r \geq 0$, $r, s \in \Omega_s$. This property holds for the AOC and the ACO approach since $\boldsymbol{I}_s$ can be reproduced exactly from $\boldsymbol{I}_r$ by implementing (2) and (3), respectively.

We now examine the important property of strong causality, which guarantees the preservation of edge positions through scale.

Proposition 1 (Strong Causality): Given an edge at position (x_1, y_1) in $\boldsymbol{I}_s$, there exists an edge at (x_1, y_1) in $\boldsymbol{I}_r$ if $r \leq s$ within the AOC and ACO scale spaces.

Proof: Recall the definitions of edges within level sets and within images presented in Section II-A for the discrete domain case assumed in this paper. If there exists an edge at (x_1, y_1) at scale s, then (x_1, y_1) is a member of a connected component [in the level set $L(\boldsymbol{I}, t)$ for some level t] with a minimal area of s. Consider a neighboring pixel position $(x_2, y_2) \in N(x_1, y_1)$ on the other side of the edge, where (x_2, y_2) is a member of the connected component [in the complemented level set $L^c(\boldsymbol{I}, t)$] with a minimal area of s. The only way in which the edge at (x_1, y_1) can be removed is to remove either of these connected components entirely, since area open and area close are connected operators and do not partially remove a connected component. But, since the connected components have a minimal area of s, they will not be removed at scales $r \leq s$. Therefore, the AOC and ACO scale spaces maintain the strong causality property.

Because the AOC and ACO scale spaces possess the strong causality property, two important qualities are guaranteed. First, new edges and hence new boundaries will not be created with increased scale. Second, the position of edges does not drift through the scale space. In scale space classification, edge movement can lead to misclassified pixels near region

boundaries. It is important to state that existing scale spaces for images do not maintain the strong causality property. The Gaussian scale space [21] has increased edge localization error with increased scale. Morphological scale spaces using standard open and close filters [15] produce edge movement due to the shape of the structuring element. For example, morphological filtering of a rectangular object with a circular structuring element leads to corner rounding. Scale spaces based on anisotropic diffusion [16] cannot guarantee the position nor causality of edges in the image. As shown in [1], anisotropic diffusion can produce false edges at increased scales, such as "staircase" artifacts.

C. Scale Space Classification

Given a scale space $\{\boldsymbol{I}\}$, with intensity $I_s(x, y)$ at position (x, y) and scale s, we wish to track $I_s(x, y)$ through scale. For fixed (x, y), $I_s(x, y)$ defines a one-dimensional signal with independent variable s, denoted by $\boldsymbol{I}(x, y)$. This 1-D signal $\boldsymbol{I}(x, y)$ represents the scale space evolution of $I(x, y)$. When sampled at a discrete domain of scales Ω_S, we call $\boldsymbol{I}(x, y)$ the scale space vector at (x, y). The scale space classifier clusters pixels via the similarity between scale space vectors.

The assumption of the scale space classifier is that pixels are classified by intensity at a range of scales. Therefore, we implicitly assume that each pixel is a member of a particular *object* at each scale s. Then, $I_s(x, y)$ is the intensity of the object at location (x, y) and scale s. Here, we assume that objects are either *ascending* or *descending*, brighter or darker than the surrounding pixels. The scale of an object is defined as follows:

Definition 1 (Object Scale): The scale of an ascending object at location (x, y) is the area of the connected component in the image with intensity equal to or greater than $t = I(x, y)$, the intensity at (x, y):

$$s_A(x, y) = |A_t(x, y)| \tag{4}$$

where $|A_t(x, y)|$ is the cardinality of the connected component set $A_t(x, y)$, defined by

$$A_t(x, y) = \{(m, n): \exists P(x, y, m, n), \forall (a, b) \in P(x, y, m, n), I(a, b) \geq t\}. \tag{5}$$

Likewise, the scale of a descending object at location (x, y) is the area of the connected component with intensity equal to or less than t, the intensity at (x, y):

$$s_D(x, y) = |D_t(x, y)| \tag{6}$$

where

$$D_t(x, y) = \{(m, n): \exists P(x, y, m, n), \forall (a, b) \in P(x, y, m, n), I(a, b) \leq t\}. \tag{7}$$

In (5) and (7), $P(x, y, m, n)$ is a connected path between (x, y) and (m, n). In the discrete case assumed here, paths are defined by 4-connectivity or 8-connectivity. Given the definitions of object scale and connected components for the ascending and descending cases, we can define the objects in the image scale space:

Definition 2 (Ascending and Descending Objects): The object at position (x, y) and scale s is defined by $A_s(x, y)$ if the object is ascending and $D_s(x, y)$ if the object is descending. An object is ascending at (x, y) if $s_D(x, y) > s_A(x, y)$, whereas an object is descending if $s_A(x, y) > s_D(x, y)$. In the case of equality, the pixel at (x, y) is a boundary point between an ascending and descending object.

For intuition into the object and object scale definitions, consider a bright, thin spike in the image on a relatively smooth, dark background. The connected component containing the spike would be smaller, typically, than that containing the surrounding background. By Definitions 1 and 2, pixels in the spike would be members of an ascending object.

The definition of an object (Definition 2) is both scale and position dependent. Let $O_s(x, y)$ denote an object at (x, y) and scale s. Note that it is possible that $(x_1, y_1) \in O_s(x_2, y_2)$, but $(x_2, y_2) \notin O_s(x_1, y_1)$. The object memberships $(x_1, y_1) \in O_s(x_2, y_2)$ and $(x_2, y_2) \in O_s(x_1, y_1)$ hold only when $I_s(x_1, y_1) = \boldsymbol{I}_s(x_2, y_2)$.

To classify pixels based on the intensity of the associated object at scale s, the objects within the scaled image $\boldsymbol{I}_s$ should have area greater than or equal to s. Using Definitions 1–2 and the definitions of the AOC/ACO scale spaces, we find that this property holds. By definition, $\boldsymbol{I}_s$ is created by an AOC or ACO on the previous (finer) representation using scale parameter s. An AOC or ACO operation with area s removes all connected components within the image level sets of area less than s. By Definitions 1 and 2, any object in $\boldsymbol{I}_s$ will have an associated scale that is greater than or equal to s.

Intraregion homogeneity should increase with increased scale. If two pixels are members of the same object at a given scale, the two pixels should also be members of the same object at a coarser scale.

Proposition 2: If (x_1, y_1) and (x_2, y_2) are members of the same object $O_{s1}(x_1, y_1)$ at scale s_1, then at scale s_2, (x_1, y_1) and (x_2, y_2) are members of the object $O_{s2}(x_1, y_1)$ if $s_2 > s_1$.

Proof for AOC and ACO Scale Spaces: Since AOC and ACO are connected operators, the connected components within each level set and complemented level set are either preserved or removed in their entirety. Consider an ascending object $O_{s1}(x_1, y_1)$ at scale s_1. Because the object is ascending, $I_{s1}(x_2, y_2) \geq I_{s1}(x_1, y_1)$. If connected components within the level sets (or complemented level sets) of $\boldsymbol{I}_{s1}$ are removed to create $\boldsymbol{I}_{s2}$, the relationship $I_{s2}(x_2, y_2) \geq I_{s2}(x_1, y_1)$ is preserved. Hence, by Definitions 1 and 2, $(x_2, y_2) \in O_{s2}(x_1, y_1)$. For the case of a descending object, the argument is similar.

Another assumption of the scale space classifier is that pixels do not have a single scale. In other words, we cannot cluster pixels based on the two scalar features of intensity and scale. For example, consider a multiscale object such as a three-layer wedding cake. A pixel inside the third (highest) level, is also a member of the second layer at a coarser scale, and is a member of the base layer at a still coarser scale. Finally, every pixel is a member of the background, at an extreme (coarse) scale.

Now, let us analyze the clustering of scale space vectors. The distance between scale space vector $\boldsymbol{I}(x, y)$ and a cluster mean $\boldsymbol{\mu}(l)$ for class l is given by the following p-metric:

$$d(\boldsymbol{I}(x, y), \boldsymbol{\mu}(l)) = \left[\sum_{\Omega_s} |I_s(x, y) - \boldsymbol{\mu}_s(l)|^p\right]^{1/p}. \quad (8)$$

for $p \in (0, \infty]$. The particular p-metric used depends on the distribution of the intensities. For example, if the intensities within a given class have a Gaussian distribution, the $p = 2$ metric (the L_2 norm) should be used [17]. For a Laplacian distribution, $p = 1$.

A fundamental motivation behind the scale space classifier is that two points in the same object should not, in general, be classified into different classes. However, since all points are members of the same object at $s = \infty$, we cannot force members of the same object at any scale to be members of the same class. We can assert that the difference in intensities between two members of the same class at a given scale is decreasing with increased scale. Therefore, the two object members are becoming more tightly clustered with increasing scale, and the difference in distances to the nearest cluster center is decreasing.

In analyzing the performance of the scale space classifier, we compare the clustering of two points for the fixed scale classifier and the scale space classifier. With respect to a set of class mean vectors, if two points move closer together, they are more likely to classified into the same class. If two points move farther apart, they are less likely to end up in the same class. So, for pairs of points in the image at a fixed scale s, we can compare the fixed scale distance of

$$\begin{aligned} &d_s(I_s(x_1, y_1), I_s(x_2, y_2)) \\ &\quad = [|\Omega_s||I_s(x_1, y_1) - I_s(x_2, y_2)|^p]^{1/p} \end{aligned} \quad (9)$$

with the scale space distance of

$$\begin{aligned} &d(\boldsymbol{I}(x_1, y_1), \boldsymbol{I}(x_2, y_2)) \\ &\quad = \left[\sum_{\Omega_s} |I_s(x_1, y_1) - I_s(x_2, y_2)|^p\right]^{1/p}. \end{aligned} \quad (10)$$

The fixed scale distance is the distance at that scale multiplied by the cardinality of the set of scales $|\Omega_s|$. The scale space distance is the same form as (8), replacing the class mean vector with another scale space vector.

With the scale space classifier, we desire two characteristic improvements over the fixed scale classifier operating on the original image $\boldsymbol{I}_0 = \boldsymbol{I}$. These characteristics are formalized by Propositions 3 and 4:

Proposition 3: If two points are members of the same object, but have different intensities initially $(I_0(x_1, y_1) \neq I_0(x_2, y_2))$, then the area morphology based scale space classifier clusters the two pixels more closely than the fixed scale classifier. That is,

$$d_0(I_0(x_1, y_1), I_0(x_2, y_2)) > d(\boldsymbol{I}(x_1, y_1), \boldsymbol{I}(x_2, y_2)). \quad (11)$$

Proof: At scale $s = 0$, we have

$$|I_0(x_1, y_1) - I_0(x_2, y_2)| = |I_0(x_1, y_1) - I_0(x_2, y_2)|. \quad (12)$$

For any scale $s'' \geq s'$, using the AOC or ACO scale spaces, we have

$$|I_{s''}(x_1, y_1) - I_{s''}(x_2, y_2)| \leq |I_{s'}(x_1, y_1) - I_{s'}(x_2, y_2)|, \quad (13)$$

since AOC and ACO will remove the smaller connected components until $(x_1, y_1) \in O_{s''}(x_2, y_2)$ and $(x_2, y_2) \in O_{s''}(x_1, y_1)$. In fact, for $s'' \geq |O_{s'}(x_2, y_2)|$, we have $I_{s''}(x_1, y_1) = I_{s''}(x_2, y_2)$, thus reducing $|I_{s''}(x_1, y_1) - I_{s''}(x_2, y_2)|$ to zero. Combining the equality of (12) with the inequality of (13), along with the inspection of the definitions for distance in (9) and (10), we see that (11) is asserted.

Note that if the initial intensities are indeed equal in Proposition 3, the inequality of (11) is altered by replacing "$>$" with "$\geq$". Thus, in the case of equal initial intensities within the same object, the scale space classifier will cluster the two points at least as closely as the fixed scale classifier.

Proposition 4: If two points have the same initial intensity $(I_0(x_1, y_1) = I_0(x_2, y_2))$, but are members of different objects of different scales, then the fixed scale classifier clusters the two pixels more closely than the scale space classifier. That is,

$$d(\boldsymbol{I}(x_1, y_1), \boldsymbol{I}(x_2, y_2)) > d_0(I_0(x_1, y_1), I_0(x_2, y_2)). \quad (14)$$

Proof: Given that $I_0(x_1, y_1) = I_0(x_2, y_2)$, $d_0(I_0(x_1, y_1), I_0(x_2, y_2)) = 0$. Let the area of the object at (x_1, y_1) be defined as $|A_0(x_1, y_1)|$ and the area of the object at (x_2, y_2) be defined as $|A_0(x_2, y_2)|$. Then define $s' = \min\{|A_0(x_1, y_1)|, |A_0(x_2, y_2)|\}$. At scale $s'' = s' + 1$, we have $I_{s''}(x_1, y_1) \neq I_{s''}(x_2, y_2)$. Therefore, (14) is asserted.

Earlier, we stated that the scale space classifier reduces intraobject classification error. We now formally define this error measure.

Definition 3 (Intraobject Classification Error): For (x_1, y_1) and scale s, the intraobject classification error is defined by the cardinality of the set of members (x_2, y_2) of the object at (x_1, y_1) and scale s where (x_1, y_1) and (x_2, y_2) are members of different classes in the final classification:

$$\begin{aligned} e_s(x_1, y_1) = &|\{(x_2, y_2) : (x_2, y_2) \in O_s(x_1, y_1), \\ &(x_2, y_2) \in \gamma_{c_2}, (x_1, y_1) \in \gamma_{c_1}, c_1 \neq c_2\}|. \end{aligned}$$

where γ_{c_1} and γ_{c_2} are homogeneous sets in the final classification and c_1 and c_2 are class labels. The total normalized intraobject classification error is given by

$$e = \frac{\sum_{\Omega_s}\sum_{\Omega} e_s(x, y)}{|\Omega_s||\Omega|}. \quad (15)$$

With the scale space classifier, the associated intraobject classification error is less than or equal to the same error using the original image or any one scale. The scale space classifier attempts to minimize the distance between pixels in the same object through scale. As shown in the proof of Proposition 3, the difference between two pixel intensities in the same object will decrease monotonically until equal. Hence, as members of the same object are clustered more tightly (relative to nonmembers), the intraobject classification error is reduced. In Section IV, results are given that demonstrate the reduction in intraobject classification error.

D. Sampling Scale Space

In the discrete case, we need to sample the scale space appropriately. Since the scale space $\{\boldsymbol{I}\}$ is a three-dimensional space, sampling (and quantization) are performed in each dimension. Typically, the generation of Ω, the sampled (x, y) domain, is fixed by the initial digitization of the image. Sampling the scale domain Ω_s is an open problem.

1) Minimum and Maximum Scale: The absolute minimum scale for any image scale space is $s = 0$. Note that in the discrete case, images at scales $s = 0$ and $s = 1$ are identical, since every existing connected component has a minimum area of one. So, $\boldsymbol{I}_1 = \boldsymbol{I}_0$. The upper limit of scale is theoretically infinite; however, a discrete-domain image with $|\Omega|$ pixels has a maximum scale of $s = |\Omega|$. For the classification problem with C classes, there should be at least one object for each class. Therefore, the maximum scale is $|\Omega|/C$ in this case.

Given *a priori* information about the scale of objects in an image, additional bounds on the scales used in classification can reduce the computational complexity of the scale space classification approach. For example, in a classification of blood cells, we would know the minimum area of a red cell and the maximum area of a white cell. Scales beneath the minimum scale $s_{\min}$ and above the maximum scale $s_{\max}$ could only add error to the classification process.

2) Sampling Intervals in Scale: For a discrete sampling of scales, the most obvious sampling would include scales $s_{\min}$, $s_{\min} + 1$, $\cdots$, $s_{\max} - 1$, $s_{\max}$. In this framework, the scale operator at scale s would reject connected components of area less than s pixels, and the next scale would reject all connected components with area less than $s + 1$ pixels. Therefore, given the discretization of the image spatial domain, each potential scale would be represented (between the minimum and maximum scales, inclusive).

Nevertheless, the straightforward sampling of the image scale may be an example of over-sampling. Consider an application in which the classification of coins is desired (as in Fig. 1). Since there are only a few different coin sizes, the classification process only requires a few scale samples. So, there seems to be a set of important scales in a particular scale space for this application. Choosing the scales in an *ad hoc* fashion is possible, but we seek an automated method to sample scale.

The idea behind our sampling scheme for the scale parameter is based on *granulometry*. For an AOC operation or an ACO operation, we can track the differences in a given parameter (such as total summed intensity) between successive scales. Define the normalized granulometry for the scale space $\{\boldsymbol{I}\}$ by

$$G(s) = \frac{1}{|\Omega|} \sum_{\Omega} |I_{s+1}(x, y) - I_s(x, y)|. \tag{16}$$

Given a series of scales $s_{\min}$, $s_{\min}+1$, $\cdots$, $s_{\max}-1$, this feature tracking provides a granulometry $G(s)$ that reveals information about the image structure. When a large object disappears in the scale space (when it becomes part of a larger flat zone), a local maximum in the granulometry may be observed. Thus, the scale at which the object is the most simplified is the scale that precedes removal of the object. So, a sampling scheme for scale can be based upon locating the maxima in $G(s)$ and using these scales in the classification process. Additionally, the maxima may be filtered based upon the magnitude in change. For example, at scale s, a maximal point that only represents a magnitude change in $G(s)$ of less than sT would be rejected as a sample point. Here, T represents the minimum allowed intensity change between classes. For C classes and K discrete intensities, we can assume $T \approx K/C$.

For the coin classification application of Fig. 1(a), the normalized granulometry using (16) is shown in Fig. 2. To facilitate the location of local maxima, we convert the discrete-domain granulometry to a continuous-domain signal using a polynomial fit, as shown in Fig. 3. In the granulometry, the two major maxima, at scales $s = 670$ and $s = 292$, correspond to the area of the large coins and the area of the small coins, respectively.

In summary, the area open–close and close–open operators are used here to create a scale space. The AOC and ACO scale spaces provide fidelity, Euclidean invariance, causality, and strong causality for edge location through scale in contrast to other scale spaces generated with linear filters or standard morphology. The AOC and ACO scale spaces allow a meaningful clustering that can be exploited in image classification. Finally, efficient implementations of scale space classification can be achieved by sampling the scale space.

To achieve a classification from an AOC or ACO scale space, a clustering technique must be applied. In the next section, two fundamental clustering methods, fuzzy c-means and k-means, are reviewed. Also, the Bayesian clustering method is discussed in order to provide another comparison to the scale space classification approach.

III. Clustering Algorithms

The objective of scale space classification is to group scale space vectors based on a similarity measure. Standard minimum-distance based classifiers are used for this purpose. The possibility of a single pixel belonging to different objects at various scales suggests the use of an unsupervised fuzzy c-means classifier. Hard classifiers such as the unsupervised k-means algorithm can then be posed as a special case of fuzzy c-means classifier. A brief overview of these classification schemes is provided here, as the classifiers are used within the scale space classification paradigm. As a comparative example, a parametric Bayesian classifier that has recently been used for multiscale clustering is also discussed.

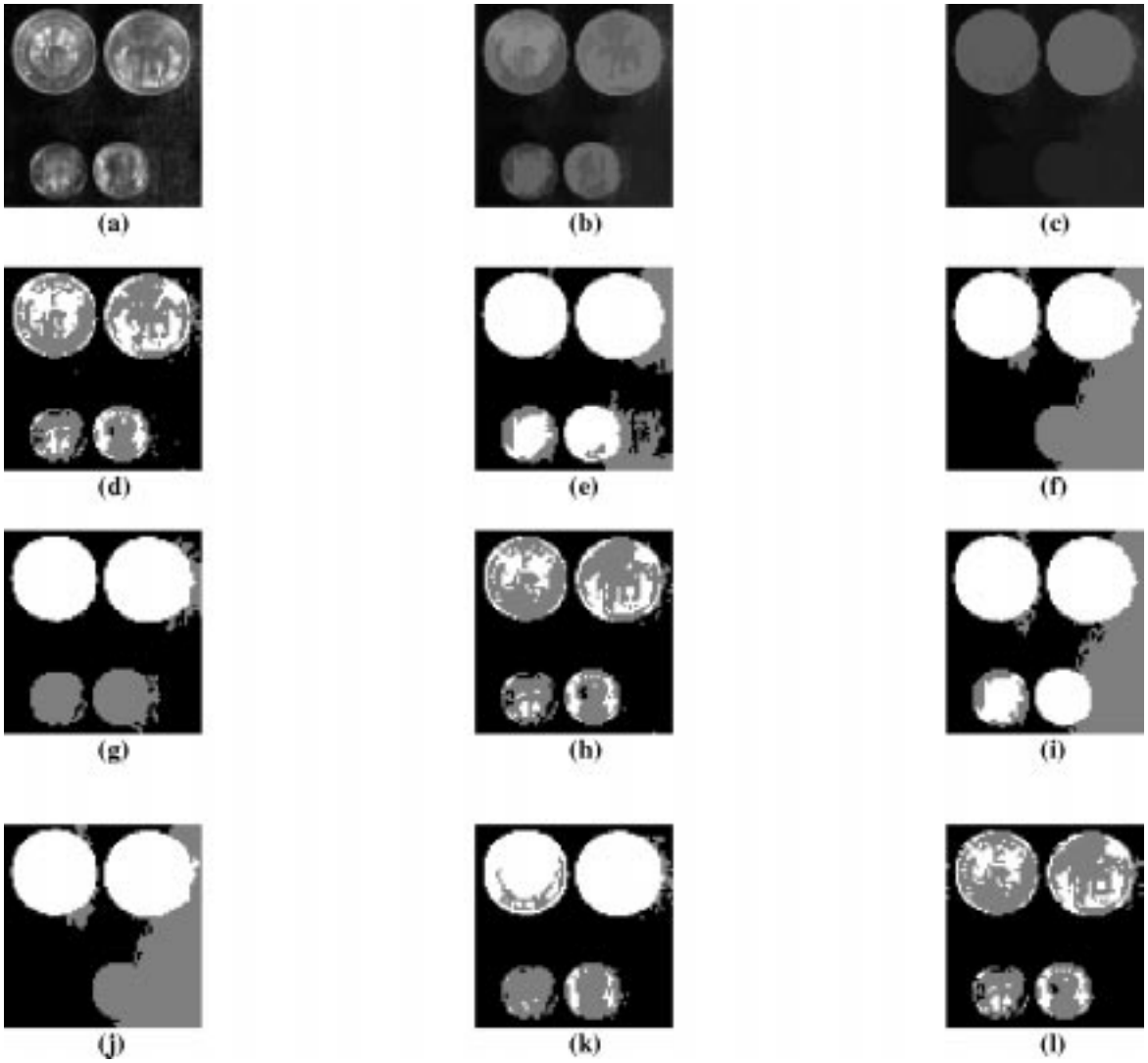

Fig. 1. (a) Original "coins" image; (b) AOC scale space image at $s = 100$; (c) AOC scale space image at $s = 500$; (d) 3-class FCM classification of (a); (e) 3-class FCM classification of (b); (f) 3-class FCM classification of (c); (g) 3-class scale space classification of (a). The scale space is constructed with (a)–(c). The scale space vectors are classified using FCM, (h) 3-class k-means classification of (a), (i) 3-class k-means classification of (b), (j) 3-class k-means classification of (c), (k) 3-class scale space classification of (a). As with (g), the scale space is constructed with (a)–(c). Here, the clustering scheme employed is k-means, and (l) 3-class Bayesian classification of (a).

A. Fuzzy c-means Classification

Within the fuzzy c-means clustering algorithm, the familiar least-squared error criterion is applied [2]:

$$J_m(U, \boldsymbol{\mu}) = \sum_{\Omega} \sum_{i=1}^{C} (u_i(x, y))^m \|d_i(x, y)\|^2. \tag{17}$$

Here, U is the fuzzy C-class partition of scale space. Given a scale space vector $\boldsymbol{I}(x, y)$ at location (x, y), the measure $\|d_i(x, y)\| = \|\boldsymbol{I}(x, y) - \boldsymbol{\mu}_i\|$ is the distance between the scale space vector and the ith cluster center $\boldsymbol{\mu}_i$. The distance is weighed by the fuzzy membership value of each scale space vector $u_i(x, y)$ corresponding to ith class. The fuzzy exponent m has the range $m \in [1, \infty]$. For every feature vector $\boldsymbol{I}(x, y)$, the error criterion $J_m(U, \boldsymbol{\mu})$ is minimized subject to the conditions

$$\sum_{i=1}^{l} u_i(x, y) = 1,\ 0 < \sum_{\Omega} u_i(x, y) < |\Omega| \text{ and } u_i(x, y) \geq 0.$$

For the scale space vector at (x, y), the fuzzy membership is updated according to

$$u_i(x, y) = 1 \Bigg/ \left[\sum_{j=1}^{C} \left(\frac{d_i(x, y)}{d_j(x, y)} \right)^{2/(m-1)} \right]. \tag{18}$$

The initial fuzzy membership value is generated using uniformly-distributed random number generator. For the classified image, the cluster center is updated for all the classes at each iteration according to

$$\boldsymbol{\mu}_i = \frac{\sum_{\Omega} (u_i(x, y))^m \boldsymbol{I}(x, y)}{\sum_{\Omega} (u_i(x, y))^m}. \tag{19}$$

The algorithm is terminated for insignificant (1–2% of the current value) changes in $\boldsymbol{\mu}_i$ between consecutive iterations. To reduce processing time, the final classified image is obtained

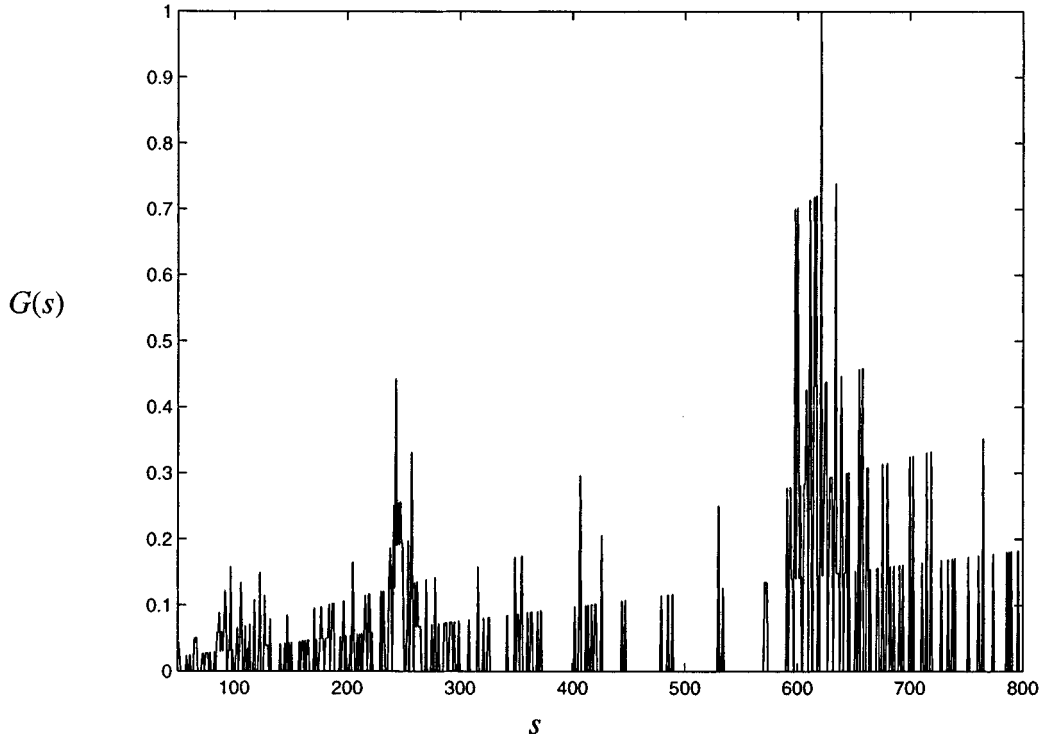

Fig. 2. Normalized granulometry for the "coins" image.

byclassifying the pixels based highest class membership value for each pixel.

B. K-Means Classification

In the case of k-means, a so-called *hard* classifier, the feature domain is not partitioned into fuzzy classes. For a fixed number of clusters, k-means minimizes the squared error function

$$\varepsilon^2 = \sum_{\Omega} \sum_{i=1}^{C} \|d_i(x, y)\|^2. \tag{20}$$

An initial seed vector representing cluster center is arbitrarily specified for each class. The scale space vectors (or pixel intensities in the fixed scale case) are classified based on the minimum Euclidean distance from the cluster centers. A new set of cluster centers is then calculated from this reassignment. The iterative process continues until the net migration of cluster centers at jth iteration, $\Delta\boldsymbol{\mu}(j)$, is insignificant compared to that of the previous iteration, where

$$\Delta\boldsymbol{\mu}(j) = \sum_{i=1}^{l} |\boldsymbol{\mu}_i(j) - \boldsymbol{\mu}_{i-1}(j-1)|. \tag{21}$$

C. Bayesian Classification

With the Bayesian classifier, an ideal partition of the image space is considered as a collection of regions of uniform or slowly varying intensity. In this case, the image intensity is modeled by a Markov random field. If the segmentation of the image is denoted by S, then the *a priori* probability density of the region segmentation S is given by a Gibbs density [14]

$$p(S) = \frac{1}{Z} \exp\left\{ -\sum_{Q} V_q(x, y) \right\} \tag{22}$$

where Z is a normalizing constant. V_q is the weight proportional to the homogeneity of a neighborhood, also described using pixel cliques Q, around the pixel (x, y).

The conditional density of each observed region is modeled as a white Gaussian process of mean signal intensityμ and noise with variance σ^2. Each region i is characterized by mean intensity μ_i. If $p_i(S(x, y)|\boldsymbol{I})$ gives the conditional probability of region i at (x, y) given each observed pixel intensity $I(x, y)$, then

$$p_i(S(x, y)|\boldsymbol{I}) \propto \exp\left\{ -\frac{1}{2\sigma^2} |I(x, y) - \mu_i|^2 \right\}. \tag{23}$$

Therefore the combined probability density has the form

$$p_i(S(x, y)|\boldsymbol{I}, Q) \propto \exp\left\{ -\frac{1}{2\sigma^2} |I(x, y) - \mu_i|^2 - \sum_{Q} V_q(x, y) \right\}. \tag{24}$$

The implementation of (24) consists of iterative maximization of p_i, followed by updating of μ_i using iterated conditional modes (ICM). The ICM algorithm suffers from dependence on the selection of initial estimates and from suboptimal local minima in the cost function. Specific implementation issues of the Bayesian classifier are discussed in Section IV where the results are compared with those of scale space classifier.

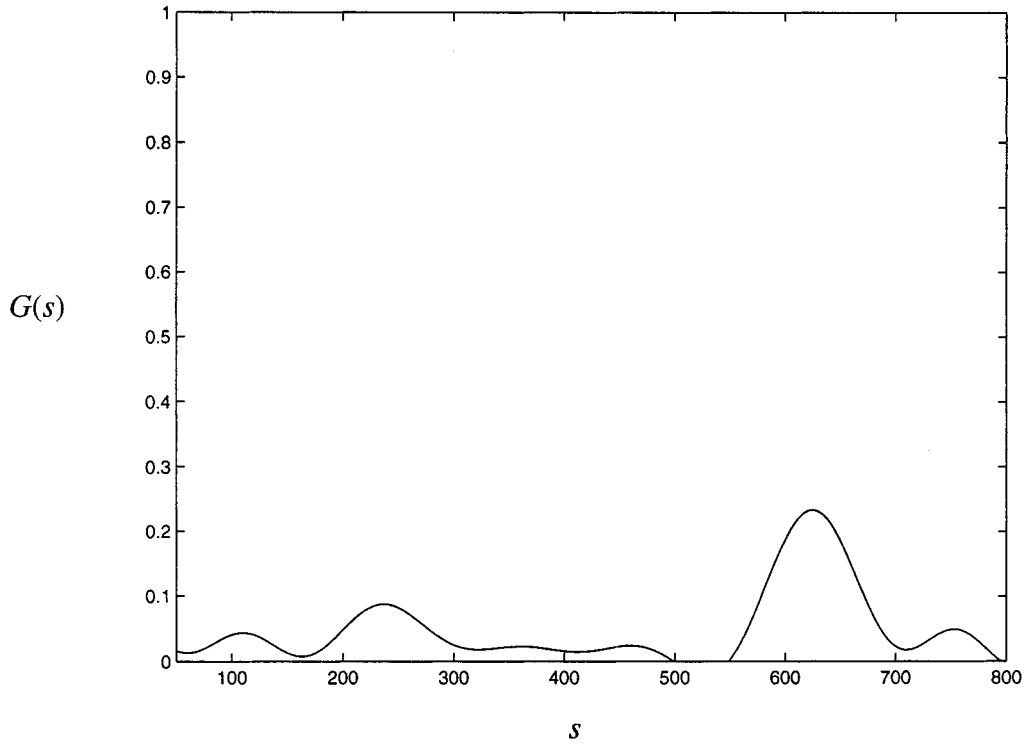

Fig. 3. Polynomial fit of the granulometry for the "coins" image.

IV. Results and Discussion

In this section, results from standard classifiers and from the scale space classifier are compared. Note that we emphasize classification applications in which object scale is important, so as to highlight the benefits of the scale space derived from area morphology. The results are given in the form of classified images and tabulated data. Table I provides the traditional measure of classification accuracy, and Table II gives the intraobject classification error (Definition 3). Four real image examples including two biomedical applications are supplied.

Fig. 1(a) gives the "coins" image with the corresponding AOC scale space images shown in Fig. 1(b) and (c) for scales 100 and 500, respectively. This image contains four coins of two different sizes and of similar intensity. The objective of extracting the larger and smaller coins separately as two different classes highlights the advantages of the scale space classifier.

Fig. 1(d)–(f) are the results from fuzzy c-means (FCM) classification of Fig. 1(a)–(c), respectively. Notice that the integrity of the coins is sacrificed in the FCM results, and that merging with the background occurs. In contrast, the scale space classification, shown in Fig. 1(g), correctly labels the two smaller coins as one class and the two larger coins as another class. The classification accuracy (Table I) for the scale space classifier is almost 20% higher than that of all fixed scale FCM results. As an aside, note that all FCM implementations here use a fuzzy exponent $m = 2$.

The scale space classification result using k-means [Fig. 1(k)] is acceptable but does not match the high quality result shown in Fig. 1(g). Fig. 1(h)–(j) are the results of k-means classification of Fig. 1(a)–(c), respectively. The scale space classifier outperforms the k-means classification at any fixed scale in both classification accuracy (Table I) and intraobject classification error (Table II). Fig. 1(l) displays the 3-class result from the Bayesian classifier. The Bayesian classifier is not able to match the performance of the scale space classifier in visual quality, classification accuracy, or intraobject classification error. For the implementation of the Bayesian classifier in this paper, the pixel clique is defined as the 8-connected neighborhood, and the initial estimates are supplied from the classification result of fuzzy c-means.

TABLE I
Comparison of Classification Accuracy (SSFCMC: Scale Space Classification using Fuzzy c-Means, SSKMC: Scale Space Classification using k-Means, FMC: Fuzzy c-Means, KM: k-Means, and BC: Bayesian Classification)

Example	SSFCMC	SSKMC	FCM	KM	BC
"Coins"	73.83%	71.27%	55.72%	53.84%	53.47%
"Blood Cells"	96.21%	99.55%	75.22%	75.22%	76.50%
"Plasma"	92.96%	90.71%	82.04%	83.21%	82.58%
"Connecting rod"	99.48%	99.42%	97.48%	97.54%	97.49%
"Connecting rod" image with noise	99.38%	99.34%	93.1%	92.6%	97.28%

TABLE II
Intraobject Classification Error (Refer to Definition 3)

Example	SSFCMC	SSKMC	FCM	KM	BC
"Coins"	133.58	139.64	171.62	167.77	168.10
"Blood Cells"	125.5	115.95	174.74	174.74	173.84
"Plasma"	365.55	399.84	416.30	419.95	463.93
"Connecting rod"	122.1	120.55	159.98	158.99	143.23
"Connecting rod" image with noise	169.98	168.78	211.7	209.53	188.89

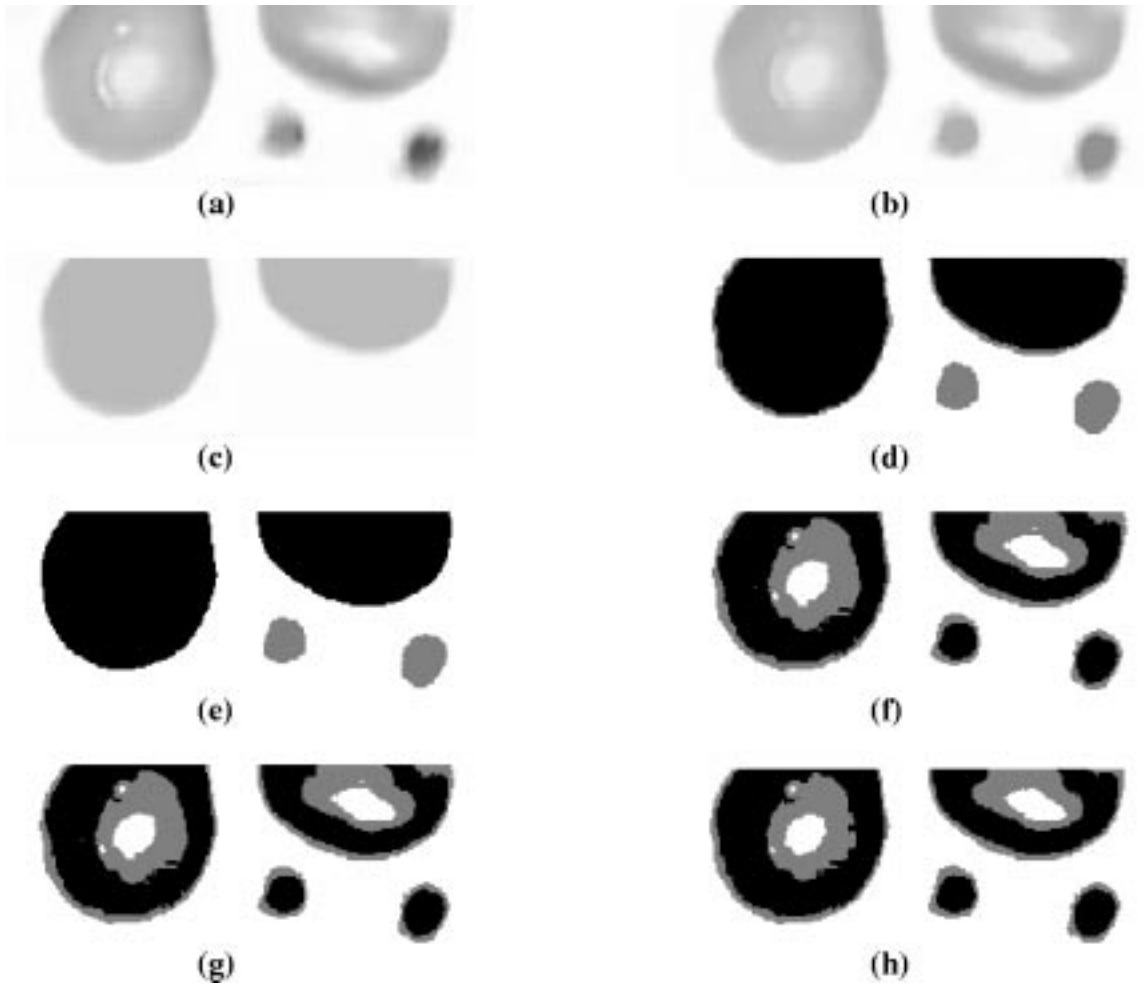

Fig. 4. (a) Original "blood cells" image; (b) AOC scale space image at $s = 200$; (c) AOC scale space image at $s = 2000$; (d) result of scale space classifier (3-class) using (b) and (c). Note that the original image is not included in the scale space, and the FCM clustering technique is used; (e) result of k-means scale space classifier (3-class) using (b) and (c); (f) 3-class FCM classification of (a); (g) 3-class k-means classification of (a); and (h) 3-class Bayesian classification of (a).

The "blood cells" image shown in Fig. 4(a) could be used to label and count different blood cells (i.e., red cells and white cells). Fig. 4(b) and (c) are the AOC scale space images sampled at scales 200 and 2000, respectively. Again, the scale space classification result of Fig. 4(d) (with FCM) and 4(e) (with k-means) are visually superior compared to the classifications of the original image as shown in Fig. 4(f)–(h). Note that the blood cells contain shiny spots of considerable area, probably due to specular reflection. However, with the area morphology based processing, these insignificant regions are removed at higher scales. The scale space classifier using k-means provides a classification accuracy of 99.55%. The inferiority of the fixed scale classifiers is evident in the spurious classes generated within the four cells. The standard FCM, k-means, and Bayesian classifiers essentially create false objects within each cell, which can be observed in the classified images and the 58% increase in intraobject classification error (see Table II).

The performance of scale space classifier is tested with another microscopic biomedical image shown in Fig. 5(a). This image is particularly difficult for classification due to irregular shapes and nonuniform illumination. The scale space classifier results in Fig. 5(d) (FCM scale space classifier) and 5(e) (k-means scale space classifier) yield an amazing 90+% accuracy. Fig. 5(f)–(h) show the failure of the fixed scale classifiers to preserve the integrity of the objects in the scene and the background. It is noteworthy that in all examples provided in the paper, the scale space classifiers give the highest classification accuracy and the lowest intraobject classification error.

Finally, we provide industrial inspection application of a mechanical part in Fig. 6(a). The specularities and intensity variations in the "connecting rod" image make the classification task challenging. A 2-class classification using the FCM classification technique is shown in Fig. 6(b). Note the gaps in the interior that preclude object/background separation. For example of the performance of the classifiers in the presence of noise, 5% salt and pepper noise is added to the image, as shown in Fig. 6(c), and the corresponding FCM classified result is shown in Fig. 6(d). AOC-scaled versions are shown in Fig. 6(e) and (g), and Fig. 6(f) and (h) are the corresponding scale space classified results. We notice that the scale space classification produces a cohesive region containing the connecting rod from both the uncorrupted and corrupted imagery. The comparison is extended using

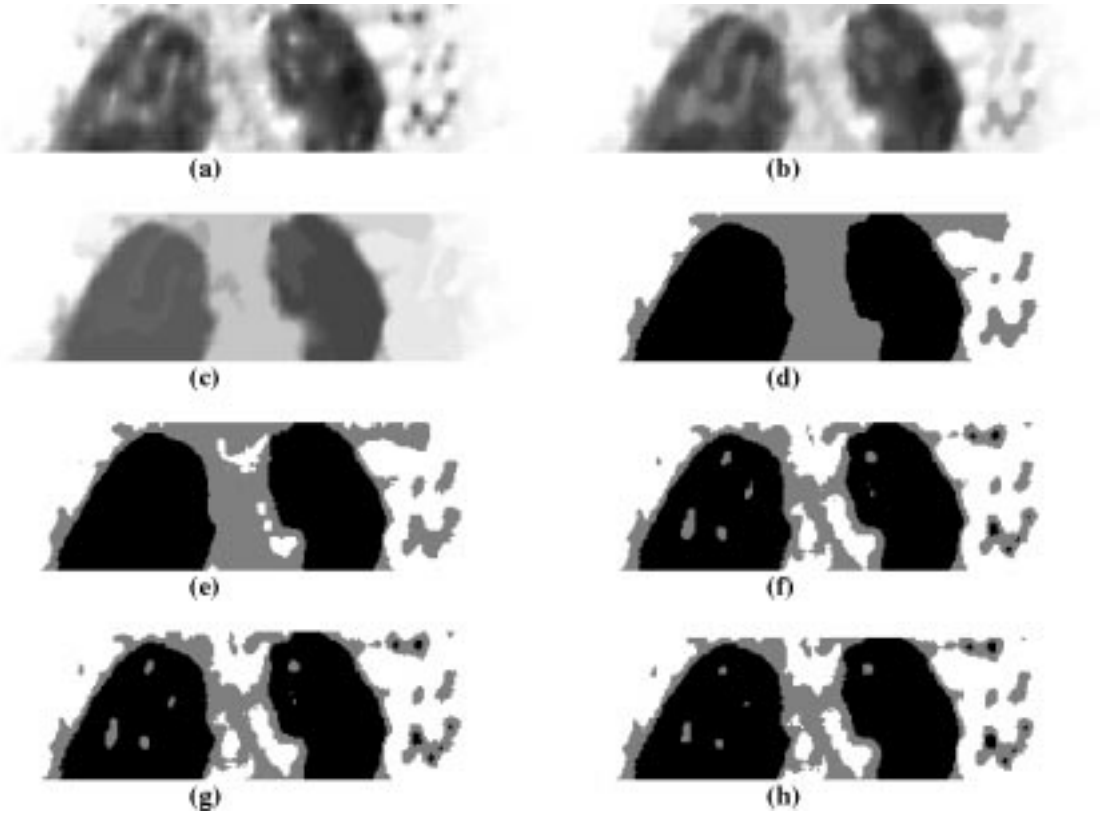

Fig. 5. (a) Original "plasma" image; (b) AOC scale spacing image at $s = 50$; (c) AOC scale space image at $s = 1500$; (d) 3-class FCM scale space classifier result where the scale space is constructed with (a)–(c); (e) 3-class k-means scale space classifier result with scale space as in (d); (f) 3-class FCM classification of (a); (g) 3-class k-means classification of (a); and (h) 3-class Bayesian classification of (a).

results of k-means and scale space k-means classifiers. Fig. 6(i) and (j) are the results of the k-means and scale space k-means on Fig. 6(a) and (e), while Fig. 6(k) and (l) are resultant from Fig. 6(c) and (g), respectively. Here, the area morphology based scale space operator is able to extract the overall shape of the object without any additional post-processing. In this example, the Bayesian classification method also is successful (see Fig. 6(m) and (n)). Because the connecting rod example does not involve classification according to object scale, the k-means, c-means and Bayesian methods yield high classification accuracy, as shown in Table I.

Although the results demonstrate the efficacy of scale space classifier, we must also note that the computational burden is increased due to scale space generation. In the scale space classifier, the classification process itself does not impose any additional computation in clustering. For an $N \times N$ image to be classified to C clusters requires $O(N^2C)$ comparisons. In case of the scale space classifier, the comparisons are vector distance measures where the length of vector depends on the number of scale samples in the scale space. We have shown that the image granulometry can be exploited to limit the number of scales utilized and to thus limit the computational overhead. Typically, to generate one AOC/ACO scale space image via level set analysis, $O(KN^2)$ comparisons are required where K is the number of gray levels (256 for 8-bit imagery). More efficient queue-based implementations of the area operators have been reported in [20], in which standard open and close filters are used as "marker images" that determine which connected components (of the image level sets) are retained in the filtering process. The marker images are then reconstructed (using geodesic dilation) to form an approximation of an area open or area close operation. This queue-based reconstruction step is $O(N^2)$.

We have implemented the area morphology operations and the clustering algorithms on a Sun Ultra 10 with 256 MB of RAM. The AOC/ACO operations were programmed using the interpretative Matlab 5, while the clustering step was written in the more expeditious C language. Using the level set analysis implementation of (2), the AOC operation on the 69x67 "coins" image required 49 s and 19 s for Fig. 1(b) and (c), respectively. With Vincent's queue-based approximation of the AOC operation [20], only 2 s and 1 s, respectively, were used to generate the same scale space images. With the 90×234 "blood cells" image, Fig. 4(b) and (c) used 65 s and 56 s, respectively, of processing time, while the fast algorithm used 2 s for each. The scaled images for the 74×276 "plasma" example needed 102 s [Fig. 5(b)] and 72 s [Fig. 5(c)] using level set analysis and just 2 s using the fast algorithm. Note that, in each case, the image at the higher scale is generated from that of the immediately lower scale. In case of the 120×120 image in Fig. 6(e), the time used to produce the scaled image is 83 s, and the cost is 210 s for Fig. 6(g); these times improve to 3 s

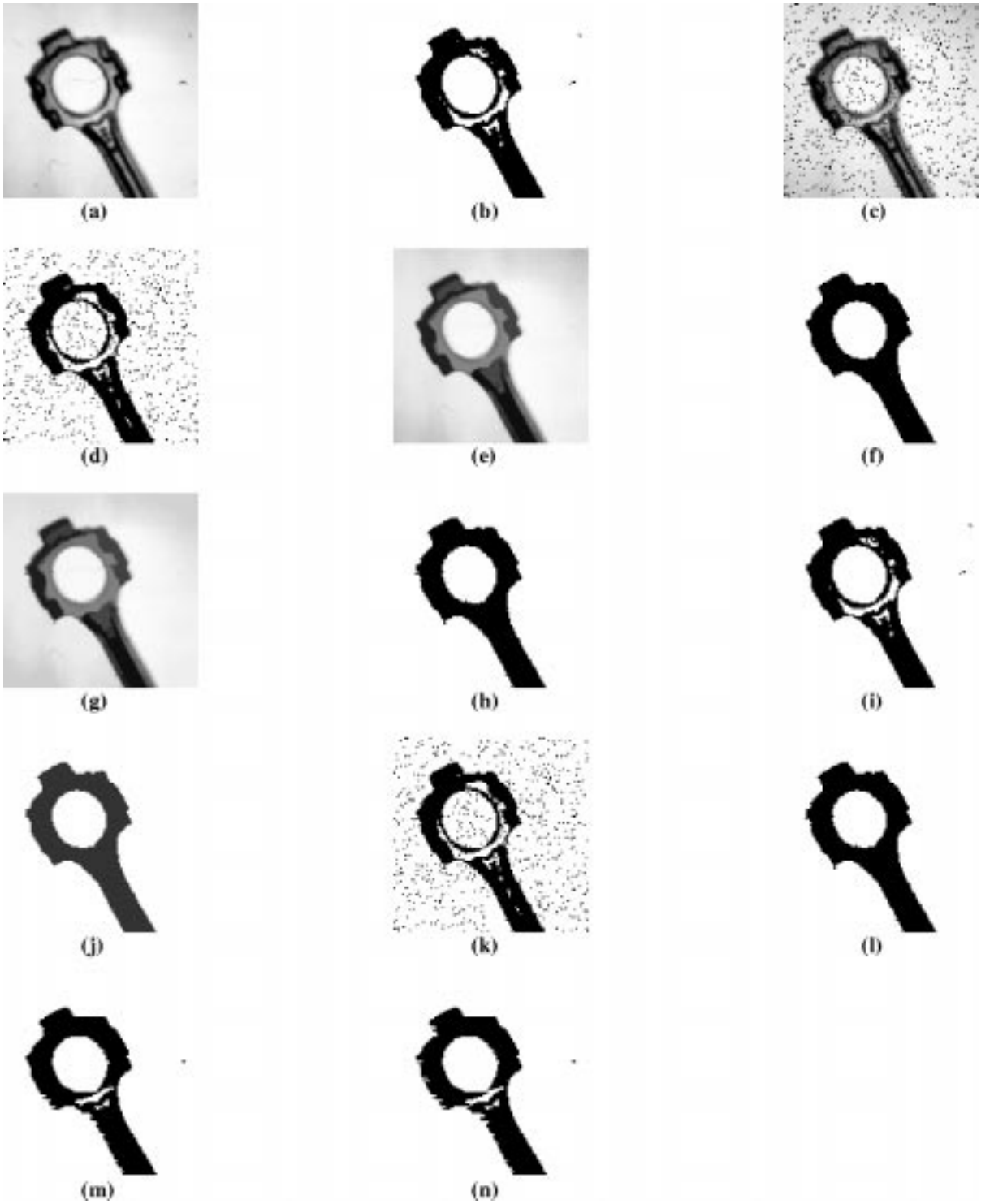

Fig. 6. (a) Original "connecting rod" image; (b) 2-class FCM classification of (a); (c) noisy version of (a) with 5% salt and pepper noise; (d) 2-class FCM classification of (c); (e)AOC scale space image of (a) at $s = 300$; (f) 2-class scale space FCM classification of (e); (g) AOC scale space image of (c) at $s = 300$; (h) 2-class scale space FCM classification of (g); (i) 2-class k-means classificaion of original image of (a); (j) 2-class k-means scale space classification of (a) at $s = 300$; (k) 2-class k-means classification of (c); (l) 2-class k-means classification of (g); (m) 2-class Bayesian classification of (a); and (n) 2-class Bayesian classification of (c).

and 5 s, respectively, with the fast algorithm. For the smaller "coins" image, the scale space classification using both fuzzy c-means and k-means requires less than 1 s, whereas for the rest of the examples, the classification step requires less than 2 s on the Ultra 10.

V. Conclusion

The paper describes an approach to area morphology-based multiscale classification in which vectors in scale space are clustered. In this framework, features from multiple scales are used

simultaneously, in contrast to existing hierarchical approaches. The AOC and ACO operators provide a well-motivated scale space for classification wherein the pixels belonging to the same object are clustered more tightly. The results are observed both in the image examples as well as the tabulated classification accuracy and intraobject classification error results. In images with objects of similar intensity and differing scale, the scale space classifier is able to distinguish between the objects where the traditional fixed scale classifier fails and produces classification errors. The results support the conclusion that the scale space approach is superior to the fixed scale approach when object scale is important in the classification task at hand.

The development of the AOC and ACO scale spaces is in itself an important contribution. The scale spaces are unique for each given input image and satisfy desirable properties that are useful for scale space generation and classification. Through scale, the AOC and ACO operators do not introduce new features such as additional regions or edges. Furthermore, the AOC and ACO scale spaces possess the strong causality property, ensuring the edges do not drift with increased scale. Assertion of the strong causality property differentiates the AOC and ACO scale spaces from existing image scale spaces generated via linear and nonlinear filtering. Therefore, the AOC and ACO scale spaces may prove to be valuable multiscale structures in applications beyond classification, such as content based retrieval and segmentation for video coding.

In future work, we wish to extend the scale space classifier to color imagery and multispectral imagery. The multispectral classification process could be applied to important problems in biomedicine and remote sensing.

REFERENCES

[1] S. T. Acton, "Multigrid anisotropic diffusion," *IEEE Trans. Image Processing*, vol. 7, pp. 280–291, Mar. 1998.

[2] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum, 1981.

[3] C. A. Bouman and B. Liu, "Multiple resolution segmentation of textured images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 99–113, Feb. 1991.

[4] C. A. Bouman and M. Shapiro, "A multiple random field model for Bayesian image segmentation," *IEEE Trans. Image Processing*, vol. 3, pp. 162–177, Mar. 1994.

[5] K. B. Eom and R. Chellappa, "Noncooperative target classification using hierarchical modeling of high-range resolution radar signatures," *IEEE Trans. Signal Processing*, vol. 45, pp. 2318–2327, Sept. 1997.

[6] J. M. Gauch, "Image segmentation and analysis via multiscale gradient watershed hierarchies," *IEEE Trans. Image Processing*, vol. 8, pp. 69–87, Jan. 1999.

[7] J. M. Gauch and S. M. Pizer, "Multiresolution analysis of ridges and valleys in grey-scale images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 635–646, June 1993.

[8] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 532–550, July 1987.

[9] P. T. Jackway, "Gradient watersheds in morphological scale space," *IEEE Trans. Image Processing*, vol. 5, pp. 913–921, June 1996.

[10] B. C. Lovell and A. P. Bradley, "The multiscale classifier," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 124–137, Feb. 1996.

[11] P. Maragos, "Pattern spectrum and multiscale shape representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 701–716, July 1989.

[12] J.-M. Morel and S. Solimini, *Variational Methods in Image Segmentation*. Boston, MA: Birkhauser, 1995.

[13] L. Najman and M. Schmitt, "Geodesic saliency of watershed contours and hierarchical segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 1163–1173, Dec. 1996.

[14] T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Trans. Signal Processing*, vol. 40, pp. 901–914, Apr. 1992.

[15] K. Park and C. Lee, "Scale-space using mathematical morphology," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 1121–1126, Nov. 1996.

[16] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans Pattern Anal. Machine Intell.*, vol. 12, pp. 629–639, July 1990.

[17] A. Restrepo and A. C. Bovik, "On the statistical optimality of locally monotonic regression," *IEEE Trans. Signal Processing*, vol. 42, pp. 1548–1550, 1994.

[18] P. Salembier and J. Serra, "Flat zones filtering, connected operators, and filters by reconstruction," *IEEE Trans. Image Processing*, vol. 4, pp. 1153–1160, Aug. 1995.

[19] L. Vincent, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 583–598, June 1991.

[20] ——, "Morphological gray scale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Processing*, vol. 2, pp. 176–201, Apr. 1993.

[21] A. P. Witkin, "Scale-space filtering," in *Proc. Int. Joint Conf. Artificial Intelligence*, 1983, pp. 1019–1021.

[22] A. Wright and S. T. Acton, "Watershed pyramids for edge detection," in *Proc. IEEE Int. Conf. Image Processing*, Santa Barbara, CA, October 26–29, 1997.

Scott T. Acton (S'89–M'93–SM'99) received the B.S. degree in electrical engineering from Virginia Tech, Blacksburg, in 1988, and the M.S. and Ph.D. degrees in electrical engineering from the University of Texas, Austin in 1990 and 1993, respectively.

He has been with AT&T, the MITRE Corporation, and Motorola, Inc. Currently, he is an Associate Professor with the School of Electrical and Computer Engineering, Oklahoma State University (OSU), Stillwater, where he directs the Oklahoma Imaging Laboratory, which is sponsored by several organizations including NASA, the Army Research Office (ARO), and Lucent Technologies. His research interests include multiscale image representations, diffusion algorithms, image morphology, image restoration, image segmentation, and content-based retrieval.

Dr. Acton is the winner of the 1996 Eta Kappa Nu Outstanding Young Electrical Engineer Award, a national award that has been given anually since 1936. At OSU, he has been selected as the 1997 Halliburton Outstanding Young Faculty Member. He serves as Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING and is a member of SPIE and Eta Kappa Nu.

Dipti Prasad Mukherjee received the B.E. degree from Jadavpur University, Calcutta, India, in 1985, the M.S. degree from the University of Saskatchewan, Saskatoon, Sask., Canada, in 1989, and the Ph.D. degree from from Indian Statistical Institute (ISI), Calcutta, in 1996.

He is currently a Faculty Member with the Electronics and Communication Sciences Unit, ISI. He was a Visiting Assistant Professor with the Oklahoma Imaging Laboratory, School of Electrical and Computer Engineering, Oklahoma State University, in 1998-99. He was UNDP Fellow with the Robotics Research Group, University of Oxford, Oxford, U.K., in 1992. His research interests are in the areas of computer vision and graphics. He has published 20 peer-reviewed journal papers and is the author of a textbook on computer graphics and multimedia.

Dr. Mukherjee was the recipient of UNESCO-CIMPA fellowships to INRIA, France in 1991, 1993, and 1995.