March, 1997

LIDS-P 2385

Research Supported By:

DARPA contract FA49620-93-1-0604

Scale-Based Robust Image Segmentation

Kim, A.
Pollak, I.
Krim, H.
Willsky, A.S.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAR 1997		**00-03-1997 to 00-03-1997**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Scale-Based Robust Image Segmentation	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Massachusetts Institute of Technology,77 Massachusetts Avenue,Cambridge,MA,02139-4307**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**7**	

SCALE-BASED ROBUST IMAGE SEGMENTATION

A. Kim, I. Pollak, H. Krim and A.S. Willsky
Stochastic Systems Group,
LIDS, MIT, Cambridge MA 02139

Abstract

Image segmentation represents an essential step in the early stages of an Automatic Target Recognition system. We propose two robust approaches fundamentally based on scale information inherent to a given imagery. The first approach is parametric in that the scale evolution of an image is statistically captured by a model which is in turn utilized to classify the pixels in the image. The second, on the other hand, nonlinearly evolves the image along some specific characteristic to unravel and delineate the various comprising entities in it.

1 Introduction

The growing interest in Automatic Target Recongnition (ATR) is primarily due to its importance in many applications ranging from manufacturing to remote sensing and surface surveillance. Analysis and classification of various entities of an image are often of great interest, and its partitioning into a set of homogeneous regions or objects is thus of importance. The mere size of some imagery constitutes a major hurdle. A case in point is Synthetic Aperture Radar (SAR) imagery for which terrain coverage rates are very high (in excess of $1\,km^2/s$) and which with daunting computational demands, make algorithmic efficiency of central importance.

A SAR image reflects a coherent integration of scatterer returns (i.e. reflectivity characteristics) within a resolution cell. The number of scatterers which coherently sum up within a cell will vary with the resolution. This leads to a variation in the underlying statistics of the image. Natural clutter for example, tends to consist of a large number of equivalued scatterers, and this in contrast to the man-made one, mostly comprising a few prominent scatterers. It is precisely this type of statistical characteristic that we are interested in capturing and subsequently using as the basis for our classification of various terrain types in SAR. In addressing such a problem, we have two basic approaches, the first aims at characterizing the statistics of the image evolution in scale and ultimately using them in the pixel classification; the second attempts to diffuse "noncomplying" observations via a nonlinear evolution to make it converge to specific desired domains of attraction.

The goal of this paper is to address a fundamental problem arizing in ATR, namely the robust and efficient segmentation of imagery into homogeneous regions, and in presence of perhaps severe noise. To proceed, we introduce in the next section a multiscale stochastic modeling framework which affords one to capture the evolution in scale of a given image process and to statistically characterize regions of interest. Two algorithms based on this framework will be described and shown to lead to efficient and accurate segmentation. In Section 4, we present our second method based on a variational approach, and which consists of nonlinearly evolving a given image along some specific geometric constraints built around an energy functional. Finally, in Section 5, we provide a number of examples substantiating the proposed algorithms using real data imagery and accurate estimates of boudaries.

[0]This report describes research supported in part by DARPA under contract FA49620-93-1-0604.

2 Stochastic Modeling

2.1 Multiscale Stochastic Models

In this subsection, we describe a general multiscale modeling framework e.g. [2] and its adaptation to classification/identification problems. Under this framework, a multiscale process is mapped onto nodes of a *qth* order *tree*, where q depends upon how the process progresses in scale.

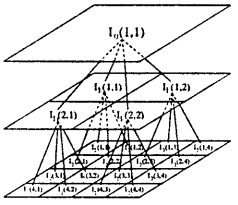

Fig. 1: Multiresolution tree.

As illustrated in Fig. 1, a *qth* order tree is a connected graph in which each *node*, starting at some root node, branches off to q child nodes. As described above, the appropriate representation for a multiscale SAR image sequence is $q = 4$, a quadtree. Each level of the tree (i.e., distance in nodes from the root node) can be viewed as a distinct scale representation of a random process, with the resolutions proceeding from coarse to fine as the tree is traversed from top to bottom (root node to terminal nodes). A coarse-scale shift operator, $\bar{\gamma}$ is defined to reference the *parent* of node s, just as the shift operator z allows referencing of previous states in discrete time-series. The state elements at these nodes may be modeled by the coarse-to-fine recursion

$$\mathbf{x(s)} \;=\; \mathbf{A(s)x(s\bar{\gamma})} \;+\; \mathbf{B(s)w(s)}. \qquad (1)$$

In this recursion, $\mathbf{A}(s)$ and $\mathbf{B}(s)$ are matrices of appropriate dimension and the term $\mathbf{w(s)}$ represents white driving noise. The matrix $\mathbf{A}(s)$ captures the deterministic progression from node $s\bar{\gamma}$ to node s, i.e., the part of $\mathbf{x}(s)$ predictable from $\mathbf{x}(s\bar{\gamma})$, while the term $\mathbf{B}(s)\mathbf{w}(s)$ represents the unpredictable component added in the progression. An attractive feature of this framework is the efficiency it provides for signal processing algorithms. This stems from the *Markov* property of the multiscale model class, which states that, conditioned on the value of the state at any node s, the processes defined on each of the distinct subtrees extending away from node s are mutually independent.

For pixel classification purposes, a multiscale model can be constructed for each specific homogeneous class. To specify each model, it is necessary to determine the appropriate coefficients in the matrices, $\mathbf{A}(s)$ and $\mathbf{B}(s)$, and the statistical properties of the driving noise, $\mathbf{w}(s)$. Once the models have been specified, a likelihood ratio test can be derived to segment the imagery into the clutter classes.

For a binary classification problem (i.e. requiring a binary hypthesis test) each pixel in the image corresponds to one of two hypotheses: the pixel is part of some texture (H_g) or another (H_f). By exploiting the Markov property associated with the multiscale models for imagery, the log-likelihood ratio test for classifying each pixel can be written as,

$$\begin{aligned}\ell \;&= \textstyle\sum_s \log\left[p_{\mathbf{x}(s)|\boldsymbol{\beta}_g}, (\mathbf{X}(s) \mid \boldsymbol{\beta}_g)\right] - \\ &\quad \textstyle\sum_s \log\left[p_{\mathbf{x}(s)|\boldsymbol{\beta}_f}(\mathbf{X}(s) \mid \boldsymbol{\beta}_f)\right], \qquad (2)\end{aligned}$$

where $\boldsymbol{\beta}_{(\cdot)} = (\mathbf{X}(s\bar{\gamma}), H_{(\cdot)})$, and $p_{\mathbf{x}(s)|\boldsymbol{\beta}_g}$ and $p_{\mathbf{x}(s)|\boldsymbol{\beta}_f}$ are the conditional distributions for the two hypothesized models. In the next subsection, we will show that this likelihood test can be efficiently computed in terms of the distributions for $\mathbf{w}(s)$ under the two hypotheses.

2.2 Scale-Autoregressive SAR Model

In this paper, we focus on a specific class of multiscale models, namely scale-autoregressive models [2] of the form

$$\begin{aligned}I(s) \;=\;& a_1(s)I(s\bar{\gamma}) + a_2(s)I(s\bar{\gamma}^2) + \ldots + \\ & a_R(s)I(s\bar{\gamma}^R) + w(s),\; a_i(s) \in I\!R \quad (3)\end{aligned}$$

where $w(s)$ is white driving noise and "s" is a three-tuple vector denoting scale (or resolution level), and spatial coordinates (q, n). For homogeneous regions of texture, the prediction coefficients (the $a_i(s)$ in (3)) are constant with respect to image location for any given scale. That is, the coefficients, $a_1(s), \ldots, a_R(s)$, depend only on the scale of node s (denoted by $m(s)$), and thus will be denoted by

$a_{1,m(s)}, \ldots, a_{R,m(s)}$. Furthermore, the probability distribution for $w(s)$ depends only on $m(s)$. Thus, specifying both the scale-regression coefficients and the probability distribution for $w(s)$ at each scale completely specify the model.

Following the procedure of state augmentation used in converting autoregressive time series models to state space models, we associate to each node s a R-dimensional vector of pixel values, where R is the order of the regression in (3). The components of this vector correspond to the SAR image pixel associated with node s and its first $R-1$ ancestors. Specifically, we define

$$\mathbf{x}(s) = \left[I(s) \quad I(s\overline{\gamma}) \quad \cdots \quad I(s\overline{\gamma}^{R-1}) \right]^T. \quad (4)$$

Thus, for a model of the form (3) or equivalently (1), ℓ in (2) can be calculated using

$$p_{\mathbf{x}(s)|\mathbf{x}(s\overline{\gamma})}(\mathbf{X}(s) \mid \mathbf{X}(s\overline{\gamma})) = p_{w(s)}(W(s)) \quad (5)$$

with $W(s)$ being the vector of residuals at scale "s". The identification of the model for each clutter class can thus be obtained for each scale m by a standard least-squares minimization,

$$\boldsymbol{\alpha}_m = \arg \min_{\boldsymbol{\alpha}_m \in \mathbb{R}^R} \left\{ \sum_{\{s \mid m(s)=m\}} [I(s) - a_{1,m} I(s\overline{\gamma}) - \ldots - a_{R,m} I(s\overline{\gamma}^R)]^2 \right\} \quad (6)$$

where

$$\boldsymbol{\alpha}_m = [a_{1,m}\ a_{2,m}\ \ldots\ a_{R,m}]^T,$$

with R being the regression model order. For most of the results presented in Section 5 a third order regression ($R=3$) for both the grass and the forest models was chosen. To obtain a statistical characterization of the prediction error residuals (the $w(s)$ in (3)) of the model at scale m, we evaluate the residuals in predicting scale m of a homogeneous test region. In particular, we use the $\boldsymbol{\alpha}_{m,\mu}$ found in (6) to evaluate all $\{\mathbf{w}(s) | m(s)=m\}$ as indicated by Eq. (5) with μ specifying "grass" or "forest".

3 Model-Based Classification

3.1 Residual-Based Classification

While we could conceivably postulate a spatial random field model for each windowed clutter category to classify its center pixel, we use to advantage the efficiency of multiscale likelihood calculation to base the classification of each individual pixel "s" on a surrounding $(2K+1) \times (2K+1)$ window $\mathcal{W}(\mathbf{s})$, where the parameter K is a nonnegative and judiciously selected integer. While a larger window provides a more accurate classification of homogeneous regions, it also increases the likelihood that the window contain a clutter boundary. Thus, keeping the window size as small as possible is also desirable. As shown in [1] in the next section, one can determine the tradeoff between classification accuracy and window size by examining the empirical distribution of ℓ over windows of various size for homogeneous regions of grass and forest.

3.1.1 Statisical Hypotheses Test

Whenever a clutter boundary is present within a test window the validity of the center pixel classification is questionable. This effect results in a classification bias near boundaries. To address this problem, we devise a method to detect the proximity of grass-forest boundaries and subsequently utilize a procedure to refine the classification. Terrain boundary proximity is detected via a simple modification of the decision made based on the test statistic ℓ. Specifically, rather than comparing ℓ to a single threshold to decide on a grass-or-forest classification, we compare ℓ to two thresholds a and b as shown in Fig. 2, and where we include a defer decision.

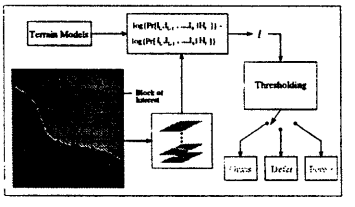

Fig. 2: Initial pixel classification.

This is tantamount to refining the decision procedure by considering smaller windows within the original window until a majority rule lifts the ambiguity,

$H_g : \ell > a$ Classify as Grass,

$a > \ell > b$ Defer decision (possible boundary presence),

$H_f : \ell < b$ Classify as Forest.

3.2 Parameter-Based Classification

An alternative approach to classifying a clutter type is to evaluate an l^2 distance between the computed model parameters for a window region $\mathcal{W}(\mathbf{s})$ and those of a template homogeneous region [3]. For each pixel "s" (or equivalently tree node) a model for a corresponding surrounding window $\mathcal{W}(s)$ is computed at several scales giving rise to what we refer to as an evolution vector statistic,

$$\mathbf{y}(\mathbf{s}) = [\boldsymbol{\alpha}_{m(s)}, \boldsymbol{\alpha}_{m(s)-1}, \cdots, \boldsymbol{\alpha}_1], \qquad (7)$$

where "s" will sweep all of the three-tuple vectors $[m(s), q, n]$ at the $m(s)^{th}$ resolution with $m(s) = 1, \cdots, L-1$ and L is the cardinality of the considered set of images. We should note at this point that in this approach, the DC component in the $\boldsymbol{\alpha}_{(.)}$ is included. The order of the regression associated with modeling $\mathcal{W}([m(s), q, n])$ from its ancestors will vary with the level $m(s)$ as defined by the function

$$O_{m(s)} = \max(R, m(s))$$

The regression vector $\boldsymbol{\alpha}(m(s))$ provides a statistically optimal description of the linear dependency of $\mathcal{W}(\mathbf{s})$ on $\{\mathcal{W}(s\gamma^j)\}_{j=1}^{O_{m(s)}}$, and $\mathbf{Y}(\mathbf{s})$ thus being a measure of the scale evolution behavior of the windowed region.

Here again the size of the window used to compute the evolution vector, is the result of a tradeoff between *modeling consistency* and *local accuracy.*

3.2.1 Statistical Classification

A characterization of the evolution vector $\mathbf{y}(\mathbf{s})$ is necessary to carry out a statistically meaningful classification of various types of terrain. Specifically, a BHT is applied to the evolution $\mathbf{y}(\mathbf{s})$ in order to classify node $\mathbf{s}$ as a member of either a region of grass or of forest. These hypotheses which are respectively designated as hypotheses H_g and H_f. The classification of pixel $m(s)$ will depend only on $\mathbf{y}(\mathbf{s})$ and the predetermined likelihoods $p(\mathbf{y}(\mathbf{s})|H_g)$ and $p(\mathbf{y}(\mathbf{s})|H_f)$.

To carry out a statistically significant hypothesis test, we need to specify the conditional probability density for $\mathbf{y}$ under each hypothesis. To do this, we extensively examined the distribution of the evolution vectors obtained from a large homogeneous region of the corresponding terrain. For both grass and forest terrain, it turns out that each component in $\mathbf{y}$ approximately has a Gaussian distribution. We consequently make the approximation that $p(\mathbf{y}|H_g)$ and $p(\mathbf{y}|H_f)$ are N-variate Gaussian densities. They are then completely specified by their mean vectors $\mathbf{m}_g$ and $\mathbf{m}_f$ and their covariance matrices K_g and K_f all of which are calculated from the training data for each hypothesis. A maximum likelihood (ML) detector is then used to classify each pixel (using the ML detector assumes equal apriori probabilities for each hypothesis and a cost function that penalizes all misclassifications equally). In implementing the ML detector, a threshold η is calculated from the likelihoods and used in the classification of each pixel through its comparison to a sufficient statistic derived from the evolution vector. Because $p(\mathbf{y}|H_g)$ and $p(\mathbf{y}|H_f)$ are assumed to be Gaussian, it is straightforward to compute the threshold η and the sufficient statistic $\ell'(\mathbf{y})$ for each evolution vector as

$$\eta = \log \frac{|K_g|}{|K_f|}, \qquad (8)$$

and

$$\ell'(\mathbf{y}) = (\mathbf{y}-\mathbf{m}_f)^T K_f^{-1} (\mathbf{y}-\mathbf{m}_f) - (\mathbf{y}-\mathbf{m}_g)^T K_g^{-1} (\mathbf{y}-\mathbf{m}_g). \qquad (9)$$

The classification of $\mathbf{I}(\mathbf{s})$, denoted as $\mathbf{C}(\mathbf{s})$, is then given by

$$\mathbf{C_s} = \begin{cases} H_g & \text{if } \eta \geq \ell\left(\mathbf{y}_{[q,n]}\right) \\ H_f & \text{if } \eta < \ell\left(\mathbf{y}_{[q,n]}\right). \end{cases} \qquad (10)$$

The construction of the evolution $\mathbf{y}$ and subsequent application of a BHT for each $s \in$

$\{s \mid m(s) = m\}$ thus provide a segmentation of $\mathbf{I}_m$. Instead of independently generating a segmentation for each image resolution for all $s \in \{s \mid m(s) < L\}$, $\mathbf{C(s)}$ can be obtained by comparing η to the average of the sufficient statistics of nodes in $\mathbf{I}_L$ which have s as an ancestor, i.e.

$$\mathbf{C}_{[1,m,n]} = \begin{cases} H_g & \text{if } \left(2^{m(s)-1}\right)^2 \eta \geq \sum_{\mathbf{s}} \ell\left(\mathbf{y(s)}\right) \\ H_f & \text{if } \left(2^{m(s)-1}\right)^2 \eta < \sum_{\mathbf{s}} \ell\left(\mathbf{y(s)}\right). \end{cases} \tag{11}$$

Although this does not yield a sufficient statistic for $\mathbf{I(s)}$, doing so is computationally more efficient than calculating a sufficient statistic as in Eq. (9) for each node at every level in the quadtree. Note that the segmentation technique described here, could easily be generalized to a larger number of terrain types.

4 Stabilized Inverse Diffusion Equations (SIDEs)

The previous two techniques rely on modeling a linear evolution of the observed imagery with an ultimate goal of pixel classification. In this section, we instead carry out a nonlinear evolution which is driven by prescribed geometric features underlying the process/imagery, and study its progression.

Towards that end, we introduce a discontinuous force function, resulting in a system of equations that has discontinuous right-hand side (RHS). As shown below, the objective is to drive an evolution trajectory onto a lower-dimensional surface which clearly has value in image analysis, and in particular in image segmentation. Segmenting a signal or image, represented as a high-dimensional vector $\mathbf{I}$, consists of evolving it so that it is driven onto a comparatively low-dimensional subspace, which corresponds to a segmentation of the signal or image domain into a small number of regions.

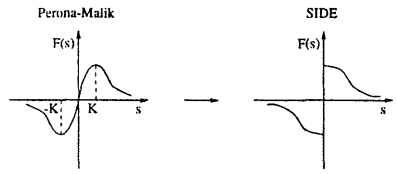

Fig 3: Force function for a SIDE.

The type of force function of interest to us here is illustrated in Fig. 3(right). More precisely, we wish to consider force functions $F(v)$ which, in addition to drving the following evolutions,

$$\begin{aligned} \dot{\mathbf{I}}(s) &= F(\mathbf{I})(s), \\ \mathbf{I}(0) &= \mathbf{I}_0, \end{aligned} \tag{12}$$

where "s" is now a continuous scale, satisfy the following properties:

$$\begin{aligned} F'(v) \leq 0 \quad & \text{for } v \neq 0, \text{ and } F(0^+) > 0, \\ & F(v_1) = F(v_2) \Leftrightarrow v_1 = v_2. \end{aligned} \tag{13}$$

Contrasting this form of a force function to the Perona-Malik function [4] in Fig. 3 (left), we see that in a sense, one can view the discontinuous force function as a limiting form of the continuous force function in Fig. 3 (left). We therefore need a special definition of how the trajectory of our evolution proceeds at the discontinuity points $F(0^+) \neq F(0^-)$). For this definition to be useful, the resulting evolution must satisfy well-posedness properties: the existence and uniqueness of solutions, as well as stability of solutions with respect to the initial data. Assuming the resulting evolutions to be well-posed, we demonstrate that they have the qualitative properties we desire, namely that they both are stable and also act as inverse diffusions and hence enhance edges. We address the issue of well-posedness and other properties in [5].

Considering the evolution (12) with $F(v)$ as in Fig. 3(right) in a SIDE, notice that the RHS of (12) has a discontinuity at a point $\mathbf{I}$ if and only if $I_i = I_{i+1}$ for some i between 1 and $N-1$. It is when a trajectory reaches such a point $\mathbf{I}_{(\cdot)}$ that we need the following definition:

$$\dot{\mathbf{I}}_i = \dot{\mathbf{I}}_{i+1} = \frac{1}{2}((F(\mathbf{I}_{i+2} - \mathbf{I}_{i+1}) - F(\mathbf{I}_i - \mathbf{I}_{i-1})). \tag{14}$$

In other words, the two observations are simply merged into a single one, resulting in Eq. 14 for $n = i$ and $n = i + 1$ (the differential equations for $n \neq i, i + 1$ do not change.).

Similarly, if q consecutive observations become equal, they are merged into one which is weighted by $1/q$ [5]. The evolution can then naturally be thought of as a sequence of stages: during each stage, the right-hand side of (12) is continuous. Once the solution hits a discontinuity surface of the right-hand side, the state reduction and re-assignment of q_{n_i}'s, described above, takes place. The solution then proceeds according to the modified equation until the next discontinuity surface, etc.

Notice that such an evolution automatically produces a multiscale segmentation of the original signal if we view each compound observation as a region of the signal. The algorithm may be be summarized as follows:

1. Start with the trivial initial segmentation: each sample is a distinct region.
2. Evolve (12) until the values in two or more neighboring regions become equal.
3. Merge the neighboring regions whose values are equal.
4. Go to step 2.

The same algorithm can be used for 2-D images, which is immediate upon re-writing Eq. (12) and an example is provided next:

5 Experiments

A number of segmentation experiments have been carried out on SAR imagery using the three techniques described above. Fig. 4 shows a segmentation as a result of a systematic pixel classification as described in Technique 1. A brute force as well as a refined segmentations are shown, and the handrawn of the eyeballed boundary is shown in white. In Fig. 5, the model-based segmentation is illastrated and in Fig. 6 the nonlinear evolution-based segmentation is shown which clearly is perhaps the most promising, albeit at a slightly higher computational cost.

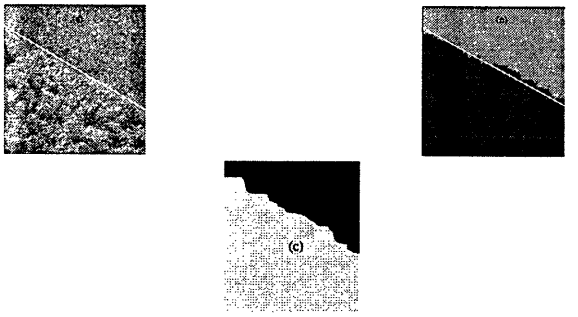

Fig. 4: Residual and Model-Based Segmentation (b and c).

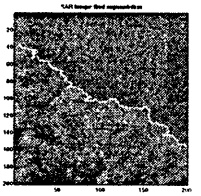

Fig. 6: Segmentation by Nonlinear Diffusion.

References

[1] C. Fosgate, H. Krim, W. Irving, W. Karl, and A. Willsky. Multiscale segmentation and anomaly enhancement od sar imagery. *IEEE Trans. on Im. Proc.*, 6(1):7–20, Jan. 96.

[2] W.W. Irving. Multiresolution approach to discriminating targets from clutter in sar imagery. In *Proc. of SPIE Symp.*, Orlando, FL., 1995.

[3] A. Kim and H. Krim. Hierarchical stochastic modeling of sar imagery for segmentation/compression. *submitted to special issue of IEEE Trans. on SP*, 1996.

[4] P. Perona and J. Malik. Scale-space and edge detection using anistropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12, 1990.

[5] I. Pollak, A.S. Willsky, and H. Krim. Image segmentation via regularized inverse diffusion. *To be submitted to IEEE Trans. on Image Processing.*