

Coding Theoretic Approach to SAR Image Segmentation

Unoma Ndili^a, Robert Nowak^a, Richard Baraniuk^a, Hyeokho Choi^a and Mário Figueiredo^b

^aRice University
Department of Electrical and Computer Engineering
Houston, Texas, U.S.A.

^bInstituto de Telecomunicações, and
Instituto Superior Técnico
1049-001 Lisboa, Portugal

ABSTRACT

In this paper, a coding theoretic approach is presented for the unsupervised segmentation of SAR images. The approach implements Rissanen's concept of Minimum Description Length (MDL) for estimating piecewise homogeneous regions. Our image model is a Gaussian random field whose mean and variance functions are piecewise constant across the image. The model is intended to capture variations in both mean value (intensity) and variance (texture). We adopt a multiresolution/progressive encoding approach to this segmentation problem and use MDL to penalize overly complex segmentations. We develop two different approaches both of which achieve fast unsupervised segmentation. One algorithm is based on an adaptive (greedy) rectangular recursive partitioning scheme. The second algorithm is based on an optimally-pruned wedgelet-decorated dyadic partition. We present simulation results on SAR data to illustrate the performance obtained with these segmentation techniques.

Keywords: Segmentation, Minimum Description Length, SAR, Multiscale, Wedgelets, CART, Adaptive Recursive Partitioning

1. INTRODUCTION

Synthetic Aperture Radar (SAR) provides all-weather, day-or-night, remote sensing for applications such as mapping, search-and-rescue, mine detection, and target recognition. SAR data processing entails forming an image from measured radar backscatter returns, followed by processing to detect and recognize 'targets' from the formed image. Intensity and textural properties in SAR imagery can therefore provide important perceptual cues to identify candidate targets. Segmentation of SAR data is important as a first step to target recognition and entails identifying regions of homogenous intensity or 'textural' properties within the image. Several supervised texture segmentation schemes have been developed for SAR data, requiring training prior to segmentation.¹⁻³ Other unsupervised schemes have achieved segmentation based only on intensity (mean) values without consideration for textural variations (constant variance models).^{4,5} To the best of our knowledge, this scheme is the first in the literature to provide an unsupervised segmentation based on both mean (intensity) and variance (texture).

Our segmentation scheme is aimed at automatically identifying regions of both varying mean and/or variance in SAR imagery. We suppose that the m -by- n image under study is composed of connected regions of pixels with each pixel assumed to be independently and identically distributed (i.i.d.) according to a Gaussian density and modeled as:

$$y(i_1, i_2) = f(i_1, i_2) + \sigma(i_1, i_2) z(i_1, i_2) \quad 0 \leq i_1, i_2 \leq m, n$$

where $z(i_1, i_2) \sim$ i.i.d. $\mathcal{N}(0, 1)$ so that $y(i_1, i_2)$ is distributed according to $\mathcal{N}(f(i_1, i_2), \sigma^2(i_1, i_2))$. Thus our image model is a Gaussian random field whose mean and variance functions both vary across the image. We assume that the mean f and variance σ^2 functions are piecewise constant. This model is intended to capture variations in both mean value (intensity) and variance (texture). This is different from the signal plus *constant noise models* used in most wavelet-based multiscale image analysis procedures, and finds applications in a wide class of natural and

E-mail^a: {unondili, nowak, richb, choi}@ece.rice.edu, Web: dsp.rice.edu, E-mail^b: mtf@lx.it.pt
This work was supported by the ONR grant N00014-99-1-0813

radar images including SAR. For the sake of completeness, we discuss a simple transformation of the SAR data to approximate a Gaussian distribution. However this step is not really necessary since our statistical model of piecewise constant mean and variance functions is generic to images.

We use a coding theoretic approach based on Rissanen’s MDL principle⁶⁻⁸ to achieve unsupervised segmentation. MDL penalizes overly complex segmentation models thereby providing a means for data dependent model selection. Conditional distributions based on our model for use with the MDL criteria are obtained for the progressive coding scheme we adopt. We propose two different approaches which involve a tradeoff in terms of optimality versus speed and flexibility. The **Adaptive Recursive Partitioning** (ARP) algorithm is a greedy scheme which achieves rectangular (not necessarily dyadic) tessellations while the **Wedgelet Decorated Dyadic Partition** (WDDP) algorithm is an optimal multiscale analysis restricted to dyadic partitions but in addition allows for wedge-splits. This algorithm is based on the dyadic CART⁹⁻¹¹ with the exception that our model enables the detection of changes in variance, as will be illustrated in our results.

Our paper is organised as follows: Section 2 briefly discusses the statistical model for SAR data and presents a simple transformation to approximate SAR data with a Gaussian density. Section 3 reviews coding theory and the MDL principle and derives the conditional density required by our Gaussian model in the MDL criteria. In Section 4 we describe the two schemes: the ARP and WDDP algorithms. Section 5 presents results of the two schemes on a test image with analysis and comparisons. We present an illustrative example on a single look SAR data from the MSTAR clutter data HB06170¹² and on a multi-look SAR image. We conclude in section 6.

2. STATISTICAL MODEL FOR SAR IMAGES

Before proceeding to examine the statistical characteristics of SAR data, we want to emphasize that the MDL principle is not concerned with obtaining a code length based on the data generating distribution. In fact, in most statistical applications, the true distribution is rarely known. Any distribution Q defined on the data, not necessarily the true distribution can be used to encode the data.^{13,14} This is because we are more concerned with model selection given a number of possible classes. This means that our MDL criteria based on a Gaussian assumption will do a reasonable job on an image with regions of constant mean and/or variance irrespective of the true distribution of the data. Having said that, we proceed in this section to examine the statistics of SAR data for the sake of completeness.

Suppose we have a random *speckled* SAR data set Z^* which is typically complex. It had been observed that the pixels y of the SAR amplitude image $Y = |Z|$ are Rayleigh distributed. In order to achieve greater accuracy in the approximation of small sample size, we model the image pixels $y \in Y$ with a two-parameter Rayleigh population having a pdf:

$$p_Y(y) = \frac{y - \mu_y}{\sigma_y^2} e^{-(y - \mu_y)^2 / (2\sigma_y^2)}, \quad y \geq \mu_y, \quad \sigma_y^2 > 0, \quad (1)$$

where μ_y and σ_y^2 are the mean and variance of the distribution. We employ the estimating equations¹⁵ $E[Y] = \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ (the sample mean) and $Var(Y) = s^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$ (the sample variance) to obtain simple moment estimators for σ_y^2 and μ_y as:

$$\hat{\sigma}_y^2 = 2s^2 / (4 - \pi)$$

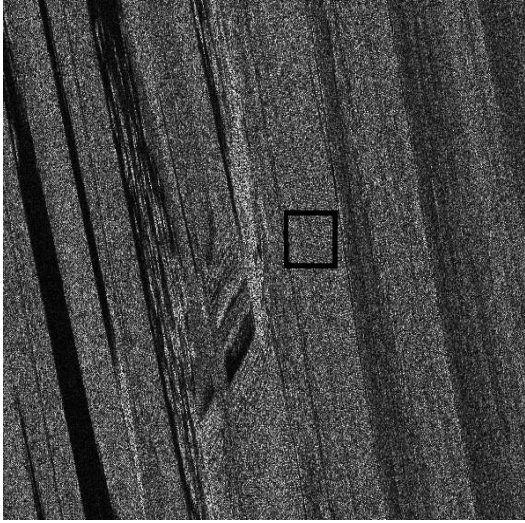
and

$$\hat{\mu}_y = \bar{Y} - \hat{\sigma}_y \sqrt{\pi/2}$$

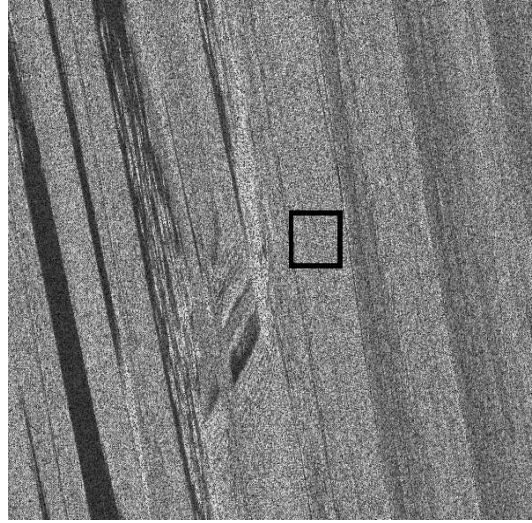
It has been shown in¹⁶ that for a variable Y with a distribution whose mean and variance are functions of the same parameter s such as we have for the Rayleigh, a simple variance-stabilizing transform such as $X = \sqrt{Y}$ is approximately Gaussian with p.d.f. given as:

$$p_X(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-(x - \mu_x)^2 / (2\sigma_x^2)}, \quad (2)$$

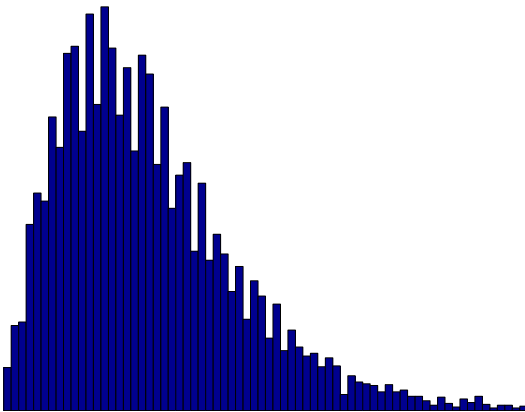
The effect of this simple transformation is illustrated in Figure 1 where (a) and (b) show an example of an original Y and transformed \sqrt{Y} SAR image from the MSTAR database.¹² Figures 1 (c) and (d) are the corresponding histograms of a homogenous region (outlined in the SAR images). We note that the Rayleigh distribution is approximately Gaussian after the square-root transformation for each homogenous region. We therefore assume X is generated by a Gaussian process and use this as input to our algorithm.



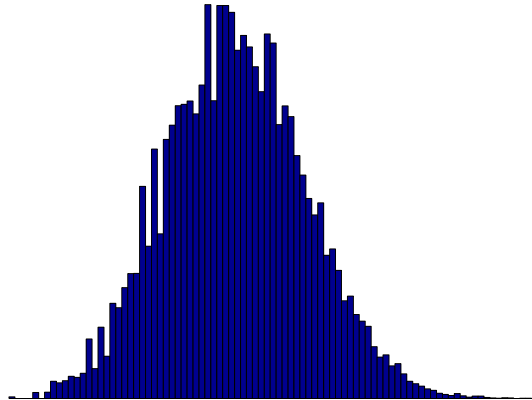
(a)



(b)



(c)



(d)

Figure 1. (a) Original Y and (b) \sqrt{Y} transformed SAR images with corresponding histograms showing their approximate (c) Rayleigh and (d) Gaussian distributions

3. CODING THEORETIC APPROACH AND MDL PRINCIPLE

Now that the SAR data has been transformed to be approximately Gaussian, we now proceed to briefly describe the MDL principle used here.

Suppose we want to optimally encode and transmit a finite set of data sequence $X^n = \{x_1, \dots, x_n\}$ of length n to some decoder. Given a probability distribution $p(X^n|\theta)$ on X^n , where θ parameterizes the distribution, the Shannon-optimal code length for prefix codes is given as $-\log p(X^n|\theta)$.¹⁷ In this paper we assume logarithm to base 2 so that the codelengths are in bits. Assume we have a set of K competing models to explain our data $\{p_i(X^n|\theta_i)\}_{i=1}^K$, the MDL criteria^{7,8} states that among these K possibilities, the ‘best’ model is the one that minimizes the description length obtained by assuming a two-part code,

$$L(X^n) = L(\hat{\theta}_k) - \log p(X^n|\hat{\theta}_k), \quad k = 1, \dots, K, \quad (3)$$

where $L(\hat{\theta}_k)$ is the code length required to describe the maximum likelihood estimate (MLE) $\hat{\theta}_k$ of the parameter θ_k such that the decoder knows the model under which the code for the data (of length $-\log p(X^n|\hat{\theta}_k)$) was obtained. MDL has been successfully used in several image analysis/segmentation problems.⁴

3.1. MDL FOR GAUSSIAN DATA

When applying MDL it is necessary to address the issue of how to encode the parameter estimate $\hat{\theta}_k$. Under a Gaussian data assumption, the elements of $\hat{\theta}_k$ are real-valued and have to be truncated to finite precision in order to yield finite code-length $L(\hat{\theta}_k)$. The standard solution is the well known $\frac{1}{2} \log n$ code-length for each components of θ_k , which is based on asymptotic approximations.^{6,13}

The MDL approach above reviewed uses two-part codes: we first encode and send a parameter estimate, then the data itself, coded according to that parameter estimate. However, if we build a code for all possible data out-comes, this code is *incomplete*.¹⁸ In fact, once the receiver has the parameter estimate, it knows that only data out-comes that could have led to this estimate are possible. To avoid incompleteness, we derive below the conditional density of the data, given the parameter estimates (or equivalently, the corresponding sufficient statistics).

3.1.1. Conditional density for Gaussian with unknown mean and variance

Let us assume that $X^n = \{x_1, \dots, x_n\}$ is a set of i.i.d. $\mathcal{N}(f, \sigma^2)$ samples to be transmitted. We adopt a progressive coding scheme in which we first describe the parameters of our prevailing model to the receiver. The sufficient statistics that have to be collect to obtain estimates of f and σ^2 are the sum $\sum_{i=1}^n x_i$ and sum of squares $\sum_{i=1}^n x_i^2$. These constitute the information that is initially sent to the decoder. We next need to describe the data under this assumed model, given that the receiver already knows the parameters initially transmitted. To avoid the above mentioned incompleteness issue, we determine the conditional density $-\log p(X^n | \sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$ based on which we have to build the code for the data.

Our Gaussian model parameterized by the mean f and variance σ^2 may be expressed in exponential family form with two natural parameters $\zeta_1 = \frac{f}{\sigma^2}$ and $\zeta_2 = -\frac{1}{2\sigma^2}$ for which $\sum x_i \equiv t_1$ and $\sum x_i^2 \equiv t_2$ are the corresponding sufficient statistics. Using Neyman's factorization, we can write

$$\begin{aligned} p(X^n | \zeta_1, \zeta_2) &= \left(\frac{-\zeta_2}{\pi} \right)^{\frac{n}{2}} e^{\left(\frac{\zeta_1}{4\zeta_2} \right)} e^{\left(\zeta_1 \sum x_i + \zeta_2 \sum x_i^2 \right)} \\ &= \underbrace{\psi(\zeta_1, \zeta_2) \exp(\zeta_1 t_1 + \zeta_2 t_2)}_{\propto p(t_1, t_2 | \zeta_1, \zeta_2)} \underbrace{1}_{\propto p(X^n | t_1, t_2)} \end{aligned}$$

Surprisingly the conditional density $p(X^n | t_1, t_2)$ is proportional to 1 showing that it is uniform on the constraint set $\mathcal{C}(t_1, t_2) = \{X^n : \sum x_i = t_1 \text{ and } \sum x_i^2 = t_2\}$. The normalizing constant is then simply the area of the constraint set \mathcal{C} which is a $(n-1)$ -sphere resulting from the intersection of a n -sphere of radius $\sqrt{t_2}$ centered at the origin with a hyper-plane at 45° in R^n . The distance from the center point $(\frac{t_1}{n}, \dots, \frac{t_1}{n})$ on the plane to the origin is $\frac{t_1}{\sqrt{n}}$ and is perpendicular to the hyper-plane, thus the radius r of this $(n-1)$ -sphere is $r = \sqrt{t_2 - \frac{t_1^2}{n}}$. Using the formula for the surface area of a sphere, the joint p.d.f. of X^n conditioned on t_1 and t_2 is (with $\Gamma(\cdot)$ denoting Euler's gamma function) is given as:

$$p(X^n | t_1, t_2) = \begin{cases} \frac{\Gamma(\frac{n-1}{2}) r^{2-n}}{2\pi^{\frac{n-1}{2}}} & \Leftarrow X^n \in \mathcal{C}(t_1, t_2) \\ 0 & \Leftarrow X^n \notin \mathcal{C}(t_1, t_2). \end{cases} \quad (4)$$

4. TWO ALGORITHMS

Now that we have the conditional distribution required for description lengths as per MDL we consider two different approaches. Suppose we want to transmit an m -by- n image. First we introduce some general notation: Given any region R of size n_R -by- m_R , let X^R describe the data in region R i.e. $x \in R$ with mean f_R and variance σ_R . Let us define $t_1^R \equiv \sum_{x \in R} x$ and $t_2^R \equiv \sum_{x \in R} x^2$. We may carve up R into H non-overlapping regions each represented as R_h for $h = 1, \dots, H$. We are now ready to describe our algorithms.

4.1. The Adaptive Recursive Partitioning Algorithm

The Adaptive Recursive Partitioning Algorithm (ARP) is a recursive algorithm that starts by examining the entire image to determine whether it is best represented as homogenous with a common mean and variance under model class 1, or else split (vertically or horizontally) into two rectangles or into four rectangles (with one common vertex) under model class 2. If a decision to split is made, we denote the resulting rectangular regions as R_h ($h = 1, 2$ or $1 \dots 4$) each having a different mean and/or variance. Each of the resulting subblocks is then recursively examined in this way to decide how it is ‘best’ described and processing on a subblock only stops when we decide on model class 1 or we reach trivial pixel partitions.

The MDL criteria is used to decide the ‘best’ model as explained above. The description length of a model consists of the code length required to indicate a given model class plus the number of bits required to encode the model parameter and the data given that model.

Model Class 1: Under model class 1, assuming sufficient statistics for the model parameters have already been transmitted by our progressive scheme, we only need to code and transmit the data as a homogeneous block. Given the conditional distribution obtained in sec (3.1.1), the description length for this model class is simply:

$$L_1(R) = -\log p(X^R | t_1^R, t_2^R)$$

Model Class 2: Under this model class we may split R into 2 or 4 rectangles. The best split point k is determined using the MDL criteria. We examine all possible split points, calculating codelengths for R given that split, and decide on the one that gives the minimum codelength. We decide to split if this codelength is less than that for the homogenous model under class 1. In that case we must then send the value of k to the decoder in addition to the sum and sum of squares of the resulting regions i.e. $t_1^{R_h}$ and $t_2^{R_h}$ to enable us describe the data in the R_h regions ($h = 1, 2$ or $1 \dots 4$), under different models using their parameterized density. Since there are $J = (m_R - 1)(n_R - 1)$ possible split positions, we require at most $\log J$ bits to code k . We note that if we split R into H rectangles ($H = 2$ or 4) we only need to send sufficient statistics for the mean and variance of $H - 1$ of these regions since we already have the statistics for R due to our progressive scheme. As an example, if there are J ways to split R into two rectangles $R_1(j)$ and $R_2(j)$, $j = 1, \dots, J$, so the description length for model class 2 is:

$$L_2 \{R(j)\} = -\log p(X^{R_1} | t_1^{R_1(j)}, t_2^{R_1(j)}) - \log p(X^{R_2} | t_1^{R_2(j)}, t_2^{R_2(j)}) + \log n_R m_R + \log J$$

In the trivial case of splitting a 2×2 subblock into 4, it is redundant to send 2 statistics (mean and variance) for just one data point. In this case we send only the value of that pixel. Now comparing codelengths, if $L_1 < \min L_2(j)$ model class 1 is chosen for the subblock and processing is stopped, otherwise we split at the point k which minimizes code lengths obtained from $L_2 \{R(j)\}$. Each rectangular region is split recursively under model class 2 until either L_1 is selected or we get to the trivial pixel level.

This algorithm was developed for Poisson data by Nowak and Figueredo⁴ and was used for the segmentation of SAR data⁵ based on intensity function only by transforming the SAR data to Poisson. Our technique captures both variance and intensity variations in the data using the Gaussian model for underlying unknown mean and variance functions. The ARP is a greedy algorithm because at each level we ignore the fact that each subblock may be further subdivided thus achieving an even shorter code length. At each stage an optimal decision is made however, the overall best possible segmentation may not be achieved.

4.2. Wedgelet-Decorated Dyadic Partition

The goal of the Wedgelet Decorated Dyadic Partition (WDDP) algorithm is to implement an optimal parsing of the observed image. Our approach here is an extension of the CART-based wedgelet algorithm first proposed by Donoho.⁹ The CART theorem¹⁰ asserts that given a function $\xi(T)$ which assigns numerical values to a quad tree T and its subtrees, and obeys a certain additivity property, the optimal subtree is obtained by a breadth-first bottom up pruning. This idea is used here: first we use a multiscale analysis to place a wedgelet-decorated tree on the observation. Due to the prohibitive complexity required to search through all possible subtrees we restrict our analysis to dyadic partitions of the observations. A wedge is a line from one edge of a region to another so that it splits the region into two polygons.⁹ Wedgelets are basis functions that enable us to split a dyadic region with a wedge at different locations and orientations. They add more flexibility to the partitioning scheme enabling non-dyadic tessellations of our observation space. To reduce the complexity of our algorithm, we do not search through

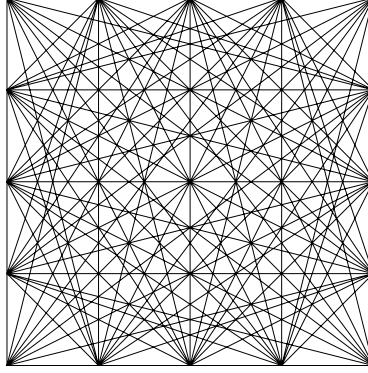


Figure 2. Wedge-splits of a dyadic block

the dictionary of all possible wedgelets at each scale since this would be computationally intensive to implement at all nodes. Instead marking off only $\tau = 4$ points on any edge of all blocks ensures that at each scale the cardinality of our wedgelets is fixed at $B = 80(\text{i.e. } 6\tau^2 - 4\tau)$. This idea is intuitively justified if we consider that wedgelets approximate arbitrary boundaries quite well if we zoom in enough, where the cardinality of our wedgelet search approaches that of the dictionary. A dyadic block with all possible wedge-splits is shown in figure2 with each line indicating a wedge. More on the wedgelet idea may be found in Donoho's original paper.⁹

The terminal nodes of the subtree are non-overlapping dyadic partitions or polygons from wedge-splits determined using the MDL criteria. This optimal tree is built by the following *hereditary* scheme: Given an n -by- n dyadic image, let $J = \log_2 n$ denote the tree depth and $j = 1, \dots, J$ indicate the scale such that the observation X^n is at the finest scale $j = 1$. A quad split on a dyadic region R produces four children denoted as $R_c = \{R_{c1}, \dots, R_{c4}\}$. Unlike ARP, we start from the bottom with 2×2 quad-split partitions and proceeding upwards inheriting at each stage the best possible description length for the subtree below a node. Each node of the tree is decorated with a description cost which is obtained from the following three models.

Model Class 1: Under this model we may represent R as a homogenous region with a mean and variance. The penalty is the number of bits required to transmit t_1 and t_2 therefore our codelength is:

$$L_1(R) = -\log p(X^R | t_1^R, t_2^R) + \log mn$$

Model Class 2: Perform a wedge split on R at B different orientations to obtain two polygons denoted as $R_1(b)$ and $R_2(b)$ for $b = 1 \dots B$. In addition to coding the individual sum and sum of squares for both regions and the data in each of the resulting regions given sufficient statistics for its parameters, we also need to describe the location of the wedge using at most $2 \log n_R$ bits. Our description length is thus:

$$L_2 \{R(b)\} = -\log p(X^{R_1(b)} | t_1^{R_1(b)}, t_2^{R_1(b)}) - \log p(X^{R_2(b)} | t_1^{R_2(b)}, t_2^{R_2(b)}) + \log n_{R_1} n_{R_2} + 2 \log n_R$$

and

$$L_2(R) = \min L_2 \{R(b)\} \quad \text{for } b = 1, \dots, B$$

Model Class 3: For this class we perform a quad-split on R into its 4 children R_c where each may be represented under a different model. The codelength of each individual child is *inherited* from previous (lower) levels as the best possible representation. The description length under this model is simply the sum of the optimal description length of its four children and is given as:

$$L_3(R) = \sum_{c=1}^4 \min \{L_1(R_c), L_2(R_c), L_3(R_c)\}$$

For each node on the tree the 'best' description length is the minimum of the three possible codelengths from the model classes i.e. $L = \min(L_1, L_2, L_3)$. We stop processing and prune our tree whenever L is either L_1 or L_2 .

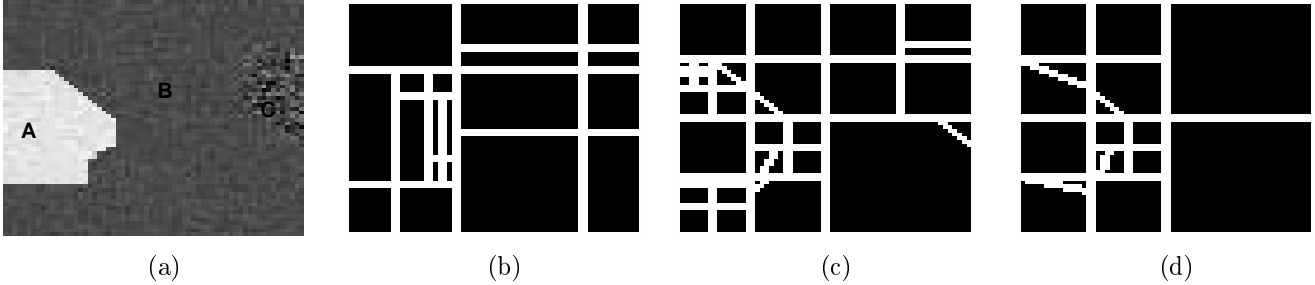


Figure 3. Results on synthetic test image (a) original image. Segmentation maps obtained using the following schemes (b) ARP (c) WDDP (d) Constant-variance WDDP

5. RESULTS AND APPLICATIONS

This section presents results of the two schemes and compares their performances. Figure 3(a) shows a synthetic test image used to illustrate the usefulness of our scheme. The patch on the left labeled as A has a different intensity from B but the same variance while C the region on the right has the same intensity but a different variance from B. The regions were made to have arbitrary shapes to illustrate the wedgelet idea. We have chosen to use segmentation maps to clearly illustrate where splitting occurs for each algorithm.

Figures 3(b) and (c) show segmentation maps for the Adaptive Recursive Algorithm (ARP) and the Wedgelet Decorated Dyadic Partition (WDDP) respectively. Both algorithms detect the change in texture on the right patch C. The ARP is not restricted to dyadic partitions and so we see rectangular tessellations of the observation, however, the WDDP does a better job approximating arbitrary shapes with wedges and is not blocky. Figure 3(d) is the segmentation map for the constant-variance wedgelet-decorated dyadic CART algorithm developed by Donoho in.⁹ This scheme performs nearly as well as the WDDP in regions with varying intensity but fails to detect vital changes in variance. The model developed here is therefore well suited for the segmentation of a wider class of images requiring both intensity and textural characterization.

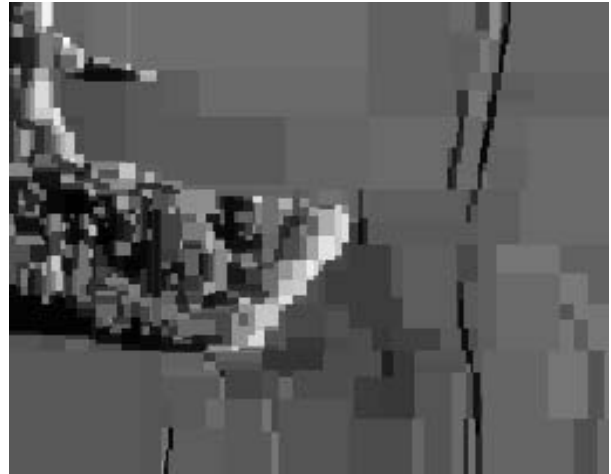
Figures 4 shows results of the various algorithms applied to original MSTAR SAR¹² amplitude data HB06170 in which changes in variance (texture) is an important feature. Figure 4 (a) is the original SAR image. Figures 4(b,c,d) show the ARP, WDDP and constant-variance CART algorithms applied to the SAR image. These results indicate that the WDDP outperforms the other algorithms. The ARP gives a blocky effect while the constant variance CART algorithm produces an under segmented image. Another illustration is shown in figure 5. The original (a) is a multi-look SAR image of agricultural landscape in Flevoland, Netherlands.¹⁹ Figures 5(b) and (c) are the resulting segmented images using ARP and WDDP respectively. We see that both algorithms correctly smooth the data in areas with homogenous regions. WDDP however does a better job in regions that have non-vertical and non-horizontal boundaries. We have tested our algorithm on several SAR images with similar results and conclude that the WDDP scheme outperforms the ARP. Both schemes are computationally fast and may be used for real-time applications.

6. CONCLUSION

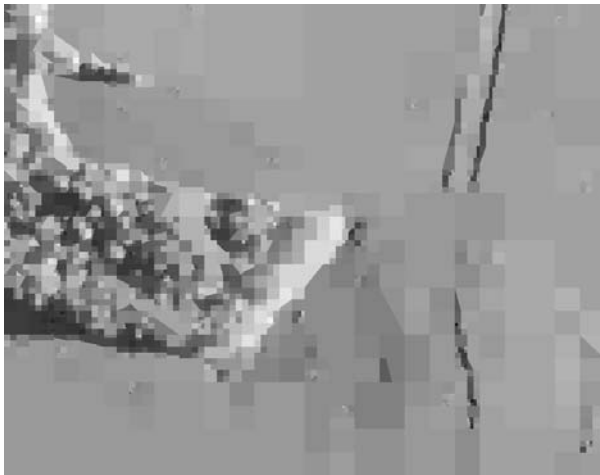
In this paper, we have investigated a novel approach to unsupervised segmentation of SAR images using the coding theoretic approach of minimum description length (MDL). Assuming piecewise homogeneous regions of an image, having a common intensity and/or textural properties, we assume a Gaussian random field which is appropriate, being parameterized by its mean and variance. Our model achieves segmentation based on intensity and variance features. This is important in target detection applications on SAR imagery. We develop two schemes: an adaptive recursive algorithm which is greedy and not restricted to dyadic partitions and a wedgelet decorated dyadic partition which is optimal. Both schemes developed here are computationally fast and unsupervised. Our results illustrate the flexibility of both schemes in detecting variations in both intensity and variance. We compare our schemes with an existing constant-variance CART algorithm and conclude that the WDDP outperforms the ARP.



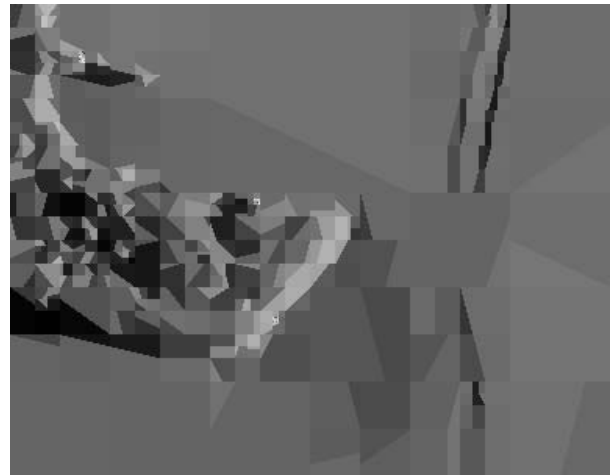
(a)



(b)



(c)



(d)

Figure 4. MSTAR SAR amplitude data HB06170 (a) original image (b) ARP-segmented image (c) WDDP-segmented image (d) constant-variance WDDP segmentation

REFERENCES

1. R. Fjrtoft, P. Marthon, A. Lops, and E. Cubero-Castan, "An optimal multiedge detector for SAR image segmentation," *IEEE Trans. Geoscience and Remote Sensing (TGARS)* **36**, pp. 793–802, May 1998.
2. V. Venkatachalam, H. Choi, and R. Baraniuk, "SAR image segmentation using wavelet-domain hidden markov tree models," *Proc. SPIE, International Symposium on Aerospace/Defense Sensing, Simulation, and Controls, Algorithms for SAR Imagery*, (Orlando, Florida), April. 2000.
3. D. Stewart, D. Blacknell, A. Blake, R. Cook, and C. Oliver, "An optimal approach to SAR image segmentation and classification.," *IEEE. Trans. on Radar, Sonar and Navigation* **147**, June. 2000.
4. R. D. Nowak and M. A. T. Figueiredo, "Unsupervised progressive parsing of poisson fields using minimum description length criteria," *Proc. ICIP II*, pp. 26–29, (Kobe, Japan), 1999.
5. V. Venkatachalam, R. Nowak, R. Baraniuk, and m. Figueiredo, "Unsupervised SAR image segmentation using recursive partitioning," *Proc. SPIE, International Symposium on Aerospace/Defense Sensing, Simulation, and Controls, Algorithms for SAR Imagery*, (Orlando, Florida), April. 2000.
6. J. Rissanen, *Stochastic Complexity in Statistical inquiry*, World Scientific Publishers, Singapore, 1989.

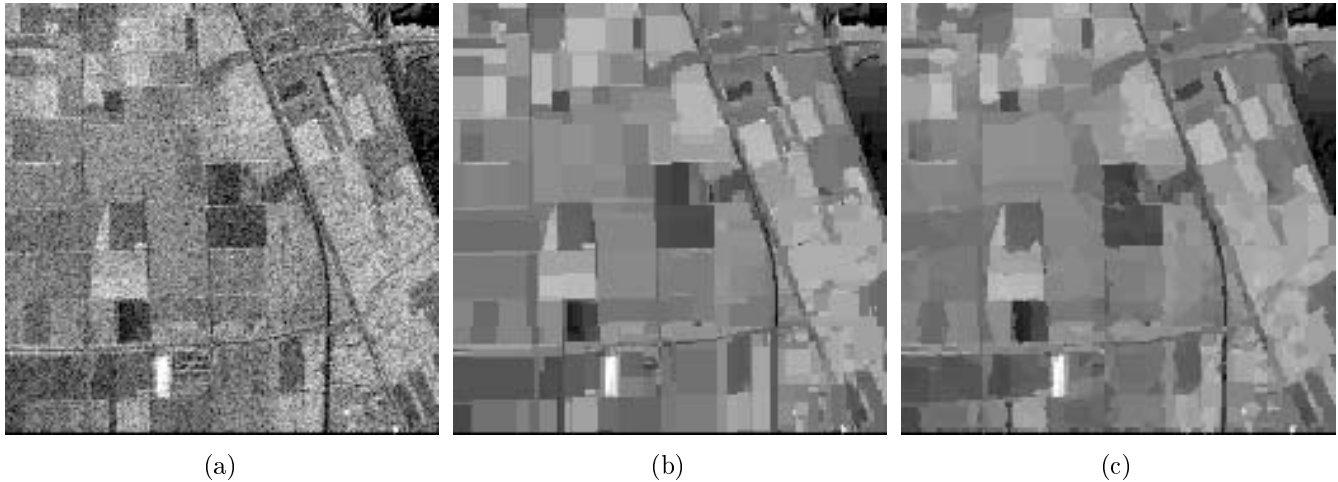


Figure 5. Multi-look SAR image of agricultural fields in Flevoland, Netherlands, (a) original image (b) ARP segmented image (c) WDDP segmented image

7. N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," *Wavelets in Geophysics*, Foufoula-Georgiou and Kumar (eds.), Academic Press, 1994.
8. H. Krim and I. Schick, "Minmax description length for signal denoising and optimal representation," *IEEE Trans. on Information Theory* **45**(3), April, 1999.
9. D. L. Donoho, "Wedgetlets: Nearly-minimax estimation of edges," *Annals of Statistics* **27**, pp. 859–897, 1999.
10. L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*, Wordsworth and Brook, Pacific Grove, CA, 1984.
11. D. L. Donoho, "Cart and best-ortho-basis: A connection," *Annals of Statistics* **2**, pp. 1870–1911, 1997.
12. MSTAR, "public clutter DBA7B02AA 2-CD set," www.mbvlab.wpafb.af.mil/public/MBVDATA, Sept. 1995.
13. M. H. Hansen and B. Yu, "Model selection and the principle of minimum description length," *Journal of the American Statistical Association*, to appear, 2001.
14. J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, Singapore, 1989.
15. R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, J. Wiley and Sons, Chichester, UK., 1980.
16. N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 1, J. Wiley and Sons, Chichester, UK., 1994.
17. T. Cover and J. A. Thomas, *Elements of Information Theory*, J. Wiley and Sons, New York, 1992.
18. J. Rissanen, "Fisher information and stochastic complexity," *IEEE Transaction on Information Theory* **42**, pp. 40–47, 1996.
19. J. Dias, T. Silva, and J. Leitao, "Adaptive restoration of speckled SAR images using a compound random markov field," *Proc. ICIP* **TA3.02**, pp. 859–897, 1998.